Contents

CHAIR’S FOREWORD 2
DIRECTOR’S REVIEW AND HIGHLIGHTS 3
ORGANISATION AND GOVERNANCE 4
Organisational chart 4
Advisory Board 5
Research Committee 8
Director 8
Chief Investigators 8
Partner Investigators 13
Special advisor 14
Information infrastructure 14
Centre Manager 14
Postdoctoral Fellows / Research Officers 15
Research Assistants 15
Masters/PhD students 15
Honours students 16
International & national interns 16
Undergraduate research trainees and volunteers 17
Visiting scholars, associates 17
STRATEGIC DIRECTION (2008-2010) 18
Vision 18
Mission 18
Aims 18
Objectives 18
Capabilities — core technologies and infrastructure 18
RESEARCH PROGRAMS 19
Program 1 –The Visible Cell®: linking genome and phenome 20
1.1 Generation of subcellular localisation data in mammalian cells 20
1.2 Project and data management for Visible Cell® 22
1.3 Data analysis and knowledge discovery in very large biological network databases 24
Program 2 – Phenotype-informed discovery of networks and systems 26
2.1 Computational discovery and modelling of gene-regulatory networks 26
2.2 Differential expression of genes towards pathway discovery 27
2.3 Discovery and modelling of RNA-based regulatory networks 28
Program 3 – Modelling dynamic cellular processes 30
3.1 Mathematical modelling of genetic regulatory networks and biochemical pathways 30
3.2 Modelling and visualising biological processes in complex spatial environments 32
Program 4 – Algorithms for graphs and networks 34
4.1 Discrete optimisation and graph clustering 34
4.2 Optimisation-based approaches to large systems 35
4.3 External-memory algorithms for analysis of massive graphs 37
KEY PERFORMANCE INDICATORS 39
1. Research findings and competitiveness 40
Publications 40
Quality of publications 40
Number of publications 40
Patents and trademarks 40
Invitations to address and participate in conferences 41
International conferences 41
National conferences 44
Invitations to visit leading international laboratories, major research groups and centres 44
Commentaries about the Centre’s achievements 46
Evidence of interdisciplinary linkages within the Centre 46
2. Research Training and professional education 47
Recruitment 47
Professional/technical training and advanced career development 47
Seminar/Symposium hosted and/or sponsored by ACB 47
Other research and training and professional education 49
Postgraduate and undergraduate courses in the Centre’s area 49
Scientists in Schools Partnership 49
3. International, national and regional links and networks 50
Visitors 50
International visitors 50
National visitors to one or more ACB nodes 50
Collaborative national and international workshops and exchanges 50
Other research linkages 51
Membership on national and international professional committees 53
Research projects with international partners 54
4. End-user links 55
Commercialisation activities 55
Government, industry and business briefings 55
Trained/ing in technology transfer and commercialisation 55
Public awareness programs 55
5. Organisational support 56
New organisations recruited to or involved in the Centre 56
Infrastructure 56
Annual cash contributions from other organisations 2008–2010 57
Annual in-kind contributions: collaborating organisations 2008–2010 57
6. National benefit 58
Measures of expansion of Australia’s capability in the priority area(s) 58
APPENDIX 1: PUBLICATIONS 60
APPENDIX 2: FINANCIAL STATEMENT 73
I am pleased to introduce you to the 2008 Annual Report of the ARC Centre for Bioinformatics (“aCB”).

2008 was a year of transition for aCB. In 2007, aCB was awarded a three-year extension by ARC, and was upgraded by ARC to an ARC Centre of Excellence. Therefore, this 2008 Annual Report provides an overview of the Centre’s first year of operation as the “ARC Centre of Excellence in Bioinformatics”.

This second round of Federal Government support now permits aCB to build on its existing research portfolio and to reinforce its capabilities and the interests of its Investigators and their research teams across both existing and new interdisciplinary research programs and between geographical nodes. New Investigators have joined the Centre from three institutions new to the aCB: Macquarie University, the University of Auckland, and the University of Tennessee. Collaborations have also continued to grow outside the Centre, as the co-authorship of grants and research publications, listed later in this report, amply demonstrates.

aCB became a key node in the EU’s Sixth Framework Information Society Technology project “Quasi-opportunistic supercomputing for complex systems in Grid environments” (DosCosGRID). The technical manager, Chris Kurovski, worked out of aCB during 2008 to manage the project from Australia.

ACB Director Mark Ragan represented The University of Queensland on the Working Group of EMBL Australia, the body that manages Australia’s associate membership in the European Molecular Biology Laboratory (EMBL).

Most importantly, aCB’s research excellence continued to be recognised internationally in 2008. Professor Michael Fellows and Dr Frances Rosamond worked in Jena, Germany for much of 2008 in support of Michael’s von Humboldt research award. Professor John Mattick was elected a Fellow of the Australian Academy of Science, and became the inaugural Gutenberg Chair at the Université Louis Pasteur in Strasbourg. Professor Geoff McLachlan became president-elect of the International Federation of Classification Societies. Professor Kevin Burrage continued his joint appointment at Oxford University as Professor of Computational Systems Biology. Professor Mark Ragan facilitated the National Technology Roadmap for Bioinformatics with the Ministry of Science Technology and Innovation in Malaysia.

I anticipate that 2009 will be another year of considerable achievement for the Centre as it continues to develop its activities under the second term of its support from ARC.

Dr Gregory R. Smith
Chair, Advisory Board
March 2009
2008 was a year of transition for our Centre. In its first phase (2003-2007), the ARC Centre of Excellence in Bioinformatics brought together researchers from diverse disciplines, institutions and regions to build new research programs in bioinformatics, linking the mammalian genome with its expression in the cellular phenotype. We learned to see problems from each others’ perspectives, work in new scientific languages, and apply unfamiliar tools. We engaged researchers outside our Centre and institutions, contributed to new partnerships and the development of national research infrastructure, and initiated our successful Winter School in Mathematical and Computational Biology. In mid-2006 we applied for a more-ambitious three-year extension of our Centre, and on 16 March 2007 we learned that ARC would fund our proposed extension – and upgrade our status to that of an ARC Centre of Excellence. These pages report on our first year as the ARC Centre of Excellence in Bioinformatics.

This second phase of our Centre differs in several respects from the first:

Firstly, we are joined by Investigators at three institutions new to our Centre: Professor Shoba Ranganathan (Macquarie University), Professor Allen Rodrigo (The University of Auckland), and Professor Michael Langston (The University of Tennessee). Associate Professor Pablo Moscato, a longstanding collaborator, has formally joined us via our founding partner, The University of Newcastle; and at The University of Queensland we welcome new Investigators Dr Brad Marsh and Professor John Mattick, while Dr Sean Grimmond, Dr Ralf Muhlbirger and Professor Bernard Pailthorpe did not continue with us into Phase II. Collaborations continue (and indeed are growing) with Sean and Bernard, and I’m confident that upcoming Annual Reports will continue to show their names as co-authors on grants and research publications, and as collaborators in infrastructure initiatives.

Two new members, whom I had the privilege of introducing in last year’s Annual Report, augmented our distinguished Advisory Board in 2008: Professor Peter Donnelly FRS, Director of the Wellcome Trust Centre for Human Genetics, and Professor of Statistical Science at Oxford University, and Professor Barney Glover, Deputy Vice-Chancellor of The University of Newcastle.

Phase II of our Centre will see a progressive shift in emphasis from method and tool development, to application; from integrating large datasets, to their common use by multiple Investigators in our Centre; from building data infrastructure, to using it to address complex problems. In Phase I our external alliances emphasized technology; while valuing their continuance, we now add new alliances in data-generation, experimental validation and (perhaps) large-scale service provision.

Establishing – and during 2008, expanding – a Centre necessarily involves legal formalities, and I’m pleased to report that the funding contract and inter-institutional agreement formally extending our Centre as the ARC Centre of Excellence in Bioinformatics were signed by all parties in late December 2008. With these in place, we can now flow research funds to our new programs and projects. In 2008 we wrapped-up projects from Phase I, and began to put our plans for Phase II into action; 2009 will be a year of consolidation and scaling-up. Early outcomes will appear from some of our new projects.

Change continues to accelerate in our external environment: 2009 is radically different from mid-2006, when we wrote the application for extension to Phase II. The National Collaborative Research Infrastructure Strategy has engaged bioinformatics not only through Bioplatforms Australia, but also through the Australian National Data Service and National Computational Infrastructure. EMBL Australia is operational, and an initiative is underway to establish a major data services mirror in Australia. The Victorian Life Science Computing Initiative has been announced, and bioinformatics is firmly on the radar in Innovation departments within several state governments. At any given point in time, several institutions in Australia are recruiting bioinformaticians – and for the most part these are not universities, but established companies, medical research institutions, primary industries bodies, biotech start-ups, consultancies and others. The National Health and Medical Research Council, the National Breast Cancer Foundation, and several state governments have funded programs or centres in bioinformatics. Bioinformatics Australia, our national society organised within AusBiotech, is now an affiliate of the International Society for Computational Biology, and – fresh from hosting the 19th International Conference on Genome Informatics (GIW-2008) in December 2008, is busily planning its fourth annual conference (to be held in Melbourne). And on the research front, next-generation DNA sequencing is onshore in a big way – we are all very soon to see how robust and scalable our bioinformatics really is.

On behalf of my colleagues in the ARC Centre of Excellence in Bioinformatics, I encourage you to peruse this report of our 2008 activities, and invite your comments.

Mark Ragan
Professor Mark Ragan
Director
ARC Centre of Excellence in Bioinformatics
March 2009
Organisation and Governance

Organisational Chart

UQ – DVC (R)  Advisory Board

Director (RC Chair + COO)

Information Infrastructure  Centre Manager

Research Committee (RC)

PROGRAM 1  The Visible Cell®: linking genome and phenome

Visible Cell® Project Manager  Coordinator Data Integration  Coordinator Cell Function

Research Team  Research Groups  Research Groups

PROGRAM 2  Phenotype-informed discovery of networks and systems

Coordinator Regulatory Networks

Research Groups

PROGRAM 3  Modelling dynamic cellular processes

Coordinator Modelling

Research Groups

PROGRAM 4  Algorithms for graphs and networks

Coordinator Networks

Research Groups
Advisory Board

The 8th Advisory Board meeting was held on Thursday 28 August 2008. The members of the Advisory Board are:

1. Dr Gregory R. Smith (Chair)
   
   **Director**
   
   SciVentures Investments Pty Ltd
   
   Melbourne VIC, Australia

   Dr Greg Smith is a cofounder and director of SciVentures Investments Pty Ltd, the manager of the SciVentures Pre-Seed Fund. He also is a director 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. Peter is a Fellow of both the Royal Society and Academy of Medical Sciences. He 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.
Prof. Simon Easteal
John Curtin School of Medical Research
The Australian National University
Canberra ACT, Australia

Prof. Simon Easteal is Deputy Director of The John Curtin School of Medical Research at the Australian National University. His research is focused on how the evolutionary dynamic between humans and their environments has shaped the complexity of human biology, given 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.

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.

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 in 2008 was Chair 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, 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 FASSA  
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’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 1993-94, and deputy chair of the ARC Research Grants Committee 1994. He chaired the National Committee of DVCs/PVCs (Research) in 2001 and was convenor 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; CRCMining; and Australian Genome Research Facility Ltd.

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

In 2006 Professor Wainwright was appointed Director of the Institute for Molecular Bioscience (IMB) at the University of Queensland, 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 research focus 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 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 90 publications, and co-editor of two books. He is the Director of the NSW Systems Biology Initiative, a member of the Advisory Board of the Australian Proteomics Computational Facility (APCF) and the Management Committee of the Ramaciotti Centre for Gene Function Analysis. Marc is a Senior Editor of the journal Proteomics, and serves on the Editorial Boards of Practical Proteomics and Bioinformatics and Biology Insights. He has six years experience in industry, and has co-founded two biotechnology companies, one of which, Proteome Systems, employs ~30 staff and is listed on the Australian Securities Exchange. 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 2008. Members of the ARC Centre of Excellence in Bioinformatics Research Committee for 2008 were:

Chair  Prof. Mark Ragan
Program 1  Dr Rohan Teasdale
Program 1  Prof. Shoba Ranganathan
Program 2  Dr Timothy Bailey
            (from September 2008)
Program 2  A/Prof. Phoebe Chen
            (to September 2008)
Program 2  Prof. Geoff McLachlan
Program 2  Prof. John Mattick
Program 3  Prof. Kevin Burrage
Program 4  A/Prof. Pablo Moscato
Secretary  Ms Lanna Wong

Director

1 Prof. Mark Ragan FLS
Institute for Molecular Bioscience
The University of Queensland
Please refer to page 6 for profile.

Chief Investigators

2 Dr Timothy L. Bailey
Institute for Molecular Biology
The University of Queensland

Dr Timothy Bailey has been active in bioinformatics and computational biology since the early 1990s, and since 2004 he has been a senior research fellow at the Institute for Molecular Biology. 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 recognitio

Director

1 Prof. Mark Ragan FLS
Institute for Molecular Bioscience
The University of Queensland
Please refer to page 6 for profile.

Chief Investigators

2 Dr Timothy L. Bailey
Institute for Molecular Biology
The University of Queensland

Dr Timothy Bailey has been active in bioinformatics and computational biology since the early 1990s, and since 2004 he has been a senior research fellow at the Institute for Molecular Biology. 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. During the past year, Dr Bailey’s group, in collaboration with the Noble group at the University of Washington, introduced a new website featuring a suite of sequence analysis tools. This “MEME suite of motif-bases sequence analysis tools” includes motif discovery algorithms (MEME and GLAM2), motif-based sequence search tools (MAST, FIMO and GLAM2SCAN) and tools for analysing the functions of DNA sequence motifs in regulatory networks (TOMTOM and GOMO). The new MEME suite website is currently used by about 1,000 biologists per month.
3 Prof. Kevin Burrage
Institute for Molecular Bioscience
The University of Queensland; and
Professor of Computational Systems Biology
COMLAB, Oxford University, UK
Prof. Kevin Burrage was 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 180 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, and the role of noise in molecular systems. 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 and 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. A/Prof. Chen founded Asia-Pacific Bioinformatics Conference and continues to chair its Steering Committee. She also serves on the national 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, New Zealand) were nominated for the Gödel Prize in 2005. They co-authored the foundational papers and the first comprehensive monograph for the field, 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 an Alexander von Humboldt Research Award, and in 2007 of an Inaugural Fellowship to the Durham University Institute of Advanced Studies, Prof. Fellows spent much of 2007 and 2008 with research groups in Europe as an invited distinguished visiting researcher. He is an Associate Editor for the Journal of Computer and Systems Sciences, and (since 2008) for the A.C.M. Transactions on Algorithms.
6 Dr Markus Hegland
Centre for Mathematics and its Applications
Mathematical Sciences Institute
The Australian National University
Dr Markus Hegland is a member of the Computational Mathematics Program at the ANU’s Centre for Mathematics and Its Applications. He is known internationally for his work on ill-posed problems, parallel solution of linear systems of equations and fast Fourier transforms, and, most recently, high-dimensional approximation. He has pioneered the application of numerical techniques, in particular sparse grids, in data mining applications. Finally he has developed chemical master equation solvers which were shown to work for up to 100 different chemical species. He was co-initiator of the Australasian Data Mining Conference (AusDM) and of the High-Dimensional Approximation Workshop (HDA), which annually attracts some of the most prominent researchers of the field world-wide to Australia.

7 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 discrimant 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, focusing on the statistical analysis of microarray gene expression data. A joint Wiley monograph on the analysis of microarray data was published in 2004. In 2006 he was awarded a Professorial Research Fellowship for the period 2007-2011 by the Australian Research Council, and in 2007 was appointed to the ARC College of Experts. He is also Fellow of American Statistical Association, Australian Mathematical Society, Royal Statistical Society, and President-elect of the International Federation of Classification Societies.

8 Dr Brad Marsh
Senior Research Fellow and Group Leader
Institute for Molecular Bioscience
The University of Queensland
Prof. John Mattick FAA 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), the CSIRO Eureka Prize for Leadership in Science (2006) and the Julian Wells Medal of the Lorne Genome Society (2009). 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 and a Fellow of The Australian Academy of Science in 2008. Prof. Mattick serves on the advisory boards of several institutes in Australia and abroad. Professor Mattick receives no direct funding from the Centre.

A/Prof. Pablo Moscato
Priority Research Centre for Bioinformatics,
Biomarker Discovery and Information-Based Medicine
The University of Newcastle and Information-based Medicine Program
Hunter Medical Research Institute

A/Prof. Pablo Moscato was founding Director of the Newcastle Bioinformatics Initiative (2002) and, in 2006, became the Co-Director and Founder of the Priority Research Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine of The University of Newcastle, Australia. He is also one of the four members of the Strategic Research Advisory Panel of the Hunter Medical Research Institute. He is a member of the editorial boards of Journal of Mathematical Modelling and Algorithms, Journal of Heuristics, and in 2008 Memetic Computing. Pablo introduced memetic algorithms in 1989 when he was a member of the Caltech Concurrent Computation Program, and has championed this field for two decades. He has shown how successful the application of these hybrid computing methodologies to large-scale problems in combinatorial optimisation and bioinformatics can be. Pablo has served on program committees of numerous international conferences in bioinformatics, heuristics and optimisation and is currently leading research programs in data analysis from high-throughput technologies as applied to the study of samples from Alzheimer disease, multiple sclerosis, melanoma, and breast and prostate cancer.
11 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 ITEE. 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.

12 Dr Rohan D. Teasdale
Institute for Molecular Bioscience
The University of Queensland
Dr Rohan Teasdale holds a Senior Research Fellowship A 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. Rohan has 65 research publications, with many papers in high-profile and leading bioinformatics and cell biology journals. These publications have focused on how individual proteins are compartmentalised within the cell and defining the protein trafficking machinery responsible for this process. In addition he has numerous publications outlining novel bioinformatics methodologies, database resources and collaborative projects.

13 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). 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 and the Institute for Infocomm Research, Singapore. 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. She was re-elected as President of the Asia-Pacific Bioinformatics Network (2008-2011).
Partner Investigators

14 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.

15 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.
16 Prof. Allen Rodrigo

Director, Bioinformatics Institute
The University of Auckland,
New Zealand

Allen Rodrigo is Professor of Computational Biology and Bioinformatics, and Associate Dean of IT at the Faculty of Science 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 and on the Editorial Board of Mathematical Biosciences. He also sits on the Scientific Advisory Board of two bioinformatics companies, is a Principal 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.

Special advisor

17 Prof. Gene Myers

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

Prof. Gene Myers was a foundation group leader at 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 research is now focusing on algorithms for fast image searching and comparison. Gene’s 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 Zürich. Gene has been an Honorary Professor in the Institute for Molecular Bioscience at UQ since 2003.

Information infrastructure

Mr Matthew Bryant
System Administrator, High-performance computing

Mr Oliver Cairncross
Project Manager, Visible Cell®

Ms Mhairi Marshall
Grid & Database Developer

Mr Tim McComb (to October 2008)
Application Developer

Dr Andrew Newman (from December 2008)
Application Developer

Mr Timothy Sullivan (to October 2008)
Application Developer

Mr David Wood
Database Administrator/Developer

Centre Manager

Ms Lanna Wong, MBA
Postdoctoral Fellows / Research Officers

Dr Marjan Asakarian-Amiri
Dr John Belward (part-time)
Dr Mikael Bodén
Dr Pamela Burrage
Dr Qingsheng Chen
Dr Adrian Coote
Dr Larry Croft
Dr Aaron Darling
Dr Melissa Davis
Dr Marcel Dinger
Dr Gabriel Fung
Dr Martin Gómez-Ravetti
Dr Nicholas Hamilton
Dr Martin Hansen
Dr John Hawkins
Dr Karin Kassahn
Dr Krzysztof Kurowski (to November 2008)
Dr Kim-Anh Lê Cao (from November 2008)
Dr Fawang Liu (part-time)
Dr Cheong Xin Chan (to April 2008)
Dr Shev MacNamara
Dr Igor Makunin
Dr Lorenzo Malquori
Dr Harald Oey
Dr Neelima Pottekkat Sidharthan
Dr Fran Rosamond
Dr Osvaldo Anibal Rosso (to August 2008)
Dr Muhammad Shoaib Sehgal
Dr Zheng Yuan

Research Assistants

Dr Cheong Xin Chan (to April 2008)
Ms Seetha Karunarathne
Mr Lloyd Flack
Mr Daniel Johnstone (from November 2008)
Ms Ranjeeta Menon
Ms Chikako Ragan

Masters/PhD students

Supervision/Associate supervision (principal advisor in ACB):

Farah Abdullah (PhD awarded, 2008)
Advisor: Prof. Kevin Burrage
Paulo Amaral (due to complete in 2010)
Advisor: Prof. John Mattick
Isye Arieshanti (MPhil)
(due to complete in 2009)
Advisor: Dr Mikael Bodén
Associate Advisors: Dr Timothy Bailey & Dr Stefan Maetcheke
Rajith Aturaliya (PhD awarded, 2008)
Advisor: Dr Rohan Teasdale
Denis Bauer (due to complete in 2009)
Advisor: Dr Timothy Bailey
Durgaprasad Bollina (PhD awarded, 2008)
Advisor: Prof. Shoba Ranganathan
Fabian Buske (due to complete in 2012)
Advisor: Prof. John Mattick
Associate Advisor: Dr Timothy Bailey
Alhadi Bustamam (due to complete in 2009)
Advisor: Prof. Kevin Burrage
Associate Advisors: Dr Nicholas Hamilton & Dr Shoaib Sehgal
Pierre Cattenoz
(due to complete in 2011)
(Co-tutelle with the University of Strasbourg)
Advisors: Prof. John Mattick, Prof. Eric Westhof (University of Strasbourg)
Elsa Chacko (due to complete in 2011)
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 2009)
Advisor: A/Prof. Phoebe Chen
Joo-Young Choi (due to complete in 2009)
Advisor: Prof. Mark Ragan
Associate Advisor: Dr Melissa Davis
Khar Heng Choo (due to complete in 2009)
Advisor: Prof. Shoba Ranganathan
Associate Advisor: A/Prof. Tan Tin Wee
Michael Clark (due to complete in 2010)
Advisor: Prof. John Mattick
Associate Advisor: Dr Rohan Teasdale
Adam Costin (due to complete in 2010)
Advisor: Dr Brad Marsh

Alex Foo (due to complete in 2010)
Advisor: Dr Brad Marsh
Jitendra Gaikwad (due to complete in 2010)
Advisor: Prof. Shoba Ranganathan
Mohammad Islam (due to complete in 2010)
Advisor: Prof. Shoba Ranganathan
Associate Advisor: Dr Abhaya Nayak
Chol Hee Jung (due to complete in 2009)
Advisor: Prof. John Mattick
John Hawkins (PhD awarded, 2008)
Advisor: Dr Mikael Bodén
Associate Advisors: Dr Rohan Teasdale & Prof. Janet Wiles
Mario Inostroza-Ponta (PhD thesis submitted, 2008)
Advisor: A/Prof. Pablo Moscato
Associate Advisor: Dr Regina Berretta
Javed M. Khan (due to complete in 2011)
Advisor: Prof. Shoba Ranganathan
Varun Khanna (due to complete in 2011)
Advisor: Prof. Shoba Ranganathan
Keith Knapp (PhD to be awarded in April 2009)
Advisor: A/Prof. Phoebe Chen
Chin Foon Kho (due to complete in 2009)
Advisor: Dr Markus Hegland
Gaurav Kumar (due to complete in 2012)
Advisor: Prof. Shoba Ranganathan
Shev MacNamara (PhD awarded, 2008)
Advisor: Prof. Kevin Burrage
Robert McLeay (due to complete in 2011)
Advisor: Dr Timothy Bailey
Scott Mann (due to complete in 2009)
Advisor: A/Prof. Phoebe Chen
Ranjeeta Menon (due to complete in 2011)
Advisor: Prof. Shoba Ranganathan
Tim Mercer (due to complete in 2009)
Advisor: Prof. John Mattick
Nurul Mohamad (MPhil)
(due to complete in 2009)
Advisor: Dr Mikael Bodén
Associate Advisor: Dr Rohan Teasdale
Katrina Monico (due to complete in 2009)
Advisor: Prof. Geoff McLachlan
Shivashankar H. Nagaraj
(PhD awarded 2008)
Advisor: Prof. Shoba Ranganathan
Associate Advisor: Prof. Robin Gasser
Satu Nahkuri (due to complete in 2009)
Advisor: Prof. John Mattick
Co-supervision (principal advisor not an ACB investigator)

Agnes Boskovitz (PhD awarded, 2008)
Associate Advisor: Dr Markus Hegland
Advisor: Prof Rajeev Goré

Michael Bode (PhD awarded, 2008)
Associate Advisor: Prof Kevin Burrage
Advisor: Prof Hugh Possingham

Luke Hammond (due to complete in 2009)
Associate Advisor: Dr Nicholas Hamilton
Advisor: Prof Jenny Stow

Duncan Mortimer
(PhD thesis submitted, 2008)
Associate Advisor: Prof Kevin Burrage
Advisor: Dr Geoffrey Goodhill

Radosav Pantelic (PhD awarded, 2008)
Associate Advisor: Dr Nicholas Hamilton
Advisor: Dr Ben Hankamer

Mitchell Stanton-Cook
(due to complete in 2009)
Associate Advisor: Dr Nicholas Hamilton
Advisor: Dr Thomas Huber

Daniel Johnstone (due to complete in 2009)
Associate Advisors: A/Prof. Pablo Moscato & Prof. Rodney Scott
Advisor: Dr Elizabeth Adrienne Milward

Honours students

Rees Butterworth
Geometric hashing for the discovery of protein binding surfaces
Advisor: Dr Mikael Bodén

Vinh Dang
Divergence in expression patterns and functions of duplicated genes in zebrafish (Danio rerio)
Advisors: Prof. Mark Ragan & Dr Karin Kassahn

James Johnson
Volunteer computing for statistical phylogenetics
Advisors: Prof. Mark Ragan & Dr Aaron Darling

Pratishtha Vedalankar
The distribution of phosphoinositide binding proteins in the mammalian endosomes
Advisor: Dr Rohan Teasdale

International & national interns

International

Pricilia Boerleinder
Hogeschool Rotterdam, University of Rotterdam (The Netherlands)
Advisor: Dr Rohan Teasdale

Khar Heng Choo
National University of Singapore (Singapore)
Advisor: Prof. Shoba Ranganathan
Associate Advisor: A/Prof. Tan Tin Wee

Nikhil Lilaria
Indian Institute of Technology, Bombay (India)
Advisor: Prof. Mark Ragan

Camille Maumet
Institut National des Sciences Appliquées (France)
Advisor: Prof. Geoff McLachlan

National

Nicole Maggacis
The University of Queensland
Advisor: Dr Rohan Teasdale
Undergraduate research trainees and volunteers

Zoran Boskovic
Research trainee
A software suite to integrate various regulatory network inference models
Advisor: Dr Mohammad Shoabi Sehgal

Andres Esteban Marcos
Research trainee
SeEvolution, 3D visualization of genomic evolution
Advisor: Dr Aaron Darling

Michal Lorenc
Research trainee
A plasma membrane visualisation
Advisor: Dr Pamela Burrage

Cindy Yan
Undergraduate trainee/volunteer
Conserved non-coding elements associated with duplicated zebrafish genes
Advisor: Dr Karin Kassahn

Visiting scholars, associates

Visiting sabbatical scholars

Prof. Michael Zuker, Professor of Mathematical Sciences, Rensselaer Polytechnic Institute (USA)

Dr Robert Charlebois, Sanofi Pasteur (Canada)

Dr Brian Dalrymple, CSIRO Livestock Industries

Prof. Bernie Degnan, School of Integrative Biology, The University of Queensland

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

Dr Dave Edwards, Australian Centre for Plant Functional Genomics

Prof. Charles Elkan, Department of Computer Science and Engineering, University of California, San Diego (USA)

Dr Martin Frith, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (Japan)

Dr David Green, High Performance Computing Group, Information Technology Services, The University of Queensland

Prof. Michael Grishkov, Purdue University (USA)

Dr Ben Hankamer, 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, Cray Inc., 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)

Caroline Hunter, School of Information Technology & Electrical Engineering, The University of Queensland

Dr Lars Jermin, 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, Seoul National University (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 for Complex Systems

Dr Alex Loukas, Queensland Institute of Medical Research

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

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

Prof. Mark Morrison, Science Leader Metagenomics, CSIRO Livestock Industries

Dr/Prof. William Noble, Department of Genome Sciences, School of Medicine, University of Washington, USA

Prof. Bernard Palteborough, CEO, Queensland Cyber Infrastructure Foundation

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

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

Dr/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 Vancouver (Canada)

Prof. Janet Williams, School of Information Technology & Electrical Engineering, The University of Queensland
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 in the first phase of the Centre (2003-2007).
• 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

Strategic Direction 2008–2010
Research Programs

Research programs highlights

The Centre’s research in 2008 was delivered through four programs, each of which was further organized as two or three projects. The following pages describe the aims of each of these projects, the Centre and other personnel involved, and the main achievements during 2008.

As described elsewhere in this Report, 2008 was a year of transition from the initial phase of our Centre to our new Centre of Excellence. We welcomed four new Chief Investigators and their research groups, while three of our former Chief Investigators did not continue with us. This renewal is strongly reflected in our realigned Research Programs and Projects. In place of three Programs we now have four, and in place of 13 Projects we now have 11.

No simple mapping exists between our former Projects and those in our new Centre of Excellence. Two have been brought forward largely intact, and three were discontinued. The other eight, however, have been re-mixed with the goal of better exploiting synergies, whether arising from our previous work or from our new Investigators and directions. We sharpen our research focus on cellular networks, separating discovery and inference (our new Program 2) from mathematical modelling (our new Program 3), with the latter incorporating modelling activity formerly managed as part of our Visible Cell® project. Two Projects in our Centre of Excellence are entirely new.

Chief Investigators in our new Programs 2 and 4 met in Newcastle on 22 August 2008 to refine program goals, agree on datasets, and develop timelines. Personnel in Programs 1 and 3 similarly met in Brisbane on 10 October 2008. Our detailed activity plans for 2009 are set out in a separate, closed document that will be submitted to ARC concurrently with this Annual Report.

In Program 1, we develop an integrative data and visualisation framework, the Visible Cell®, that allows biologists and modellers to interact with mammalian cells reconstructed from 3D electron tomographic or fluorescence confocal images, annotate cell structures using federated data (sequences, structures, pathways, networks, literature), and conduct quantitative analyses. We also develop tools and an interface to manage the analysis and integration of very large biomolecular data.

Information flows from genome to phenome through complex molecular networks and systems, and in Program 2 we examine the signals, pathways and regulatory mechanisms by which these networks operate and interconnect. Algorithmic and statistical approaches are being focused on understanding which pathways and networks are up- or down-regulated in complex disease including cancer, and on the discovery of new pathways and sub-networks. This Program includes innovative approaches to the discovery and modelling of RNA-based regulatory networks in mammalian cells.

Computational modelling and simulation allow us to describe the operation of these networks in dynamic, quantitative terms. Key biomolecules often occur transiently, in extremely small numbers and/or in highly restricted locations. Cellular structures move too, driven by mechanical or thermodynamic forces that are highly susceptible to mathematical modelling. In Program 3 we construct computational models of molecular networks and of the subcellular structures through which they transduce information to the cellular phenome. One of the main goals of our extended Centre will be to embed these models in the Visible Cell®, making it a dynamic environment for hypothesis-generation and -testing.

Representing molecular networks and systems mathematically requires us to work with large numbers of graphs, and Program 4 develops powerful algorithmic and combinatorial approaches that will make our models computationally tractable. These include data-reduction strategies, implementation of scalable algorithms for problems on very large graphs, and novel computational techniques to make very large problems computable on actual hardware.

Research programs 2008-2010

Program 1
The Visible Cell®: linking genome and phenome
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 Discrete optimisation and graph clustering
4.2 Optimisation-based approaches to large systems
4.3 External-memory algorithms for analysis of massive graphs
Program 1 – The Visible Cell®: linking genome and phenome

1.1 Generation of subcellular localisation data in mammalian cells

Investigators:
- Dr Rohan Teasdale (The University of Queensland)
- Prof. Mark Ragan (The University of Queensland)
- Prof. Kevin Burrage (The University of Queensland)
- Prof. Jane Hunter (The University of Queensland)
- Prof. John Mattick (The University of Queensland)
- Prof. Shoba Ranganathan (Macquarie University)
- Prof. Mark Smythe (The University of Queensland)
- Prof. Xiaofang Zhou (The University of Queensland)

In collaboration with:
- Dr Christopher Bouton (Pfizer RTC, Cambridge, MA USA)
- Prof. Kevin Burrage (University of Queensland)
- Prof. Jane Hunter (The University of Queensland)
- Prof. John Mattick (The University of Queensland)
- Dr Shoba Ranganathan (Macquarie University)
- Prof. Mark Smythe (The University of Queensland)
- Prof. Xiaofang Zhou (The University of Queensland)

Researchers:
- Research Associates:
  - Dr Melissa Davis (The University of Queensland)
  - Dr Nicholas Hamilton (The University of Queensland)
  - Ms Seetha Karunaratne (The University of Queensland)
  - Dr Karin Kassahn (The University of Queensland)
  - Dr Zheng Yuan (The University of Queensland)
- PhD students:
  - Ms Joo-Young Choi (The University of Queensland)
  - Mr Chang Jin Shin (The University of Queensland)
  - Ms Josefine Sprenger (The University of Queensland)

Aims:
Because eukaryotic cells are highly compartmentalised, a protein’s function is tightly enmeshed with its location (more precisely, the distribution of its location). Knowing the latter can greatly improve prediction of its function, while conversely, a comprehensive protein inventory is critical to understanding the function of cellular compartments. This project aims to generate unique primary data on subcellular localisation of proteins in mammalian cells, and make these data available via value-added data resources for further integration, analysis, computational modelling, simulation and visualisation.

Achievements (2008):
(a) We developed novel statistical tests for distinguishing subcellular imaging, and used them in the automatic discovery of proteins whose localisation has changed under e.g. drug treatment. We showed these tests to be more sensitive than any other available, and to be able to detect subtle changes such as endosomal redistribution. A paper describing the methods is currently under review for BMC Bioinformatics, and software supporting the methodology is being made available under open-source license (Hamilton, Teasdale).
(b) A paper describing the iCluster high-throughput image visualisation tool appeared in BMC Bioinformatics and was rated highly accessed. iCluster enables large image sets (~1500 images) to be viewed and interacted with in 3D. A presentation on iCluster at the inaugural Hunter Imaging Workshop (March 2008) led to an invited review by Hamilton on quantification of subcellular imaging for the journal Traffic; the manuscript is under review (Hamilton, Teasdale).
(c) We developed high-throughput microarray-based screening of ~12,000 ncRNAs and ~30,000 mRNAs to determine their nuclear or cytoplasmic location, and are initiating medium-throughput semi-automated in situ hybridisation approaches, combined with computational algorithms developed by Nick Hamilton, to identify ncRNAs associated with known and novel cellular compartments. In association with Dr Bodo Lange (Max Planck Institute for Molecular Genetics, Berlin) we identified three ncRNAs associated with centrosomes, and in association with Dr David Spector (Cold Spring Harbor Laboratories) showed that a ncRNA induced upon muscle differentiation is an essential structural component of a sub-nuclear domain called paraspecles (Genome Research 2009) (Mattick, Teasdale).
(d) We applied the ELPP pipeline, generating subcellular localisation data for an additional 50 proteins within the mouse proteome. Within 2008 we redirected this project to a focused set of proteins predicted to encode phosphoinositide-binding domains (Teasdale, Karunaratne).
(e) We defined the nuclear proteome by integrating data from the LOCATE database with data generated from the ELPP and literature-mining (Teasdale, Karunaratne). This major body of work was published in the high-impact journal Genome Biology.
(f) We maintained the LOCATE Subcellular Localisation Database, and added predictive subcellular localisation data based on the most-accurate prediction algorithms (Teasdale, Sprenger).
(g) We predicted features of proteins relevant to their subcellular localisation, e.g. transmembrane domains (TMDs) and localisation signals (Teasdale, Yuan).
(h) We are developing resources to enable a systems biology approach targeted at the mammalian endosome. By combining protein-protein and gene expression networks we are developing target sets of interest for shRNA knockdown phenotype assays (Teasdale, Sprenger).
(i) We are currently developing experimental methodologies to determine the subcellular distribution of ncRNA with the cell (Mattick, Teasdale).

(j) By computational analysis of five fish genomes, postdoctoral researcher Dr Karin Kassahn identified mechanisms by which genes that have arisen by whole-genome duplication have partitioned existing function or taken on new function. Novel regulatory control was identified by new patterns of temporal and/or spatial expression during embryogenesis. A comprehensive manuscript was submitted in December 2008 (Ragan).

(k) PhD student Chang Jin Shin completed the mapping of orthologs from proteomes of model organisms where subcellular localisation is well-known, to the human proteome, and identified six parameters (three biological, three annotation-related) that strongly influence the quality of the mapping. The LOCATE database, developed within ACB by CI Rohan Teasdale and colleagues, provided data for validation. A manuscript was accepted at BMC Systems Biology for publication in 2009. Analysis of interologs was largely completed during 2008, and a manuscript has been invited for a special issue of Proteomics (Ragan).

(l) Because determining ortholog sets is computationally expensive, we collaborated with CI Mike Fellows to explore FPT versions of VERTEX COVER algorithms (Project 4.1). A joint manuscript was submitted for publication in a Tier A mathematics journal (Fellows, Ragan).

(m) Because interacting proteins must at least transiently or conditionally be physically co-located, subcellular localisation constrains inference of protein-protein interaction (PPI). Postdoctoral researcher Muhammad Shoaib Sehgal implemented subcellular localisation information within a multi-stage Bayesian classifier for inference of network function; MATLAB code was developed and made available to collaborators, and a manuscript was initiated (Burrage, Ragan).

(n) Data resources including for protein-protein interactions and pathways were migrated to Resource Description Framework (RDF). With Professor Jane Hunter and Dr Christopher Bouton we demonstrated the use of Semantic Web technologies SPARQL and SWRL to query and reason over subsets of these data. Although the BioMANTA collaboration with Pfizer was concluded, these technologies were transferred to the core Visible Cell® environment (Project 1.2) leading to an Applications Note in Bioinformatics (2009) focused on semantic capabilities of the Visible Cell® visualisation environment Illoura™ (Cairncross, Marsh, Ragan).

(o) Separately from the BioMANTA collaboration, MSc student Joo-Young Choi and postdoctoral researcher Melissa Davis identified data sources and standards for description of small molecules (e.g. drugs and drug-like compounds), converted them to RDF, and demonstrated the use of SPARQL and SWRL to query jointly over small-molecule and biomolecular data. Ms Choi demonstrated that complex rule sets, e.g. Lipinski’s Rule of Five, can be represented as SWRL rules and used within a reasoning framework. Posters were presented at two international conferences. We continue to work with IMBcom to protect intellectual property in this project (Ragan).
1.2 Project and data management for Visible Cell®

**Aims:**
A defining characteristic of our Visible Cell® is its spatial matrix: an actual mammalian cell, imaged at high resolution using the advanced cryo-electron microscopy facilities at UQ. For each cell this matrix can be up to 10 TB in size, depending on how we compress certain data. In the first phase of our Centre, we designed, constructed and implemented a computational and data infrastructure that can handle TB image files and links to scaleable data grid environments.

This Project aims to integrate cellular reconstruction and bioinformatic data by incorporating a specific ontology and novel methods of processing semantic data, yielding a world-first Visible Cell® prototype environment. This environment will streamline querying, analysis and visualisation of cellular reconstructions enriched with external sequence, structure, pathway, network and functional data. It will provide data and parameters to drive simulations based on the enriched data, support semantic querying and other application in systems biology (Project 1.1), and utilise national e-research and other infrastructure initiatives.

**Achievements (2008):**
(a) Neelima Sidharthan, Timothy Pan and Brad Marsh established an in-house program for identifying factors that regulate/modulate key functions of the mammalian pancreatic beta cell, using a natural product discovery/multi-modal screening approach.
(b) Andrew Noske, Peter van der Heide and Brad Marsh developed methods and generated primary data that provide proof-of-concept for a 3D Visible Cell® atlas. This will involve technology development and proof-of-concept for high-resolution (<5 nm) and intermediate resolution (10-20 nm) whole-cell electron microscope tomography.
(c) The same group developed mathematical methods for pre- and post-processing cellular electron tomography data that expedite segmentation, quantitative analysis and computational simulation.
(d) To provide high-speed access to arbitrary subsets of large topographic data, in 2008 the Visible Cell® development team led by Oliver Cairncross coded and implemented an export utility from Oracle into IMOD format. We can now build a tomogram file (or a subsection of one) much more quickly than can be done using the command-line utilities in IMOD.
(e) In 2008 we completed the implementation of the Illoura™ data structure in RDF. Oliver Cairncross and Tim McComb developed and implemented a reasoner in LISP and SPARQL that makes spatial data that do not natively have semantic value available, in semantic format, on-demand to inference engines. This solution appears to be general for spatially resolved data, not only in bioscience. We demonstrated semantic querying jointly over spatial (cell tomographic) and biomolecular pathway data. An Applications Note was submitted to Bioinformatics and will appear in early 2009.
(f) We explored out-of-memory execution of queries over tomographic data using new Semantic Web functionality in Oracle 11g, but have postponed decision on whether to integrate it into Illoura™.
(g) David Wood, Oliver Cairncross and Tim McComb implemented a bridge between the Illoura™ visualiser and SRS, allowing SPARQL queries to seamlessly incorporate information found in federated data sources. This bridge is described in some detail in the Illoura™ web pages.
(h) We continue work to extend the query engine so queries can be distributed across multiple machines and query handlers; this will add scalability to the query engine and will allow e.g. the offloading of a computationally complex routine to a purpose-built query handler. In 2008 we re-implemented some computationally intensive aspects of key queries in C++.
(i) We shifted functionality in the VTK visualisation library from Illoura™ to external J3D code. In view of other priorities, we chose not to undertake major modification of the user interface during 2008.

(j) With the wrap-up of BioMANTA, we did not further augment the Visible Cell® ontology to support queries of protein accessibility for intercellular components, or refine protein-protein interaction maps. We continue to develop the core architecture to enable better interaction with external data stores, modelling environments and software.
1.3 Data analysis and knowledge discovery in very large biological network databases

**Investigators:**
- Prof. Shoba Ranganathan (Macquarie University)
- Prof. Xiaofang Zhou (The University of Queensland)

**In collaboration with:**
- Prof. Alon Halevy (University of Washington)
- Dr Brad Marsh (The University of Queensland)
- Prof. Mark Ragan (The University of Queensland)
- Dr Rohan Teasdale (The University of Queensland)
- A/Prof. Wei Wang (University of North Carolina)

**Aims:**
The Visible Cell® project typifies the data-analysis and knowledge-discovery challenges of omic bioinformatics, including demands for efficiency with very large data, complex data types and relationships, and multiplicity of sources and formats—problems exacerbated by the explorative nature of data analysis in biology. This project addresses these issues by (1) development of a portal that will investigate a new “pay-as-you-go” approach to data integration, allowing biologists to specify just-enough data mappings for a task. It thereby offers low upfront overhead and the ability to evolve with changes in data sources. The key to this incremental, collaborative approach will be a tool that supports online community annotation, automatic consistency enforcement for data mappings, dataset and resource discovery, dynamic dataset membership, data uncertainty and lineage analysis; and (2) data mining that models data and metadata as graphs, and automates the detection of frequent sub-graphs. Initial application will be to RDF diagrams, to sub-structures of protein surfaces indexed by frequency of geometric graphs, and to semi-automatic data linking (where topological structure similarity among annotations labelled differently due to semantic heterogeneity can be automatically detected and reported for the human user to verify). These tools will allow users of our Visible Cell® to specify only the biological data needed to address each question, without our having to incur the upfront cost of fully inter-relating all data across the hundreds of data sources we are mirroring.

**Achievements (2008):**

(a) Dr Gabriel Fung was co-located with Visible Cell® core development team (Project 1.2) for around 3 months in early 2008 to establish common objectives and mutual understanding of problems and technologies. We identified some major datasets (bio-literature, protein data) and scenarios (literature mining, name entity resolution, data integration) for problem formulation, characterisation, prototyping and testing. We identified potential users, and surveyed basic user requirements for usability and scalability. Several papers and posters appeared in 2008, and a manuscript dealing with entity resolution was prepared (Fung, Zhou).

(b) We constructed an integrated database linking SCL data from the LOCATE database (developed within the project) with protein interaction and metabolic reaction network data for human and mouse, and conducted preliminary analyses examining the utility of applying protein interaction and metabolic networks to the problem of SCL prediction. Details of the integrated database and some of the preliminary analysis were presented as a poster at GIW-2008. We are currently using this work as the basis for a network-based SCL prediction scheme, and are using InParanoid ortholog-prediction algorithm to extend both experimental and literature-based SCL data for human and mouse in the LOCATE database to a number of other species (Ranganathan).

(c) We identified conserved functional modules in human, fly and yeast protein interaction networks using the PHUNKEE network-comparison algorithm, developed previously by members of the project. We analysed SCL of the conserved functional modules using data from the LOCATE database, and demonstrated that the modules from these species mostly represented processes taking place in the cell nucleus, e.g. DNA repair. This work was presented as a poster at ECCB-2008 in Italy (Ranganathan).
(d) We constructed another integrated database designed specifically for network-based protein function prediction across multiple species. The database integrates protein interaction, sequence similarity relationships, function and SCL data from a variety of sources. We conducted a detailed preliminary analysis of this data for the purpose of evaluating various techniques for function prediction. We developed a sophisticated network-based function prediction methodology based on this initial work; the function-prediction analysis is ongoing. This work will eventually be integrated with the SCL work mentioned in (b) above (Ranganathan).

(e) We reviewed the literature on protein function and disease in relation to biological networks, and initiated writing a manuscript “Protein networks, function and disease” (Cootes, Ranganathan).
Program 2 – Phenotype-informed discovery of networks and systems

2.1 Computational discovery and modelling of gene-regulatory networks

Investigator:
Dr Timothy Bailey
(The University of Queensland)

In collaboration with:
A/Prof. Phoebe Chen
(Deakin University)
Prof. Geoff McLachlan
(The University of Queensland)
Prof. John Mattick
(The University of Queensland)
A/Prof. Osamu Maruyamu
(University of Kyushu)
A/Prof. Pablo Moscato
(The University of Newcastle)
A/Prof. Andrew Perkins
(The University of Queensland)

Researchers:

Research Associates:
Dr Mikael Bodén
(The University of Queensland)
Dr John Hawkins
(The University of Queensland)

Research staff:
Mr David Wood
(Queensland Facility for Advanced Bioinformatics)

PhD students:
Ms Denis Bauer
(The University of Queensland)
Mr Robert McLeay
(The University of Queensland)
Mr Thomas Whittington
(The University of Queensland)

Aims:
This project aims to:
(a) Construct gene-regulatory sub-networks incorporating regulation by transcription factors (TFs) and non-coding RNAs (ncRNAs).
(b) Build static computational models of the transcriptional output of a gene under the control of TFs and ncRNAs.
(c) Develop useful software tools implementing the above prediction and modelling tasks, and make them available to the international research community.

Achievements (2008):
(a) Thomas Whittington developed and published in Nucleic Acids Research an approach for making tissue-specific predictions of functional transcription factor binding sites. His method combines tissue-specific chromatin modification data (histone methylation) with standard in silico binding site prediction.
(b) Dr John Hawkins developed and published theoretical analyses of the power of comparative genomics-based transcription factor binding site prediction methods. This work confirmed the high degree to which functional binding sites are not fixed in the genome, but “drift” or are lost during evolution. Dr Hawkins implemented a new sequence search algorithm that takes these facts into account in order to improve in silico prediction accuracy. The first part of this work was published in RECOMB, and the second part has been submitted to ISMB, the premier conference in bioinformatics.
(c) Dr Mikael Bodén developed an algorithm for predicting the association of Gene Ontology terms with transcription factors. This aids in the discovery and characterisation of transcription factor motifs, which are vital components in building gene-regulatory models. This work was published in Nucleic Acids Research. Fabian Buske integrated this algorithm into the MEME suite of motif-based sequence analysis tools, making its functionality available via the web to biologists worldwide. Fabian also extended the method to utilize conservation information, and a manuscript is in preparation.
(d) Robert McLeay studied improved ways to discover transcription factor binding motifs in chromatin immunoprecipitation data. This work revealed that the sensitive statistical methods for scoring motifs suggested by other researchers do not work as reported. A simpler scoring method developed by our colleague Dr Martin Frith is superior. Robert is preparing a manuscript describing these results.
(e) Denis Bauer continued the development and testing of a model of transcriptional output of the Drosophila EVE gene. She demonstrated that the model can predict regulatory modules in other species, and explored numerous model optimisation strategies. Denis also created, published and made freely available STREAM, the first workbench for constructing static regulatory models. This work resulted in articles in Bioinformatics and BMC Bioinformatics, and a third manuscript has been submitted for review.
2.2 Differential expression of genes towards pathway discovery

**Investigator:**
Prof. Geoff McLachlan  
(The University of Queensland)

**In collaboration with:**
Dr Regina Berretta  
(The University of Newcastle)  
A/Prof. Phoebe Chen  
(Deakin University)  
A/Prof. Pablo Moscato  
(The University of Newcastle)  
Dr Muhammad Shoaib Sehgal  
(The University of Queensland)

**Researchers:**

**Research Associate:**
Dr Kim-Anh Lê Cao  
(The University of Queensland)

**Research Assistant:**
Mr Lloyd Flack  
(The University of Queensland)

**Aims:**
This project aims to:
(a) Extend statistical methods based on estimates of the empirical null appropriate for non-independent samples, and develop guidelines for their use.
(b) Develop pathway-aware modifications to the minimum class entropy approach to univariate quantisation developed jointly by Chen and Moscato, and to the combinatorial approach pioneered by Moscato.
(c) Thereupon undertake a unique cross-comparison of these three approaches based on public-domain microarray datasets, comparing results against annotated pathways to find the conditions under which each approach best recovers known genetic regulatory pathways or modules.
(d) More generally, investigate the benefits of pathway-centric methodologies for improving classification in data from microarray expression studies.

**Achievements (2008):**
(a) We developed and applied new statistical methods to analyse biological data produced by CRIs in the Centre, in particular by the Mattick Laboratory (McLachlan).
(b) During 2008, Moscato and Berretta collaborated on a study using a prostate cancer gene expression data (True et al., PNAS 2006). The overarching aim was to investigate the link between the TP53 network, its robustness, and its putative role to stratify the samples into groups. Analysis of the TP53 network of Kuhn et al. (first proposed in 1999) led to a new "pathway-aware" method to uncover the consequences of altered signalling events. A manuscript will be submitted in 2009 (Moscato).
(c) The Newcastle group analysed four genome-wide microarray datasets in prostate cancer; they will be integrated with other datasets available centrally in ACB, with the goals of helping new members of our groups to understand the available information, and facilitating collaborative analysis using both statistical and combinatorial optimisation approaches (McLachlan, Moscato).
(d) We considered the formal development of statistical methods for inferring networks, including measures of uncertainty about each link in the inferred network. These methods will be applied to inferring gene networks and miRNA networks using corresponding expression data (McLachlan).
(e) We continued with the development of user-friendly software implementing EMMIX-FDR, EMMIX-GENE and EMMIX-WIRE for use in the statistical analysis of gene-expression and other high-throughput data (McLachlan).
2.3 Discovery and modelling of RNA-based regulatory networks

**Investigators:**
- Prof. John Mattick  
  (The University of Queensland)
- A/Prof. Phoebe Chen  
  (Deakin University)
- A/Prof. Pablo Moscato  
  (The University of Newcastle)

**In collaboration with:**
- Dr Tim Bailey  
  (The University of Queensland)
- Dr Regina Berretta  
  (The University of Newcastle)
- Prof. Geoff McLachlan  
  (The University of Queensland)
- Dr Colleen Nelson  
  (Translational Research Institute, Princess Alexandra Hospital)
- Dr Carlos Riveros  
  (The University of Newcastle)

**Researchers:**
- **Research Associates:**
  - Dr Qingfeng Chen  
    (Deakin University)
  - Dr Marcel Dinger  
    (The University of Queensland)
- **Research Assistant:**
  - Mr David Wood  
    (Queensland Facility for Advanced Bioinformatics)
- **PhD student:**
  - Mr Feng Chen  
    (Deakin University)

**Aims:**

(a) Apply microarray technologies to discover novel non-protein-coding RNAs (ncRNAs) involved in the control of in mammalian systems under selected conditions of development, health and disease.

(b) Infer and characterise RNA-based regulatory networks involved in these systems, and identify the molecular interactions by which this control is transduced.

(c) Characterise candidate ncRNAs transcripts in human, and identify their homologs in other mammalian and vertebrate genomes and transcriptomes.

(d) Identify evolutionarily conserved secondary structures in ncRNAs and construct hypotheses concerning their function.

**Achievements (2008):**

(a) Using the V1 and V2 microarray chips, we generated novel expression signatures encompassing both protein-coding and non-protein-coding transcripts in tissue samples from breast cancer (Mattick), as well as a variety of differentiating cell systems including embryonal stem cells, neural stem cells, T-lymphocytes, muscle cells, epithelial cells, and the developing gonadal ridge. We are developing our collaboration with Dr Colleen Nelson (QUT) to extend these studies to prostate cancer, as well as with Dr Charles Perou (University of North Carolina) and Dr Jeff Rosen (Baylor College of Medicine) to further the analysis of ncRNA expression and function in breast development and breast cancer.

(b) We developed a database of ncRNA expression (Dinger et al., *Nucleic Acids Research*, 2009) and recruited a dedicated postdoctoral scientist (due to start in April 2009) to integrate and deposit data from all available sources (including in-house experiments) into a central data store, annotate them, via linkage to genomic coordinates, with predicted binding sites for transcription factors, miRNAs, ncRNAs and other novel regulators, and make them available for analysis using a range of approaches available within the Centre (Bailey, Chen, McLachlan, Moscato), and with advanced network inference and visualisation tools provided by the Queensland Facility for Advanced Bioinformatics.

(c) Using data sources including RNAdb (Mattick) and NONCODE (Liu et al.), we formalised and standardised the description of ncRNA secondary-structure features including bulges, stems and loops. We applied data-mining technologies to extract these features, and related them to annotated function. We developed filters specific to organisms, developmental states and diseases, and used these reduced data for further data-mining. From this we developed a causality matrix that relates a putative ncRNA to an identified ncRNA class or group. A manuscript was accepted for publication in *IEEE Transactions on Knowledge and Data Engineering* (Chen).

(d) We developed methods to predict the function of ncRNAs by similarity of expression pattern with ncRNAs of known function, characterised these ncRNAs based on the annotation of their transcripts, and cluster them into groups. A manuscript was submitted (Chen).
During 2008 we explored external funding possibilities for a project that would use V1 and V2 chips to generate novel expression signatures encompassing both protein-coding and non-protein-coding genomic regions in tissue samples from prostate cancer, melanoma and breast cancer (Mattick, Moscato). An agreement now exists to put this forward for 2010, while we try to collect more evidence and pilot data on the role of non-coding RNA in the diseases currently studied by Moscato at the Newcastle node of aCB. At present, thanks to a collaboration also involving Dr Marcel Dinger (UD) and Dr Carlos Riveros (CIBM-Newcastle), we are trying to establish a pilot study from existing Illumina datasets on the role of non-protein-coding genomic regions as markers of disease subtypes in prostate cancer, melanoma and breast cancer (Mattick, Moscato). The datasets have been obtained thanks to an existing collaboration with Prof. Peter Hersey (CIBM-Newcastle) or are in the public domain. We expect the analysis to be first priority by mid-2009, when the Newcastle node concludes a set of three manuscripts, currently in preparation, on protein-coding regions and their differential expression in melanoma metastasis.

In a separate project, Moscato and his group are developing a method to find a minimum set of putative transcription factors that can be linked to a set of differentially expressed genes, i.e. to identify putative subsets of gene-expression regulators, of minimal cardinality, that “explain” a major transcriptional change between groups of samples. The method is general and can be applied to regulation by miRNAs. The Bailey group will provide lists of predicted target genes for selected human transcription factors with known binding motifs; target predictions will use the comparative genomics-based methodology developed by Centre postdoctoral researcher John Hawkins (Project 2.1).
Program 3 – Modelling
dynamic cellular processes

3.1 Mathematical modelling of
genetic regulatory networks and
biochemical pathways

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

In collaboration with:
Prof. Manuel Barrio
(Valldolid University, Spain)
Prof. Simon Davis
(Oxford University, UK)
Prof. James Dehnam
(The University of Newcastle)
Dr Nicholas Hamilton
(The University of Queensland)
Prof. John Hancock
(University of Texas, Houston)
Prof. Des Higham
(University of Strathclyde, UK)
Dr Tatiana Marquez Lago
(ETH, Switzerland)
Dr André Leier
(ETH, Switzerland)
Dr Grant Lythe
(University of Leeds, UK)
Dr Brad Marsh
(The University of Queensland)
Dr Carmen Molina-Paris
(University of Leeds, UK)
A/Prof. Pablo Moscato
(The University of Newcastle)
Dr Tianhai Tian
(Monash University)
Prof. Jordi Villa
(University of Barcelona, Spain)

Researchers:
Research Associates:
Dr John Belward
(The University of Queensland)
Dr Pamela Burrage
(The University of Queensland)
Dr Fawang Liu
(The University of Queensland)
Dr Shev MacNamara
(The University of Queensland)
Dr Muhammad Shoai Seghal
(ARC Centre of Excellence in Bioinformatics)

Research staff:
Mr Tim Sullivan
(ARC Centre of Excellence in Bioinformatics)

PhD students:
Mr Alhadi Bustamam
(The University of Queensland)
Ms Chin Foon Khoo
(Australian National University)

Aims:
(a) We will explore quantitative
mathematical models of genetic
regulatory networks and kinase
signalling pathways using stochastic
differential equations and other
mathematical formalisms, and simulate
these models in computer code, in
some cases marked-up in MathML so
that they can be imported, via SMBl,
into the Visible Cell® environment
(Program 1).
(b) We will develop new stochastic models
for understanding the behaviour of
molecular clocks in mouse, and somite
formation in zebrafish, that incorporate
delayed effects that arise from delays in
transcription and translation.
(c) We will develop collaborations with
immunologists in the UK with the aim
of studying the immune system via
modelling of the interrelationships
among T cells, D cells and B cells.
Achievements (2008):

(a) We developed new stochastic simulation techniques for solving a wide class of important stochastic differential equation problems (Burrage, Lythe), and applied some of these simulation techniques to examine how the effects of intrinsic and extrinsic noise can accelerate and stabilise juxtracrine pattern formation (Burrage, Rudge).

(b) We developed multiscale methods for biochemical kinetics based on tau leaping and other approaches, with the aim of improving performance so that larger biochemical problems can be solved over longer time periods (Burrage, Villa).

(c) We developed delayed stochastic models for molecular clocks in mouse (*heate*1 gene) and somite formation in zebrafish (Burrage, Leier, Marquez Lago). These techniques are of a multiscale nature, so we can simulate much larger problems than was previously the case.

(d) We developed efficient and effective Krylov and steady-state techniques for solving the Chemical Master Equation for the mitogen-activated protein kinase cascade (Burrage, Hegland, MacNamara, Sijde). This is the first time that such large chemical-kinetic problems have become solvable by these techniques.

(e) We developed spatial and temporal models for investigating the human immune system, and the interrelationships among T cells, D cells and B cells (Burrage, MacNamara, Davis, Lythe, Molina-Paris).

(f) We developed new, computationally efficient techniques for studying the properties of large molecular-interaction networks (Seghal, Burrage, Bustamam).

(g) During 2008, Moscato and PhD student Mario Inostroza-Ponta developed a parallel, scalable algorithm for non-hierarchical clustering. Gene-expression datasets with relatively good annotation (e.g. yeast) were studied for validation of the technique in a real scenario with very good success. Later the method was used in cancer datasets to find deregulated sets of genes and their associated putative common regulators (transcription factors and miRNAs). The methods were discussed in Mario’s PhD thesis, which was examined and approved in late 2008. Several publications are likely to arise from these results.

(h) In a collaboration involving Prof. James Denham (Newcastle), Moscato and PhD student Mario Inostroza-Ponta analysed data from a prostate cancer trial. The purpose was to categorize longitudinal radiation-induced rectal toxicity data obtained from men participating in TROG 96.01, a randomised controlled trial for locally advanced prostate cancer. Our algorithm showed seven well-defined clusters of rectal symptoms, with 25 seen immediately following radiation and 7 at years 1, 2 and 3 following radiation. The conclusion is then that the widely used CTC grading does not accord with the degree of rectal “distress” and “problems” at all time points, and reveals serious shortcomings for clinical use. On the positive side, our results open opportunities for us to develop genomic profiling and detailed patho-physiological studies in prostate cancer management. A paper in *Radiotherapy & Oncology* was published electronically (24 October 2008) ahead of print publication in 2009 (Capp, Inostroza-Ponta et al.).
3.2 Modelling and visualising biological processes in complex spatial environments

Investigators:
- Prof. Kevin Burrage (The University of Queensland)
- Dr Markus Hegland (Australian National University)

In collaboration with:
- Dr Nicholas Hamilton (The University of Queensland)
- Prof. John Hancock (University of Texas, Houston)
- Dr Tatiana Marquez Lago (ETH, Switzerland)
- Dr André Leier (ETH, Switzerland)
- Dr Brad Marsh (The University of Queensland)
- Mr Dan Nicolau Jr (Oxford University)
- Dr Rohan Teasdale (The University of Queensland)
- A/Prof. Peter Thorn (The University of Queensland)

Researchers:

Research Associates:
- Dr John Belward (The University of Queensland)
- Dr Pamela Burrage (The University of Queensland)
- Dr Fawang Liu (The University of Queensland)
- Dr Shev MacNamara (The University of Queensland)
- Dr Muhammad Shoaib Seghal (ARC Centre of Excellence in Bioinformatics)

Research staff:
- Mr Tim Sullivan (ARC Centre of Excellence in Bioinformatics)

PhD students:
- Mr Alhadi Bustamam (The University of Queensland)
- Ms Chin Foon Koo (Australian National University)

Aims:
The core goal of this theme is to develop novel methodologies for geometric, stochastic and statistical modelling of cellular processes based, where possible, on live cell imaging data in order to deepen our understanding of the processes being observed, and extend the range of inference that can be made from such data.

Some biological processes map readily to familiar mathematical formalisms (e.g., genetic regulatory networks as graphs). It is often less obvious how to model dynamic processes that involve complex cellular structures and spatial relationships. Within the ARC Centre of Excellence in Bioinformatics we have established genuine collaborations of experimentalists and mathematicians that have yielded new modelling techniques, new mathematics for generating experimentally verifiable hypotheses, and (collaboratively with IMB's Virtual Membrane strategic program) models for the kinetics of lipid raft mobility, tubular-vesicle formation from endosomes, macropinosome biogenesis and vesicle recycling. These models are parameterised using live cell imaging, high-resolution confocal and EM data.

(a) We will analyse the large numbers of 3D movies of processes such as macropinocytosis and Golgi events that are currently being captured in the laboratory of Dr Rohan Teasdale, with the aim of getting deeper insights into the nature of the complex spatial and temporal dynamics of these cellular processes.

(b) We will extend this work to whole beta-cell tomograms at 5-nm resolution (Program 1) and to 4D live images, producing and applying novel protocols, automated image-analysis and visualisation tools, and voxel-based stochastic simulators. In this way, the ultrastructural matrix of our Visible Cell® will by 2010 become model-driven.

(c) We will extend our virtual membrane models so that simulations can be run on larger problems and over more-appropriate longer time scales.
Achievements (2008):

(a) In collaboration with Group Stow (IMB), a paper describing OBCOL, a system for segmenting and quantifying multi-fluorophore 3D and 4D cell imaging, was submitted to Cytometry A. OBCOL is unique in that it is the first system that can quantify whole-cell images to give detailed co-localisation information at the organelle level. In future work, the OBCOL quantification methods will form the basis for data generation for mathematical modelling. Supporting software is being released under an open source license (Hamilton).

(b) We developed methods that automatically quantify the recruitment and expulsion in live cells of phosphoinositides from endosomes, an important subcellular organelle implicated in a number of major diseases, and submitted a manuscript (Kerr et al.). These methods enable the first detailed view of the rates of recruitment and transformations of the phosphoinositides, as well as their relationships and timings. Currently these data are being used as a basis for mathematical modelling of the phosphoinositide transformations (Fawang Liu, John Belward). The systems have not previously been modelled, and preliminary differential equation models created show much promise in replicating the behaviours and timings found in the experimental data (Hamilton, Teasdale, Burrage).

(c) We developed new computational methods for modelling the transport and kinetics of processes in crowded spatial environments, particularly the plasma membrane. These new methods are based around the concept of fractional differential equations (Liu, Burrage), and attempt to capture complicated spatial effects by a deterministic fractional partial differential equation framework. This is a very general approach that could be important wherever there is molecular crowding. This general description of non-classical diffusion more easily interprets experimental diffusion studies than do existing approaches.

(d) We developed new parallel simulation techniques that allow us to perform very complex simulations of processes on the plasma membrane by decomposing the problem into a number of sub-problems that communicate asynchronously. This allows us to tackle much larger simulation problems than via a purely sequential simulation (P. Burrage).

(e) We developed new stochastic simulation techniques for chemical reactions in spatially complex media such as the plasma membrane. We extended the Stochastic Simulation Algorithm so it can be applied in a spatially constrained environment (Nicolau, Burrage). This approach will allow us to evaluate the potential of various interactions (notably barriers, lipid rafts, and molecular crowding) to explain experimental observations of protein confinement and anomalous diffusion on the membrane.

(f) We developed a novel optimisation algorithm called extremotaxis that was inspired by how some bacteria move, and applied it to several very difficult optimisation problems with impressive results (Nicolau, Maini, Burrage).

(g) We carried out computational simulations within the Visible Cell® (Burrage, Hegland, MacNamara, Sullivan, Sehgal) focusing on simulations of granular transport and insulin release at the plasma membrane, and development of temporal stochastic and deterministic models that explore the dynamics between glucose uptake and insulin release.
Program 4 – Algorithms for graphs and networks

4.1 Discrete optimisation and graph clustering

Investigator:
Prof. Michael Fellows
(The University of Newcastle)

In collaboration with:
Prof. Mike Langston
(University of Tennessee and ORNL)

Researcher:
Research Associate:
Dr Frances Rosamond
(The University of Newcastle)

Aims:
(a) Develop and implement algorithms for discrete clustering problems related to the VERTEX COVER problem, especially FPT algorithms that can be applied to very large datasets due to effective data-reduction subroutines.
(b) Integrate discrete clustering models with problems concerning motif identification for sequence-based datasets.
(c) Develop general algorithmic approaches to classes of discrete clustering problems, such as MINIMUM EDGE DELETION EDITING to a graph of maximum degree d (the VERTEX COVER problem is the special case d=0).
(d) Develop algorithms for problems on coloured graphs, such as the graph-based Biochemical Motif Identification problem (a relaxed form of coloured subgraph isomorphism), and problems on convex recolouring of trees applicable in phylogenetic modelling and analysis.
(e) Integrate novel algorithmic strategies for identifying key nodes in networks, with broader strategies for experimentally verifying interpretations of the data suggested by the algorithmic outputs.
(f) Develop algorithms for multiple-interval graph-based analysis of sequence data.
(g) Develop new algorithms for realistic graph-based modelling of haplotype data in the presence of errors.

Achievements (2008):
(a) We made some first applications of FPT algorithmics to sequence consensus problems; these have applications in voting systems, and internet search engines as well as in the study of gene sequences. Results were published at prominent conferences.
(b) We found linear kernelisation algorithms for several strong generalisations of the VERTEX COVER problem, including the problem of deleting k vertices so that the resulting graph has maximum degree at most d (VERTEX COVER is the case of d=0). Using the systematic technology of crown reduction (developed in recent years in Centre programs) we were able to achieve linear vertex kernels for every fixed d in quadratic time. These results were accepted at STACS 2009.
(c) We continued to investigate graph motif problems (modelled by coloured graphs) with several bioinformatics applications. Our results were published at CPM 2008, and a full version prepared and submitted to Algorithmica.
(d) Our 2007 results on multiple interval graph problems relevant to sequence analysis were prepared for, submitted and accepted to Theoretical Computer Science (published January 2009).
(e) A paper on FPT algorithms for analysing haplotype data in a realistic model where plausible haplotypes are constrained, was prepared and submitted to CPM 2009.
(f) New applications (to new datasets) of our discrete clustering algorithms were obtained by the implementations of our algorithms by PI Mike Langston.
4.2 Optimisation-based approaches to large systems

Aims:

(a) Develop a comprehensive methodology based on combinatorial optimization approaches for the identification of "oncosystems", networks of highly correlated gene expression patterns in cancer samples as measured from microarray and miRNA datasets.

(b) Address the open problems including the development of powerful semi-supervised methods that identify subgroups representing potentially under-sampled phenotypes.

(c) Develop more-robust pattern recognition methods that allow platform-independent recognition of molecular profiles characterising phenome subtypes, and will combine unsupervised and semi-supervised approaches to identify pathways that are differentially expressed according to disease state.

Achievements (2008):

(a) We developed a metaheuristics-based method for identification of large molecular signatures in microarray datasets. The method will be tested with real datasets for which an assumed solution is known (e.g. the NCI-60 panel), with simulated data, and with real datasets to which a supervised method has been already applied (e.g. the 2006 dataset on prostate cancer by Hood et al.). We will attempt to identify further datasets suitable for analyses to validate the method (Moscato, Berretta).

(b) We developed a novel methodology for identifying, from gene-expression data, biomarkers that correlate with disease progression. This method uses of information-theory-based quantifiers, Shannon entropies, and generalised statistical complexity quantifiers based on Jensen-Shannon divergence (Berretta, Rosso and Moscato). A large manuscript covering methodology, implementation and results for both prostate cancer and melanoma is in preparation. The method will greatly enrich the capabilities of this group in support of Projects 2.2 and 2.3.

(c) We verified our hypothesis that there exist large molecular correlates to sudden transitions in Shannon entropy and statistical complexity, and have been able to understand dysregulated biological processes, using Pathway Studio. We also investigated the relevance of the \((\mu_k)\times-k\)-FEATURE SET approach for biomarker selection. Using a dataset on protein abundance in blood, we showed that we have established the existence of a five-protein molecular signature that can predict the onset of clinical Alzheimer’s disease two to five years before it is declared; our results were published in \textit{PLoS ONE} (Gómez-Ravetti, Moscato).

(f) We tested new algorithmic ideas based on the Jensen-Shannon divergence and Shannon entropy to study gene-expression data. As a case study, we carried out a computational linguistic study involving a large corpus of English Renaissance literature: the 185 texts (136 plays and 49 poems between 1580-1640) include 30 texts unquestionably written by Shakespeare.
We also included A Lover’s Complaint, a poem which generally appears in Shakespeare collections but whose authorship is in dispute. Our statistical complexity quantifiers combine the power of Jensen-Shannon’s divergence with the entropy variations as computed from a probability distribution function of observed word-use frequencies. We found that for a given entropy, poems display higher complexity than plays, that Shakespeare’s work falls into two distinct entropy clusters, and that his work is remarkable for its homogeneity and for its closeness to overall means. A manuscript (Rosso, Craig, Moscato) will appear in Physica A (2009).

(g) We analysed background electroencephalography (EEG), recorded with scalp electrodes, in children with childhood absence epilepsy (CAE) and in control individuals. Functional activity between electrodes was evaluated using wavelet decomposition in conjunction with the Wootters distance, then pairs of electrodes with differentiated behaviour between CAE and controls were identified using a test statistic-based feature selection technique. This approach identified clear differences between CAE and healthy control background EEG in the frontocentral electrodes, as measured by Principal Component Analysis. The findings of this pilot study have strong implications for clinical practice. A paper (Rosso et al.) will appear in Journal Neuroscience Methods (2009), and a follow-up study has been initiated to explore the use of frailty–background EEG feature set to identify a smaller, yet more predictive, set of electrode pairs to discriminate cases from controls.

(h) We dealt with randomness-quantifiers and concentrate on their ability to discern chaos in time-series used in connection with pseudorandom number generators (PRNG). Workers in the field are motivated to use chaotic maps for generating PRNGs because they are simple to implement. Our analysis shed light on the importance of the two main statistical characteristics of a chaotic map, namely its invariant measure and the mixing constant. This helps to answer two questions that arise in applications: which is the best PRNG among those available, and if a given PRNG turns out not to be good enough and a randomisation procedure must still be applied to it, which is randomisation procedure is best? A paper (De Micco et al.) will appear in Philosophical Transactions of The Royal Society A (2009).

(i) We evaluated multifractality degree (MD) in a collection of developed and emerging stock market indices. Empirical results suggest that MD can serve as a quantifier to characterise the stage of market development. We develop a model to test the relationship between stage of market development and MD, and found robust evidence that the relationship is negative, i.e. MD is associated with a less-developed market. Thus, an inefficiency ranking can be derived from multifractal analysis. We also established a link with previous volatility time series. A paper (Zunino et al.) appeared in Physica A.

(j) Two important stochastic processes, the fractional Brownian motion and fractional Gaussian noise, were analysed within the framework of the Tsallis permutation entropy. This entropic measure, evaluated after using the Bandt & Pompe method to extract the associated probability distribution, was shown to be a powerful tool to characterise fractal stochastic processes. It allows for a better discrimination of the processes than the Shannon counterpart for appropriate ranges of values of the entropic index. Moreover, we find the optimum value of this entropic index for the stochastic processes under study. A paper (Zunino et al.) appeared in Physica A (2008).

(k) We worked out theoretical curves for the permutation entropy of the fractional Brownian motion and fractional Gaussian noise by using the Bandt & Shiu theoretical predictions for their corresponding relative frequencies. Comparisons with numerical simulations show excellent agreement. We could also theoretically validate the entropy gap in the transition between these processes, observed previously via numerical results. We analysed behaviour of the permutation entropy of the fractional Gaussian noise for different time delays. A paper (Zunino et al.) appeared in Physics Letter A (2008).

(l) Even in healthy individuals, erythrocytes in capillaries continually change their shape in a synchronised erratic fashion. We studied this cell behaviour from the perspective of a bounded correlated random walk, assuming that diffraction data involve both deterministic and stochastic components. We obtained photometric data ektacytometry over several million shear-elongated cells, using a home-made device called an erythrodeformeter. As we have only a scalar signal and no governing equations, we reconstructed behaviour in an artificial phase space. The results suggest that on a random-walk approach, samples from healthy controls exhibit significant differences from those from diabetic patients; thus we can claim to have linked nonlinear mathematical tools with clinically relevant rheological properties of erythrocytes from diabetics. A paper (Korol et al.) appeared in Open Medical Informatics Journal (2008).

(m) Behaviour from the perspective of a walk approach, samples from healthy controls exhibit significant differences from those from diabetic patients; thus we can claim to have linked nonlinear mathematical tools with clinically relevant rheological properties of erythrocytes from diabetics. A paper (Korol et al.) appeared in Open Medical Informatics Journal (2008).
### 4.3 External-memory algorithms for analysis of massive graphs

**Aims:**

Since 2003, the meaning of large graph has changed by at least an order of magnitude. The gene co-expression graphs analysed by Mike Langston and Pablo Moscato typically have 25K-200K nodes and ~1M edges, while problems in protein-protein interaction graphs and MACHOS (Wong & Ragan) can be 1-2 orders of magnitude larger yet.

At least initially, this project will constitute a watching brief on technologies potentially useful for the analysis of graphs of this order. We will study the feasibility of translating to bioinformatics new algorithmic advances in advanced telecommunications, including external memory algorithms and novel data structures now being implemented by AT&T for clique and quasi-clique computations in multi-digraphs of ~200M edges, and biclique partitioning in massive weighted bipartite graphs as potentially applied to multiple microarray datasets.

**Achievements (2008):**

Although this project constitutes as a watching brief and no specific funds have been associated, it has already produced some interesting results thanks to a collaboration between Moscato and other members of the Newcastle ACB node, with Drs Luciana Buriol and Matt Ritt at Universidade Federal de Rio Grande do Sul, Brazil, where MSc student Leomar Costa has implemented an external memory algorithm version of the clustering method developed by Moscato and Inostroza-Ponta (2005-2008). This algorithm uses STXXL (Standard Template Library for Extra Large Data Sets), a free, open-source program available under the Boost Software License 1.0. We are now evaluating the feasibility of using this algorithm for problems involving genome-wide association studies and microarray data analysis. The Newcastle node is likely to host some of our Brazilian collaborators in 2009, to explore the use of the algorithm in some datasets in melanoma and multiple sclerosis.

---

**Investigators:**

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

**In collaboration with:**

Dr Tim Bailey  
(The University of Queensland)

Prof. Kevin Burrage  
(The University of Queensland)

A/Prof. Phoebe Chen  
(Deakin University)

Prof. Mike Langston  
(University of Tennessee and ORNL)

Prof. Mark Ragan  
(The University of Queensland)

Dr Isidore Rigoutsos  
(IBM TJ Watson Research Center)

Prof. Xiaofang Zhou  
(The University of Queensland)

Dr Luciana Buriol  
(University. Fed. de Rio grande, Brazil)

Dr Matt Ritt  
(Univ. Fed. de Rio Grande do Sul, Brazil)

**Researchers:**

**PhD student:**  
Mr Mario Inostroza-Ponta  
(The University of Newcastle)

**MSc student:**  
Mr Leomar Costa  
(Univ. Fed. de Rio Grande do Sul, Brazil)
19th International Conference on Genome Informatics (GIW-2008)

Gold Coast (1–3 December 2008)

ACB was local host and a sponsor of the 19th International Conference on Genome Informatics (GIW-2008), held on Gold Coast from 1–3 December 2008. GIW, held annually under the auspices of the Asian Association of Societies of Bioinformatics (AASBi), is the world’s oldest continuing conference series in bioinformatics. Until GIW-2006, annual GIW conferences were held in Tokyo, then later Yokohama. At that point AASBi decided that GIW should circulate through the AASBi countries. GIW-2007 was held at the Biopolis in Singapore, and GIW-2008 was hosted in Australia under sponsorship of Bioinformatics Australia, our national bioinformatics society organised within AusBiotech.

Mark Ragan served as conference Chair, while ACB Chief Investigators Timothy Bailey, Kevin Burrage, Phoebe Chen, Geoff McLachlan, John Mattick, Pablo Moscato, Shoba Ranganathan, Rohan Teasdale and Xiaofang Zhou were members of the National Organisation Committee. The conference was opened by Professor Peter Andrews AO, Queensland Chief Scientist.

The Australian bioinformatics community, companies and institutions provided brilliant support for GIW-2008. Financial sponsorship and support was secured from the Department of Innovation, Industry, Science and Research (DIISR), Queensland Government, The University of Queensland, CSIRO, Alexander Technology, SGI, the Computational Biology Research Center of the National Institute of Advanced Industrial Science and Technology (Japan), ARC Research Network in Enterprise Information Infrastructure, Australian Centre for Plant Functional Genomics, Australian Genome Research Facility Ltd, National ICT Australia, Queensland Facility for Advanced Bioinformatics, The Queensland Cyber Infrastructure Foundation (QCIF), Queensland University of Technology, Sydney Bioinformatics, AARNet, and Cisco Systems. The institutional and commercial displays added a further dimension that vitalised the conference and added great value for all participants. A portable tiled-wall OptIPortal operated by AARNet and QCIF was particularly popular.

GIW-2008 featured six leading keynote speakers. International keynote speakers were Julio Collado-Vides (UNAM Universidad Nacional Autónoma de México), Ming Li (Canada Research Chair in Bioinformatics, University of Waterloo), Yi-xue Li (Director, Shanghai Center for Bioinformation Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences; and Shanghai Jiaotong University) and Cliona Molony (Rosetta Inpharmatics/Merck Research Labs, Seattle, USA). Rounding out the roster of distinguished keynote speakers were Sean Grimmond (Institute for Molecular Bioscience, The University of Queensland – a CI in the first phase of ACB) and CI John Mattick.

Open round-table discussion forums were held on the topics Innovation and commercialisation in genome-scale informatics and Development of genome-scale bioinformatics in China and India. Please see pages 70 to 72 for summaries of these sessions.

GIW-2008 attracted 220 registrants, almost half from overseas (Canada, China, France, India, Japan, Korea, Mexico, Malaysia, New Zealand, Russia, Singapore, Taiwan, USA). Within Australia, all mainland states (plus ACT), almost all major bioinformatics groups, and many genomics and other “omics” groups were represented.

Prof. Allen Rodrigo, University of Auckland, chairing a session at GIW-2008.
Key Performance Indicators

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 2008:

The ARC Centre of Excellence in Bioinformatics again organised and sponsored our annual national Winter School in Mathematical and Computational Biology. The 2008 Winter School was hosted by our node at The University of Newcastle, the first time Winter School has been held elsewhere than at The University of Queensland. Winter School is designed to introduce topics in mathematical and computational biology to 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 fifth Winter School attracted 82 participants from 35 institutions in six countries, including PI Michael Langston.

2008 was a good year for publications from the Centre, with 102 C1 journal articles (59 of which were co-authored with overseas collaborators) and 37 E1 conference papers.

Prof. Michael Fellows continued his longstanding activity in science communication and popularisation. His books Computer Science Unplugged! (written with Prof. Tim Bell, University of Canterbury and Prof. Ian Witten, University of Waikato) and This is MEGA-Mathematics! (with Nancy Casey) convey sophisticated concepts such as intractability, sorting networks, and cryptography. They have won several international awards for science popularisation, and have been translated into Chinese, French, Japanese, Korean, Norwegian, Polish, Russian, Spanish, Swedish and Urdu, with more translations underway. Professor Christopher M. Bishop (UK Microsoft Research) used material from Computer Science Unplugged! in his 2008 Faraday Christmas Lectures; this lecture series, a tradition in London since 1825, is sponsored by the Royal Institution of Great Britain and the Leverhulme Trust.

During 2007-2008 Prof. Fellows and Dr Frances Rosamond were based largely in Jena (Germany) with support from his von Humboldt Research Award awarded in 2007. Mike and Fran and visited a number of research centres across Europe during 2008, taking time out from research collaboration to lead a number of workshops for schoolchildren or teachers.

Prof. John Mattick was elected a Fellow of the Australian Academy of Science. John was appointed as inaugural Gutenberg Chair at the Université Louis Pasteur de Strasbourg, and was based in Strasbourg for a number of weeks.

Prof. Geoff McLachlan was elected President-Elect of the International Federation of Classification Societies.

Prof. Kevin Burrage took up a joint appointment at Oxford University as Professor of Computational Systems Biology, and was based at Oxford for several months during 2008. Kevin took the lead in developing plans for student exchange between the Institute for Molecular Bioscience and the systems biology program at Oxford. The first student exchanges could take place during 2009.

In recognition of the excellence of his research carried out within ACB, Dr Nick Hamilton was appointed a Group Leader at the Institute for Molecular Bioscience.


Dr Melissa Davis received a Dean’s Award for Excellence for her PhD thesis Defining the membrane organisation of eukaryotic proteins. Her thesis research was advised by Dr Rohan Teasdale.

Dr Stefan Maetschke received a Dean’s Award for Excellence for his PhD thesis (submitted in late 2007) Topological models of transmembrane proteins for subcellular localization prediction. Prof. Geoff McLachlan and Dr Mikael Bodén were his Associate Advisors.

Dr Timothy Bailey’s students Denis Bauer and Fabian Buske won the best student paper award for their paper “Predicting SUMOylation Sites” at the Third International Association for Pattern Recognition (IAPR) Conference on Pattern Recognition in Bioinformatics (PRIB 2008) held in Melbourne.

2008 ANNUAL REPORT
ARC CENTRE OF EXCELLENCE IN BIOINFORMATICS
1. Research findings and competitiveness

Publications

Quality of publications

ACB researchers continue to perform well in peer-reviewed publication. Our Investigators and their research groups were authors of 102 Category C1 journal articles in 2008, of which 54 (53%) appeared in journals with 2007 impact factor ≥3.0, 30 (29%) with IF ≥5.0, and 8 (8%) with IF ≥10.0. In all, 59 (58%) of our C1 journal articles were co-authored with overseas collaborators.


Number of publications

The table on this page summarises the published output by our investigators and research groups that appeared with a 2008 publication date. We do not count papers that at the end of December 2008 were available only electronically in advance of print publication, or were in press, under review, submitted or in preparation. For details, please refer to Appendix 1: Publications and Papers.

Patents and trademarks

No patents were awarded to our investigators during 2008.

Centre students Joo-Young Choi and Elizabeth Skipington attended the BioBusiness Day Out held by IMBcom Pty Ltd on 13 November 2008. The introductory workshop provides first-year PhD students with an overview of the concepts of bio-business.

**Key Performance Indicators**

<table>
<thead>
<tr>
<th>Category</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 Book – authored research</td>
<td>1</td>
</tr>
<tr>
<td>A3 Book – edited</td>
<td>2</td>
</tr>
<tr>
<td>B Book chapter</td>
<td>14</td>
</tr>
<tr>
<td>C1 Journal articles – articles in scholarly refereed journal</td>
<td>102</td>
</tr>
<tr>
<td>C2 Journal article – other contribution to refereed journal</td>
<td>2</td>
</tr>
<tr>
<td>C3 Journal article – non-refereed article</td>
<td>1</td>
</tr>
<tr>
<td>C5 Unpublished reports (including commercial consultancies)</td>
<td>3</td>
</tr>
<tr>
<td>E1 Conference – full written paper – refereed proceedings</td>
<td>37</td>
</tr>
<tr>
<td>E2 Conference – full written paper – non-refereed proceedings</td>
<td>2</td>
</tr>
<tr>
<td>E4 Conference – edited volume of conference proceedings</td>
<td>2</td>
</tr>
<tr>
<td>E4 Conference – unpublished presentations</td>
<td>8</td>
</tr>
<tr>
<td>G Computer software</td>
<td>3</td>
</tr>
<tr>
<td>Other academic outputs</td>
<td></td>
</tr>
<tr>
<td>– Demos</td>
<td>1</td>
</tr>
<tr>
<td>– Newsletter</td>
<td>1</td>
</tr>
<tr>
<td>– Posters</td>
<td>29</td>
</tr>
<tr>
<td>– Postgraduate theses</td>
<td>9</td>
</tr>
<tr>
<td>TOTAL</td>
<td>217</td>
</tr>
</tbody>
</table>
Invitations to address and participate in conferences

International conferences

Chief and Partner Investigators
Dr Timothy Bailey
- Program Committee, 3rd International Association for Pattern Recognition (IAPR) Conference sponsored conference on Pattern Recognition in Bioinformatics (PRIB 2008), Melbourne (October)
- Organising Committee, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)

Prof. Kevin Burrage
- International Conference on Computational Cell Biology, Hinxton, UK (March)
- Second Bertinoro Systems Biology Workshop (BSB 2008), Bertinoro, Italy (April)
- Lymphocyte Kinetics in Health and Disease, London, UK (May)
- eResearch Australasia 2008, Melbourne (September)
- Program Committee, Sixth Conference on Computational Methods in Systems Biology, Rostock, Germany (October)

A/Prof. Phoebe Chen
- Chair, Steering Committee, Sixth Asia-Pacific Bioinformatics Conference (APBC2008), Kyoto (January)
- Chair, Steering Committee, 14th International Conference on Multimedia Modelling (MMM2008), Kyoto (January)
- Program Committee, 24th International Conference on Data Engineering (ICDE2008), Cancun, Mexico (April)
- Program Committee, 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2008), Osaka (May)
- Program Committee, Second International Conference on Bioinformatics Research and Development (BIRD 2008), Vienna (July)
- IASTED International Conference on Internet and Multimedia Systems and Applications (IMSA 2008), Kailua-Kona, Hawaii, USA (August)
- Keynote, 21st Republic of China Image Processing and Pattern Recognition Society Conference on Computer Vision, Graphics and Image Processing, Taiwan (August)
- Publicity Co-Chair, 3rd International Association for Pattern Recognition (IAPR) Conference on Pattern Recognition in Bioinformatics (PRIB 2008), Melbourne (October)
- Organising Committee, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)
- Program Committee, 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, Florida (December)
- Program Committee, 10th Pacific Rim International Conference on Artificial Intelligence (PRICAI2008), Hanoi (December)

Prof. Michael Fellows
- Technical Symposium on Computer Science Education (SIGCSE 2008), Portland, Oregon, USA (March)
- Fourth International Conference on Algorithmic Aspects in Information and Management (AAIM 2008), Shanghai (June)
- International Conference on Discrete Mathematics (ICDM 2008), Mysore, India (July)
- Program Committee, 2nd Annual International Conference on Combinatorial Optimization and Applications (COCO2008), Newfoundland, Canada (August)
- Program Committee, 9th International Conference for Young Computer Scientists (ICYCS 2008), Hunan, China (November)

Dr Markus Hegland
- Keynote, Chemnitz Symposium on Inverse Problems, Chemnitz, Germany (September)
- Session Organiser, 7th Australia-New Zealand Mathematics Convention, Christchurch, New Zealand (December)

Prof. Geoff McLachlan
- 18th International Conference on Computational Statistics (Compstat 2008), Porto, Portugal (August)
- German Classification Society 32nd Annual Conference (GfKl-2008), Hamburg (July)
- XXIV International Biometric Conference, Dublin, Ireland (July)
- Program Committee, 3rd International Association for Pattern Recognition (IAPR) Conference, Melbourne (October)
- Organising Committee, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)

Dr Brad Marsh
- Keynote, Ninth International Symposium on Parallel Architectures, Algorithms, and Networks (ISSPAN’08), Sydney (May)
- American Diabetes Association Annual Meeting 2008, San Francisco (June)
- International Congress of Histochemistry & Cytochemistry (ICH2008), Gdansk, Poland (August)
- The 2008 Golgi Meeting: Membrane Trafficking in Global Cellular Responses, Pavia, Italy (September)
- Session Chair, 9th Asia-Pacific Microscopy Conference (APM09), Jeju Island, Korea (November)
- Plenary, Dutch Society for Microscopy Annual Meeting, Lunteren, The Netherlands (November)

Prof. John Mattick
- Plenary, Foundation des Treilles Meeting “New catalytic and regulatory functions of RNA in eukaryotes”, Nice, France (March)
- Plenary, Genomes to Systems Conference, Manchester, UK (March)
- Opening Plenary, Cell Culture Engineering XI, Cooilam (April)
- Invited speaker, 2008 ASMB (American Society for Biochemistry and Molecular Biology) Annual Meeting, San Diego, USA (April)
- Invited speaker, Queenstown Epigenetics Meeting, Queenstown, New Zealand (June)
• Invited speaker, EMBO New Members Meeting, Tampere, Finland (September)
• Plenary, European BioPerspectives, Hannover (October)
• Keynote, The RIGHT Symposium: RNA Interference Technology as Human Therapeutic Tool, Brussels, Belgium (November)
• Opening Keynote, 8th Annual Australasian Microarray and Associated Technologies (AMATA) Conference, Dunedin, New Zealand (November)
• Plenary and National Organising Committee, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)

A/Prof. Pablo Moscato
• Program Committee, Sixth European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (EVOBIO 2008), Napoli, Italy (March)
• Program Committee, Eighth European Conference on Evolutionary Computation in Combinational Optimisation (EvoCOP 2008), Napoli, Italy (March)
• Scientific Program Committee, Special Session on Memetic Algorithms, 2008 IEEE World Congress on Computational Intelligence, Hong Kong (June)
• Program Committee, Congress on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Hong Kong (June)
• Program Committee, Brazilian Symposium on Bioinformatics (BSB 2008), Santo André, Brazil (August)

Prof. Mark Ragan
• Program Committee, Sixth Asia-Pacific Bioinformatics Conference (APBC2008), Kyoto (January)
• Second Bertinoro Systems Biology Workshop (BSB 2008), Bertinoro, Italy (April)
• Australia–Brazil Workshop on Biotechnology Innovation for Agriculture, Brasília (May)
• 16th International Conference for Intelligent Systems for Molecular Biology (ISMB2008), Toronto (July)
• Program Committee, eResearch Australasia 2008, Melbourne (September)

• Chair, Organising Committee, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)

Prof. Shoba Ranganathan
• Program Committee, 8th Asia-Pacific Bioinformatics Conference (APBC 2008), Kyoto, Japan (January)
• Program Committee, 2nd International Conference on Bioinformatics Research and Development (BIRD) 2008, Vienna, Austria (July)
• Chair, Program Committee, 7th International Conference on Bioinformatics (InCoB 2008), Taipei, Taiwan (October)
• Program Committee, IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2008, Philadelphia, USA (November)
• Program Committee and Workshop Chair, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast, Australia (December)

Dr Rohan Teadale
• Cellular Microbiology – EMBO Meeting, Villars-sur-Ollon, Switzerland (September)
• 6th International Drug Discovery Science and Technology (IDDST) 2008, Beijing (October)

Prof. Xiaofang Zhou
• Program Committee, 17th International World Wide Web Conference (WWW2008), Beijing (April)
• Program Committee, 28th ACM SIGMOD International Conference on Management of Data (SIGMOD 2008), Vancouver (June)
• General Co-Chair, Fourth International Conference on Advanced Data Mining and Applications (ADMA 2008), Chengdu, China (July)
• Program Committee and Tutorial Chair, 34th International Conference on Very Large Databases (VLDB 2008), Auckland (August)
• Keynote, 9th International Conference on Web Information Systems Engineering (WISE 2008), Auckland (September)

Dr Adrian Cootes
• Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Postdoctoral fellows, Research Officers and students
Mr Paulo Amaral
• 29th Lorne Genome Conference, Lorne, Australia (July)
• XX International Congress of Genetics, Berlin, Germany (July)

Dr Mikael Bodén
• Asia-Pacific Bioinformatics Conference (APBC 2008), Kyoto (January)
• Program Committee, 3rd International Association for Pattern Recognition (IAPR) Conference on Pattern Recognition in Bioinformatics (PRIB 2008), Melbourne (October)
• Program committee, 9th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2008), Daejeon, Korea (November)

Dr Pamela Burrage
• Second Bertinoro Systems Biology Workshop (BSB 2008), Bertinoro, Italy (April)
• eResearch Australasia 2008, Melbourne (September)

Ms Elsa Chacko
• 16th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) 2008, Toronto, Canada (July)

Mr. Khar Heng Choo
• Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Dr Adrian Cootes
• 7th European Conference on Computational Biology (ECCB) 2008, Cagliari, Italy (September)
• Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)
Dr Aaron Darling
- Statistical Genetics for Metagenomics Workshop, Berlin (July)
- International Congress of Genetics, Berlin (July)
- 16th Arrowhead Meeting on Microbial Genomics, Lake Arrowhead, California (September)
- 6th RECOMB Comparative Genomics Satellite Meeting, Paris (October)
- Program Committee, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)

Dr Melissa Davis
- Asia-Pacific Bioinformatics Conference (APBC 2008), Kyoto (January)

Dr Mercer Dinger
- Keystone Symposium: RNAi, MicroRNA, and Non-Coding RNA, Whistler, Canada (March)

Mr. Jitendra Gaikwad
- Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Dr Karin Kassahn
- Society for Molecular Biology and Evolution 2008, Barcelona (June)

Mr. Javed Khan
- 12th Annual International Research in Computational Molecular Biology (RECOMB) Conference 2008, Singapore (March/April)
- Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Mr. Varun Khanna
- Strasbourg Summer School on Chemoinformatics 2008, Obernai, France (June)
- Program Committee Member, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)
- International Conference on New Developments in Drug Discovery from Natural Products and Traditional Medicines 2008, Mohali, India (November)

Mr. Gaurav Kumar
- Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Dr Krzysztof Kurowski
- eResearch Australasia 2008, Conference, Melbourne (September)

Dr Shev MacNamara
- Society for Mathematical Biology Conference, Toronto, Canada (August)
- SIAM Conference on the Life Sciences, Montreal, Canada (August)

Ms Ranjeeta Menon
- Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Dr Timothy Mercer
- Queenstown Epigenetics Meeting, Queenstown, New Zealand (June)
- Queenstown Molecular Biology Meeting, Queenstown, New Zealand (September)

Mr. Shivashankar Nagaraj
- Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Mr Satu Nahkuri
- First International EURASNET Conference on Alternative Splicing, Krakow, Poland (May)

Mr Andrew Noske
- Systems Biology Symposium, Queenstown, New Zealand (September)

Dr Osvaldo Rosso
- II International Conference Net-Works 2008, Pamplona, Spain (June)
- Organising Committee, XVI Conference on Nonequilibrium Statistical Mechanics and Nonlinear Physics (MEDYFIND'08), Punta del Este, Uruguay (December)

Mr. Lawrence Wee
- Program Committee, 7th International Conference on Bioinformatics, Taipei, Taiwan (October)

Dr Simon Wong
- 16th International Conference on Intelligent Systems for Molecular Biology (ISMB2008), Toronto (July)

Centre staff (Mr Oliver Cairncross, Mr David Wood), postdoctoral fellows (Dr Melissa Davis, Dr Nicholas Hamilton, Dr Karin Kassahn, Dr Shev MacNamara, Dr Muhammad Shoaib Sehgal) and students (Gaurav Kumar, Elizabeth Skippington, Joo-Young Choi, Isye Arieshanti, Alhadi Bustamam) participated in the 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December).
National conferences

Chief Investigators

Prof. Kevin Burrage
• 2008 Winter School in Mathematical and Computational Biology, Newcastle (July)

A/Prof. Phoebe Chen
• First Australia Video Conference (AusVideo), Brisbane (October 2008)

Dr Markus Hegland
• Organiser of local two-day meeting as part of the “Gene Golub around the World Day”, Canberra (February)
• Treasurer, Computational Techniques and Applications Conference (CTAC 2008), Canberra (July)

Prof. Geoff McLachlan
• Joint organiser of Program on Bioinformatics and Statistical Genetics, Australian Statistical Conference, Melbourne (July)
• Symposium Celebrating Prof. Sue Wilson’s 34 Years, Canberra (December)

Dr Brad Marsh
• Session Chair, Australian Conference on Microscopy & Microanalysis, Perth (February)

Prof. John Mattick
• Plenary, Australian Genome Research Facility Tenth Anniversary Genomics Symposium, Melbourne (April)
• Opening Plenary, The 18th St Vincent’s & Mater Health Research Symposium, Sydney (September)

A/Prof. Pablo Moscato
• 2008 Winter School in Mathematical and Computational Biology, Newcastle (July)

Prof. Mark Ragan
• 2008 Winter School in Mathematical and Computational Biology, Newcastle (July)

Prof. Shoba Ranganathan
• 2008 Winter School in Mathematical and Computational Biology, Newcastle (July)

Prof. Xiaofang Zhou
• 2008 EII Winter School in Data Quality and Management, Kioloa, NSW (August)
• 200 EII Early Career Development Workshop, Kioloa, NSW (August)

Postdoctoral fellows, Research Officers and students

Mr. Jitendra Gaikwad
• Biodiversity Information Standards (TDWG) annual Conference 2008, Perth (October)

Dr Nick Hamilton
• The 8th Hunter Meeting, Hunter Valley, NSW (April)

Mr Andrew Noske
• Session Chair, Mitochondrial Imaging and Dynamics, “AussieMit” workshop, Melbourne (November)

Dr Cheong Xin Chan and Dr Aaron Darling attended Adelaide Conference on Mathematical Evolutionary Biology 2008, Adelaide (March)

Centre postdoctoral fellows Dr Osvaldo Rosso and Dr Martin Gómez-Ravetti, and students Vinh Dang, Nikhil Lilaria and Elizabeth Skippington attended the 2008 Winter School in Mathematical and Computational Biology, Newcastle (July)

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

Chief Investigators

Dr Timothy Bailey
• Noble Laboratory, University of Washington, USA (November)

Prof. Kevin Burrage
• Computer Science Department, University of Rostock, Germany (November)
• Mathematics Department, University of Nottingham, UK (November)
• Mathematics Department, University of Leeds, UK (November)
• Systems Biology Centre, University of Edinburgh, UK (December)
• Biomedical Science Park, Barcelona, Spain (December)

A/Prof. Phoebe Chen
• Department of Computer Science, Fudan University, China (March)
• National Tsing Hua University, Taiwan (November)
• Academia Sinica Taiwan (November)
• National Chiao Tung University, Taiwan (November)
• National Taiwan Ocean University, Taiwan (November)
• National Chung Hsing University, Taiwan (November)

Prof. Michael Fellows
• As Humboldt Research Award invited guest lecturer: Universität Berlin, Universität Düsseldorf, Universität Jena, Universität Karlsruhe, Universität Kiel, Universität Lubeck, Universität Rostock, Universität Saarbrücken, and Max-Planck Institut für Informatik (Saarbrücken), Germany
• Prof. Henning Fernau, Universität Trier, Germany (March)
• Centre International de Rencontres Mathématiques, Marseille (April)
• Humboldt-Universität zu Berlin, Germany (April)
• Prof. Michael Langston, University of Tennessee (December)
Dr Markus Hegland
- Institute for Numerical Simulation, University of Bonn, Germany (August, September)
- Department of Information Technology, University of Uppsala, Sweden (September)

Prof. Geoff McLachlan
- Invited fellow in the programme Statistical Theory and Methods for Complex, High-Dimensional Data held at the Isaac Newton Institute for Mathematical Sciences, Cambridge University (March-June)
- University of Warwick, UK (May)
- University of Reading, UK (June)

Dr Brad Marsh
- Walter & Eliza Hall Institute of Medical Research, Melbourne (June)
- Systems Biology Laboratory, Hagedorn Research Institute, Novo Nordisk, Denmark (September)

Prof. John Mattick
- As Leonardo Lecturer, Istituto Scientifico San Raffaele and Università Vita-Salute San Raffaele, Milan, Italy (March)
- Cambridge Biological Society, Cambridge UK (March)
- As Gutenberg Professor, University of Strasbourg, France (June)
- Friedrich Miescher Institute (FMI), Basel Switzerland (June)
- Institut de Génétique et de Biologie Moléculaire et Cellulaire (IBMC), Strasbourg, France (June)
- Institut de Biologie Moléculaire et Cellulaire (IBMC), Centre National de La Recherche Scientifique, Strasbourg France (June)
- Université Pierre et Marie Curie, Paris, France (June)
- Max Planck Institute for Molecular Genetics, Berlin, Germany (June)
- Department of Biology, University of Copenhagen, Denmark (September)
- Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany (October)
- Institut de Génétique et de Biologie Moléculaire et Cellulaire (IBMC), Strasbourg, France (October)

A/Prof Pablo Moscato
- Instituto de Ciencias Matematicas e de Computacao, Universidade de Sao Paulo, Sao Carlos, Sao Paulo, Brazil (September)
- Departamento de Bioquimica, Universidad Nacional de Quilmes, Buenos Aires, Argentina (September)
- Departamento de Computacion, Universidad de Buenos Aires, Buenos Aires, Argentina (September)

Prof. Mark Ragan
- Murzin group (Structural Classification of Proteins database), MRC Laboratory of Molecular Biology and Cambridge Centre for Protein Engineering, Cambridge UK (January)
- European Bioinformatics Institute, Hinxton, UK (January)
- EMBRAPA (Brazilian Agricultural Research Corporation) headquarters, laboratories and field station, Brasilia, Brazil (May)
- Ministry of Science, Technology and Innovation, Shah Alam, Malaysia (August)

Prof. Shoba Ranganathan
- Bioinformatics Institute, Singapore (September)
- Universiti Teknologi Malaysia (September)
- Sanger Centre, Hinxton, UK (September)
- European Bioinformatics Institute, Hinxton, UK (September)
- Academia Sinica, Taiwan (October)

Prof. Xiaofang Zhou
- Nagoya University, Japan (November 2008–January 2009)
- Institute of Software, Chinese Academy of Sciences, Beijing, China (January)
- Aalborg University, Denmark (March)
- Chinese Academy of Science Institute of Software, Beijing, China (April)
- Peking University, Beijing, China (April)
- Renmin University, Beijing, China (April)
- IBM Research, Beijing, China (April)
- GIS Company, Wuhan, China (April)
- Wuhan University, Wuhan, China (April)
- Huazhong University of Science and Technology, Wuhan, China (April)

Postdoctoral fellows, Research Officers and students
Mr Javed Khan
- Institute for Infocomm Research, Singapore (April)

Dr Osvaldo A Rosso
- Departamento de Fisica i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Terrassa, Barcelona, Spain (June)
- Departamento de Fisica y Matemática Aplicada, Universidad de Navarra, Spain (June)
- Departamento de Informática e Engenheria de Sistemas, Universidad de Zaragoza, Zaragoza, Spain (June)
- Instituto del Desarrollo Humano, Universidad Nacional de General Sarmiento, Buenos Aires, Argentina (December)
- Escuela de Ciencia y Tecnología, Universidad Nacional de San Martín, Buenos Aires, Argentina (December)
- Laboratorio de Sistemas Complejos, Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina (December)
- Servicio de Neurología Pediátrica, Hospital Italiano, Buenos Aires, Argentina (December)
- Instituto de Cálculo, Universidad de Buenos Aires, Argentina (December)
- Centro Investigaciones Ópticas, Universidad Nacional de La Plata, La Plata, Argentina (December)
- Instituto de Fisica, Universidad Nacional de La Plata, La Plata, Argentina (December)

Dr Martin Gómez-Ravetti
- Hong Kong University of Science and Technology, Hong Kong (April)
- Federal University of Minas Gerais, Brazil (May)

Dr Christophe Lefèvre, Prof. Mark Ragan and A/Prof. Phoebe Chen (front row from left) with Deakin Bioinformatics Group members.

2008 ANNUAL REPORT
ARC CENTRE OF EXCELLENCE IN BIOINFORMATICS

Key Performance Indicators
Commentaries about the Centre’s achievements

Professor John Mattick was one of seventeen of Australia’s leading scientists honoured on 19 March by election to the Australian Academy of Science. John’s nomination cites his work on the structure of genetic systems in higher organisms. Election to the Academy recognises a career that has significantly advanced, and continues to advance, the world’s scientific knowledge.

John was also quoted in two articles in the New York Times, “The promise and power of RNA” and “Now: the rest of the genome” (11 November 2008). The same issue depicted John in a cartoon accompanied by the quotation “This [the many new roles found for RNA] is potentially the biggest change in our understanding of biology since the discovery of the double helix”.

Dr Aaron Darling appeared on the Channel 7 news program “Catalyst” regarding the genetic basis of antibiotic resistance in bacteria (Wednesday, 13 February 2008).

Dr Brad Marsh’s work on high-resolution cell tomography of mammalian cells was featured in the leading textbook Physical Biology of the Cell (Phillips, Kondev & Theriot, Eds) and in the 150th anniversary edition of Gray’s Anatomy. His work was also featured in a Queensland Weekender feature article on diabetes management and research (30 October 2008).


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 11 projects and sub-projects active during 2008, 8 (73%) included personnel with advanced degrees in different disciplines; 20 of our 28 PDFs (71%), 13 of our 55 postgraduate students (24%), and 4 international interns worked in these 11 projects. Of our 57 papers and posters during 2008 with at least two Centre investigators or students as co-authors, 38 (67%) of those co-authors hold advanced degrees in different disciplines. Two-thirds (20 of 30) of our co-authored C1 journal publications were interdisciplinary in this sense. Our Winter School is intentionally interdisciplinary, and all ACB students and postdoctoral researchers are strongly encouraged to attend.

Mark was also briefly quoted in The Australian Financial Review (13 October 2008) as part of coverage of the visit of Professor Larry Smarr (Director, California Institute for Telecommunications and Information Technology) to Australia. Professor Smarr, the Australian American Leadership Dialogue 2008 Scholar, toured Australia to make the case for national investment in Grid fabric, specifically ultra-fast broadband networks linking major research infrastructure and high-performance computers, with visualisation based on tiled-wall LCD displays such as the OptIPortal.

Professor Michael Langston was featured in the Spring 2008 issue of The University of Tennessee’s faculty newsletter, available online at http://www.cs.utk.edu/~langston/faculty-focus.pdf

Dr Nick Hamilton coordinated ACB presentations for the OptIPortal visualisation cluster at the GIW-2008 conference in December 2008.

Key Performance Indicators
2. Research Training and professional education

Recruitment
In 2008, ACB recruited:

- Dr Adrian Cootes, Research Officer into Project 1.3 (Data analysis and knowledge discovery in very large biological datasets) with Prof. Shoba Ranganathan.
- Dr Gabriel Fung, Research Officer into Project 1.3 (Data analysis and knowledge discovery in very large biological datasets) with Prof. Xiaofang Zhou.
- Dr John Hawkins, Research Officer into Project 2.1 (Computational discovery of gene-regulatory motifs and motif clusters) with Dr Timothy Bailey.
- Mr Andrew Newman, Research Assistant into Project 1.2 (Data and process management for the Visible Cell®) with Prof. Mark Ragan.
- Dr Kim-Anh Lê Cao, Research Officer into Project 2.2 (Differential expression of genes towards pathway discovery) with Prof. Geoff McLachlan.
- Dr Shev MacNamara, Research Officer into Project 3.1 (Mathematical modelling of genetic regulatory networks) with Prof. Kevin Burrage.
- Dr John Belward and Dr Fawang Liu, Research Officers into Program 3 (Modelling dynamic cellular processes) with Prof. Kevin Burrage.
- Mr Daniel Johnstone and Dr Martin Gómez-Ravetti, Research Assistants into Project 4.2 (Optimisation-based approaches to large systems) with A/Prof. Pablo Moscato.

Professional/technical training and advanced career development

Seminar/Symposium hosted and/or sponsored by ACB

9th International Symposium on Parallel Architectures, Algorithms and Networks
In 2008 ACB sponsored the travel of CI Dr Brad Marsh, a keynote speaker at the 9th International Symposium on Parallel Architectures, Algorithms and Networks, Sydney (7–9 May 2008).

3rd IAPR Conference on Pattern Recognition in Bioinformatics
ACB financially supported student participation in the 3rd International Association for Pattern Recognition (IAPR) Conference on Pattern Recognition in Bioinformatics (PRIB 2008), Melbourne (15–17 October 2008).

19th International Conference on Genome Informatics
Please see page 38.

2008 Winter School in Mathematical and Computational Biology (7–11 July 2008)
Our 2008 Winter School in Mathematical and Computational Biology was held first time in Newcastle, hosted by CI A/Prof. Pablo Moscato with the local support of colleagues from ACB, the University of Newcastle’s School of Electrical Engineering and Computer Science, and the Newcastle Priority Research Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine. Winter School 2008 featured 14 speakers (9 international, 5 national), and attracted 81 participants from 35 institutions across Australia and from Finland, India, Spain, UK and USA. Of these 78% were researchers or postgraduate students, and 22% advanced undergraduates. One-third identified a biological background, with most others from mathematics, statistics, computer science or IT. About 41% represented Australian universities. Again, we received highly positive written feedback from the participants about the topics and quality of presentations.

Each of the five days focused on a different theme:

Day 1: Entropy and statistical complexity
Day 2: Data mining, machine learning and new technologies
Day 3: Combinatorial optimisation and applications
Day 4: Network pathway analysis in bioinformatics
Day 5: Advances in proteomics and genomics

Our distinguished invited speakers were:

- Prof. Vladimir Brusic
  Harvard University, USA
- A/Prof. Carlos Cotta
  University of Malaga, Spain
- Prof. Charles Cantor
  Chief Scientific Officer, Sequenom, USA
- Dr Martin Gómez-Ravetti
  ACB and The University of Newcastle, Australia
- Prof. Michael Langston
  ACB and The University of Tennessee, USA
- A/Prof. Joerg Mattes
  The University of Newcastle, Australia
- A/Prof. Pablo Moscato
  ACB and The University of Newcastle, Australia
- Dr Ferrante Neri
  University of Jyväskylä, Finland
- Dr Michael Norman
  CEO, Scapa Technologies, UK
- Prof. Shoba Ranganathan
  ACB and Macquarie University, Australia
- Dr Carlos Riveros
  The University of Newcastle, Australia
- Dr Osvaldo Rosso
  University of Buenos Aires, Argentina
- Prof. Rodney Scott
  The University of Newcastle, Australia
- Dr Muhammad Shoib Sehgal
  ACB and The University of Queensland, Australia
- Prof. Constantino Tsallis
  Brazilian Center for Physics Research Brazil
- Prof. Michael Zuker
  Rensselaer Polytechnic Institute, USA
A public lecture was held on Day 4 featuring Professor John Forbes, Director of Research for the Australian New Zealand Breast Cancer Trials Group (ANZ BCTG), Professor of Surgical Oncology at the University of Newcastle, and Director, Department of Surgical Oncology at the Calvary Mater Hospital.

2008 Winter School in Mathematical and Computational Biology, Newcastle. From left: Vinh Dang, Nikhil Lilaria, Dr Muhammad Shoaib Sehgal, Prof. Shoba Ranganathan, Prof. Michael Zuker (Rensselaer Polytechnic Institute, USA), Prof. Michael Langston (University of Tennessee) and Prof. Mark Ragan.

2008 Winter School in Mathematical and Computational Biology, Newcastle. From left: Dr Ferrante Neri (University of Jyväskylä, Finland), Prof. Constantino Tsallis (Santa Fe Institute, USA and Brasilian Center for Physics Research, Rio de Janeiro, Brazil), Ms Armita Zarnegar (PhD student, University of Ballarat), Dr Osvaldo Rosso (Universidad Nacional de Buenos Aires, Argentina and Centre for Bioinformatics, Biomarker Discovery and Information-based Medicine, The University of Newcastle (CIBM)), Dr Carlos Riveros (CIBM), A/Prof. Pablo Moscato, Dr Michael Norman (Chief Executive Officer, Scapa Technologies, Scotland); and A/Prof. Carlos Cotta (University of Malaga, Spain).
Other research and training and professional education

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

Prof. Kevin Burrage
- Doctoral training programs, University of Oxford (March)

Prof. Michael Fellows
- International Summer School on Fixed Parameter Tractable Algorithms (FPT2008), Shanghai (June)

Dr Nick Hamilton
- Future Research Leaders Program, University of Queensland (November)

Dr Markus Hegland
- Course on Approximation Theory at the AMSI Summer School, Monash University (January)

Prof. Geoff McLachlan
- Programme on Statistical Theory and Methods for Complex, High-Dimensional Data, Isaac Newton Institute for Mathematical Sciences, Cambridge (January-June)

Prof. John Mattick
- Centre for Military & Veterans Health Think Tank, Canberra (May)
- Queensland Institutes of Health Symposium, Coolum (August)
- New South Wales Supreme Court Annual Conference, Shoal Bay (August)

Postgraduate and undergraduate courses in the Centre’s area

Our investigators and researchers participated in the following for-credit courses and programs in 2008:

Dr Timothy Bailey and Dr Mikael Bodén re-designed and taught the University of Queensland undergraduate course “Advanced Bioinformatics”.

Dr Timothy Bailey
- BIOL 3014 Advanced Bioinformatics (lecturer/coordinator)

Dr Mikael Bodén
- BSc Bioinformatics dual major (convener)
- BInfTech Bioinformatics dual major (convener)
- BIOL 3014 Advanced Bioinformatics (lecturer/coordinator)

A/Prof. Phoebe Chen
- SIT781 Introduction to Bioinformatics (unit chair and lecturer)
- SIT772 Database and Information Retrieval (unit chair and lecturer)
- SIT362 Advances in Interactive Media (unit chair and lecturer)

Dr Gabriel Fung
- INFS4203/7203 Data Mining

Dr Markus Hegland
- 3359 Topics in Computational Mathematics

Dr Karin Kassahn
- BIOL3010 Ecological and Evolutionary Genetics (two guest lectures)

Dr Brad Marsh
- BIOL3006 Molecular Cell Biology (lecturer)
- BIOC3004 Structural Biology: Macromolecular Structure (lecturer)
- Undergraduate Research Opportunities Program (URDP) scholarship student projects

A/Prof. Pablo Moscato
- COMP4120 Knowledge Discovery and Data Mining (lecturer/Coordinator)
- COMP2270, COMP270 Formal Languages and Automata (lecturer/Coordinator)

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

Prof. Shoba Ranganathan
- CBMS223 Biochemistry and Molecular Biology I (Unit coordinator and lecturer)
- CBMS301 Technology Miniproject (Unit coordinator)
- CBMS880 Selected Topics in Biotechnology (lecturer)

Dr Mohammad Shoai Sehgal
- COMP3702 Artificial Intelligence (lecturer)

Prof. Xiaofang Zhou
- INFS4203/7203 Data Mining

Scientists in Schools Partnership

Dr Brad Marsh is the Centre’s point of contact for the CSIRO Scientists in Schools Partnership program. He gave two presentations explaining basic cell biology and how we image insulin-secreting cells in 3D to the combined 6/7 advanced learning class at Gracerville State School, Gracerville, Brisbane (April 2008). Brad and science teacher Ms Sandra Davey assisted the class in extracting DNA from strawberries, and loaned them some laboratory equipment and glassware for the experiments (August 2008). Brad also organised an on-site visit and tour of the IMB and electron microscopy facilities for the class (30 students, teacher, and accompanying parents) in August 2008.

Dr Melissa Davis was also involved with the CSIRO Scientists in Schools Partnership. Working with Ms Andrea O’Brien, faculty organiser of the Ironside Science Club (Ironside State School, Brisbane), Melissa made several presentations through the year on molecular biology and genetics, and gave three presentations at their 2008 Science Day.
3. International, national and regional links and networks

Visitors

International visitors
- Wagner Aioffi, Universidade Federal de Minas Gerais (UFMG), Brazil
- Prof. Vladimir Brusic, Harvard University, USA
- Dr Julio Collado-Vides, UNAM Universidad Nacional Autonóma de México, Mexico
- Mr Graham Cameron, Associate Director, EMBL-EBI, UK
- Prof. Charles Cantor, Chief Scientific Officer, Sequenom, USA
- Dr Cheong Xin Chan, University of Iowa, USA
- Prof. Vicki Chandler, University of Arizona, USA
- A/Prof. Carlos Cotta, University of Malaga, Spain
- Dr Paul Horton, Computational Biology Research Center, AIST, Japan
- Prof. David Keyes, Columbia University, New York
- Prof. Ming Li, Canada Research Chair in Bioinformatics, University of Waterloo, Canada
- Dr Yi-xue Li, Shanghai Center for Bioinformation Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai Jiaotong University, China
- Dr Cliona Molony, Head of Statistical Genetics, Rosetta Inpharmatics/Merck Research Labs, USA
- Prof. Nancy Moran, University of Arizona, USA
- Dr Ferrante Neri, University of Jyväskylä, Finland
- Dr Michael Norman, CEO, Scapa Technologies, UK
- Prof. Howard Ochman, University of Arizona, USA
- Prof. Linda Petzold, University of California Santa Barbara
- Dr Juergen Plitzko, Max Planck Institute of Biochemistry, Martinsried, Germany
- Prof. Mike Powell, University of Cambridge
- Dr Xiangtong Qi - Hong Kong University of Science and Technology (UST), Hong Kong.
- Prof. Ulrich Rüde, Universität Erlangen-Nürnberg
- Prof. Allen Rodrigo, Bioinformatics Institute, New Zealand
- Dr João Sarubbi - CEFET/MG, Brazil
- Prof. Ernst Stephan, Leibniz University, Hannover
- Dr Martin Swain, School of Biomedical Sciences, University of Ulster, UK
- George Teodoro, Universidade Federal de Minas Gerais (UFMG), Brazil
- Prof. Constantino Teallis, Brazilian Center for Physics Research, Brazil
- Prof. Alexis Vandenbon, University of Tokyo, Japan
- Prof. Michael Zeker, Rensselaer Polytechnic Institute, USA

National visitors to one or more ACB nodes
- Dr Ian Atkinson, James Cook University
- Dr Eric Chu, Monash University
- A/Prof. Christine Clarke, Westmead Institute for Cancer Research, Sydney
- Prof. Larry Forbes, University of Tasmania
- Dr Marcus Heisler, EMBL Australia
- Dr Elizabeth Murchison, Australian National University
- Dr Ken Pang, Ludwig Institute, Melbourne
- Professor David Ravine, University of Western Australia
- Dr William Ritchie, Centenary Institute, Sydney
- Prof. Ian Sloan, University of New South Wales
- Dr Rohan Williams, The John Curtin School of Medical Research, Australian National University
- Prof. Rob Womersley, University of New South Wales

Collaborative national and international workshops and exchanges

In addition to the 2008 Winter School and Student Symposium, our CIs, PdFs, PhDs presented at or participated in workshops as follows:

Chief Investigators
- Prof. Kevin Burrage
- Numerical Modelling of Complex Dynamical Systems, Leiden, Netherlands (May)
- A/Prof. Phoebe Chen
- Eighth International Workshop on Data Mining in Bioinformatics (BIOKDD 2008), Las Vegas, USA (August)
- Co-Chair, Second International Workshop on Conceptual Modelling for Life Sciences Applications (CMLSA2008), Barcelona, Spain (October)
- 2008 International Computer Symposium (ICS’08), Workshop on Medical and Bio-informatics, Taipei, Taiwan (November)
- Prof. Michael Fellows
- 34th International Workshop on Graph-Theoretic Concepts in Computer Science (WG2008), Durham, UK (July)
- Second International Frontiers of Algorithmics Workshop (FAW 2008), Changsha, China (June)

Dr Markus Hegland
- “Days honouring the lives of Gene Golub and Ron Mitchell”, Canberra (February)
- Sessions on information based complexity and on linear algebra and inverse problems at FODC 09, Hong Kong, City University (March)
- Computational Techniques and Applications Conference (CTAC 2009) conference, Canberra, (July)
- Conference on Modeling, Simulation and Optimization of Complex Processes (MSOCP 08), University of Heidelberg, Germany (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. Please see also to page 44, Invitations to visit leading international laboratories.

National collaborating institutions:
- Agilent Technologies
- ARC Centre for Complex Systems
- ARC Research Network in Enterprise
  Information Infrastructure
- ARC Special Research Centre for Functional
  and Applied Genomics
- Australian Centre for Plant Functional
  Genomics
- Australian e-Health Research Centre,
  CSIRO
- Australian Microscopy & Microanalysis
  Research Facility
- Australian Regenerative Medicine Institute,
  Monash University
- Australian Stem Cell Centre
- Bio21 Molecular Science and
  Biotechnology Institute
- CSIRO Mathematical and Information
  Sciences
- Elizabeth Macarthur Agricultural Institute,
  NSW Department of Primary Industries
- Energy Edge Pty Ltd
- Flinders University of South Australia
- Griffith University
- IBM Australia
- Invitrogen Life Technologies
- James Cook University
- Ludwig Institute for Cancer Research
- Microsoft Australia
- Monash University
- Murdoch Childrens Research Institute
- Museum Victoria
- National ICT Australia
- Nielson Media
- Queensland Brain Institute
- Queensland Cyber Infrastructure
  Foundation
- Queensland Institute of Medical Research
- Queensland University of Technology
- Rulequest Research Pty Ltd
- Swinburne University of Technology
- Tasmania ICT Centre, CSIRO
- The Bionic Ear Institute
- The University of Melbourne
- The University of Sydney
- The University of Western Australia
- The Walter and Eliza Hall Institute of
  Medical Research
- University of Technology Sydney
- Westmead Hospital, Sydney

Prof. Geoff McLachlan
- High Dimensional Statistics in Biology Workshop, Cambridge (March-April)
- Quantitative Biology and Applied Statistics Seminar, The University of Reading, UK (May)
- CRiSM (Centre for Research in Statistical Methodology) Seminar, The University of Warwick, UK (May)
- Second Annual Workshop on Statistical Methods for Genetic Analysis, QUT (November)

Dr Brad Marsh
- Session Chair, Mitochondrial Imaging and Dynamics, ‘AussieMit’ Workshop, Melbourne (November)

A/Prof. Pablo Moscato
- 2008 NSW Sydney Bioinformatics Research Symposium, Sydney (November)

Prof. John Mattick
- Foundation des Treilles Meeting on “New catalytic and regulatory functions of RNA in eukaryotes,” Nice, France (March)

Prof. Shoba Ranganathan
- Workshop on Education in Bioinformatics and Computational Biology (WWEBCB) 2008, Taipei, Taiwan (October)

Prof. Xiaofang Zhou
- 1st International Workshop on Information-explosion and Next Generation Search (INGS), Shenyang, China (April)

Postdoctoral fellows, Research Officers and students

Mr. Khar Heng Choo
- Workshop on Education in Bioinformatics and Computational Biology (WWEBCB) 2008, Taipei, Taiwan (October)
International collaborating institutions:
Aalborg University, Denmark
Academia Sinica, Institute of Information Science, Taiwan
Academia Sinica, Institute of Software, China
Aghia Sophia Children’s Hospital, Greece
AITIA International Informatics Inc., Hungary
Albert Einstein College of Medicine, USA
Albert Ludwigs Universität Freiburg, Germany
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Hungary
Allen Institute for Brain Science, Seattle, USA
Arkansas State University, USA
AT&T Laboratories, USA
Banco Central do Brazil, Brazil
Baylor College of Medicine, USA
Bioinformatics Institute of Singapore, Singapore
Broad Institute of MIT and Harvard, USA
Buenos Aires Science Research Commission, Argentina
Burnham Institute for Medical Research, USA
California Institute of Technology, USA
Centro de Regulación Genómica, Spain
Chao Family Comprehensive Cancer Center, USA
Chinese University of Hong Kong, China
Chonnam National University, Korea
Clausthal University of Technology, Germany
CNR-Fondazione scuola di S. Giorgio, Venezia, Italy
Cold Spring Harbor Laboratory, USA
Collegium Budapest, Hungary
CombiMatrix Corporation, USA
Computational Sciences Center of Emphasis, Pfizer, USA
Cranfield School of Management, UK
Dalhousie University, Canada
Dana Farber Cancer Institute, USA
Dokuz Eylül University, İzmir, Turkey
Dublin Institute of Technology, Ireland
Durham University, UK
Eidgenössische Technische Hochschule Zürich, Switzerland
Eötvös Loránd University, Hungary
Florida State University, USA
Freie Universität Berlin, Germany
Friedrich-Schiller-Universität Jena, Germany
Genome Institute of Singapore, Singapore
Hacettepe Universitesi, Turkey
Harbin Institute of Technology, China
Harvard University, USA
Hälsöö Islands, Iceland
Hohai University, China
Hong Kong University of Science and Technology, China
Hospital Italiano, Buenos Aires, Argentina
Hospital Ramos Mejía, Buenos Aires, Argentina
Howard University, USA
Imperial College London, UK
Institut national de Recherche en Informatique et en Automatique, France
Instituto Balseiro, Centro Atómico Bariloche, Argentina
Institute of Biotechnology, Bulgarian Academy of Sciences, Bulgaria
Institute for Infocomm Research, Singapore
Isaac Newton Institute for Mathematical Sciences, UK
ISIS Pharmaceuticals, USA
Istituto Italiano di Tecnologia, Genoa, Italy
Joint Genome Institute, US Department of Energy, USA
John Radcliffe Hospital, Oxford, UK
Johns Hopkins University, USA
Karolinska Institutet, Sweden
Kölçür University, Turkey
Kwangyang Institute of Science and Technology, Korea
Kyoto University, Japan
Lawrence Berkeley National Laboratory, USA
Lerner Research Institute, USA
Max-Planck-Institut für Informatik, Germany
Max-Planck-Institut für Molekulare Genetik, Germany
Max-Planck-Institut für Molekulare Biologie und Genetik, Germany
McGill University, Canada
Morendun Research Institute, UK
Nanjing University, China
Nanyang Technological University, Singapore
National Institute of Advanced Industrial Science and Technology, Japan
National Institutes of Health, USA
National University of Singapore, Singapore
National Yang-Ming University, Taiwan
Oak Ridge National Laboratory, USA
Osaka University, Japan
Otto-von-Guericke Universität Magdeburg, Germany
Parahyangan Catholic University, Indonesia
Peking University, China
Platform Computing SARL, France/Germany
Pontificia Universidad Católica de Chile, Chile
Pontificia Universidad Católica de Valparaíso, Chile
Poznan Supercomputing and Networking Center, Poland
Purdue University, USA
Queen Silvia Children’s Hospital, Göteborg, Sweden
Raytheon Inc., USA
RIKEN Genomic Sciences Center, Japan
RIKEN Wako Institute, Japan
Salfield Systems, USA
Sandia National Laboratory, USA
Scripps Institution of Oceanography, USA
Sichuan University, China
Singapore Management University, Singapore
SRA International Inc., USA
Technion – Israel Institute of Technology, Israel
Technischen Universität Berlin, Germany
Temasek Life Sciences Laboratory, Singapore
Tsinghua University, China
The Open University, UK
Universidad Autónoma de Madrid, Spain
Universidad de Buenos Aires, Argentina
Universidad de la Republica, Uruguay
Universidad de los Andes, Chile
Universidad de Navarra, Spain
Universidad de Valladolid, Spain
Universidad de Valparaíso, Chile
Universidad de Zaragoza, Spain
Universidad Favaloro, Argentina
Universidad Nacional de Córdoba, Argentina
Universidad Nacional de General Sarmiento, Argentina
Universidad Nacional de La Plata, Argentina
Universidad Nacional de Mar del Plata, Argentina
Universidad Nacional de Quiímes, Argentina
Universidad Nacional de Rosario, Argentina
Universidad Nacional de San Martín, Argentina
Universidad Nacional de Tucumán, Argentina
Universidade Católica de Brasília, Brazil
Universidade Federal de Rio Grande do Sul, Brazil
Universidade Federal de Minas Gerais, Brazil
Università di Perugia, Italy
Università di Roma “La Sapienza”, Italy
Universitat de les Illes Balears, Spain
Universitat Politècnica de Catalunya, Spain
Universitat Pompeu Fabra, Spain
Universität Bonn, Germany
Universität Lübeck, Germany
Universität Trier, Germany
Université des Sciences d’Orsay, France
Université Pierre et Marie Curie, France
Université de Rennes, France
Université de Strasbourg, France
Universiteit van Amsterdam, The Netherlands
University of Utrecht, The Netherlands
University of Bergen, Norway
University of Alberta at Birmingham, USA
University of Auckland, New Zealand
University of British Columbia, Canada
University of California at Berkeley, USA
University of California at Irvine, USA
University of California at San Diego, USA
University of California at Santa Barbara, USA
University of Cambridge, UK
University of Canterbury, New Zealand
University of Chicago, USA
University of Colorado, USA
University of Dar es Salaam, Tanzania
University of Dayton, USA
University of Dublin, Ireland
University of Florida, USA
University of Georgia, USA
University of Haifa, Israel
University of Illinois at Chicago, USA
University of Iowa, USA
University of Kentucky, USA
University of Leeds, UK
University of Liverpool, UK
University of Maryland at Eastern Shore, USA
University of Minnesota, USA
University of Montana, USA
University of North Carolina, USA
University of Oxford, UK
University of Reading, UK
University of Southampton, UK
University of Texas at Austin, USA
University of Ulster, UK
University of Vermont, USA
University of Victoria, Canada
University of Waikato, NZ
University of Warwick, UK
University of Washington, USA
University of Waterloo, Canada
University of Western Ontario, Canada
University of Wisconsin, USA
Uppsala University, Sweden
Victoria University of Wellington, New Zealand
Wake Forest University, USA
Xiamen University, China
Yale University, USA

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 2008 with executive, management and professional committees, advisory and review committees, and similar continuing bodies (excludes fellowships, journal editorial boards and learned societies):

Prof. Kevin Burrage
- Research Committee, Oxford Centre for Integrative Systems Biology

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

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

Dr Markus Hegland
- International panel of assessors, Deutsche Forschungsgemeinschaft SPP 1243
- Steering Committee, Computational Mathematics Group

Prof. Geoff McLachlan
- President-Elect, International Federation of Classification Societies
- College of Experts, Australian Research Council

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

Prof. John Mattick
- Council of Scientists, Human Frontier Science Program

A/Prof. Pablo Moscato
- Strategic Advisory Committee, Hunter Medical Research Institute.
- Deputy Leader, Information-based Medicine Program, Hunter Medical Research Institute.

Prof. Mark Ragan
- Vice-president, Bioinformatics Australia (AusBiotech)
- Chair, Executive Group, Association of Asian Societies of Bioinformatics
- Committee on Global Collaborations, European Life Sciences Infrastructure for Biological Information (ELIXIR)
- Steering Committee, Australian National Data Service
- Board of Management, Australian Research Enabling Environment (ARCHER)
- National Working Group and Queensland Coordinator, EMBL Australia
- Chair, Management Committee, and Member, Industrial Research Advisory Panel, Queensland Facility for Advanced Bioinformatics
- Advisory Board, QosCosGrid (Quasi-Opportunistic Supercomputing for Complex Systems in Grid Environments) project, European Sixth Framework
- External Examiner, Masters of Biotechnology Programme, Universiti Malaya (2005-2010)

Prof. Shoba Ranganathan
- President, Asia-Pacific Bioinformatics Network
- Chair, International Conference in Bioinformatics Executive Committee
- Chair, S* Life Science Informatics Alliance
- Steering Committee, Bioinformatics Australia

Dr Rohan Teasdale
- Steering Committee, International Conference on Bioinformatics
- Steering Committee, Bioinformatics Australia
- National Association of Research Fellows of NHMRC

Prof. Xiaofang Zhou
- Executive Committee, 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
- 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 2008:

- Prof. Limsoon Wong, National University of Singapore & Dr. J. Li, Nanyang Technological University (Singapore)
  Development of knowledge discovery for high-dimensional biomedical data

- Prof. Michael Fellows
  - Prof. Jan Arne Telle & Prof. Pinar Heggernes, University of Bergen (Norway)
    PARALGTD: Parameterized Algorithms and FPT Kernelization
  - Prof. Rolf Niedermeier, Friedrich-Schiller-Universität Jena (Germany)
    Parameterized graph clustering algorithms for bioinformatics applications

- Dr Markus Hegland
  - Prof. Michael Griebel, University of Bonn (Germany)
    MAP density estimation with Gaussian Process Priors
  - Prof. Per Lotstedt, University of Uppsala (Sweden)
    Numerical DME solvers
  - Prof. Martin Gutknecht, ETH Zürich (Switzerland)
    Efficient Algorithms for Large Linear Additive Models
  - Dr Jochen Garcke, Technische Universität Berlin (Germany)
    Sparse grid combination technique

- Dr Karin Kassahn
  - Dr Stefano Guindon, University of Auckland (New Zealand)
    Estimating selection in vertebrate protein families

- Prof. Geoff McLachlan
  - Dr Jill Mesirov, Broad Institute of MIT and Harvard (USA)
    Automated high-dimensional flow cytometric data analysis

- Prof. Brad Marsh
  - Prof. Michael Hayden and A/Prof. Bruce Verchere, University of British Columbia (Canada)
    Structure-function studies of intracellular cholesterol accumulation in ACBA1 KO mice
  - A/Prof. Maryann Martone, National Center for Microscopy and Imaging Research, University of California at San Diego (USA)
    Cell Centered DataBase (http://ccdb.ucsd.edu/CCDBWebSite/index.html)
  - Prof. David Mastronarde, Boulder Laboratory for 3D Electron Microscopy of Cells, University of Colorado (USA)
    Supermontaging: reconstructing large cellular volumes by lateral stitching of adjacent tomograms
  - Prof. Christopher Rhodes, University of Chicago (USA)
    Elucidating regulated autophagy in pancreatic islet beta cells from mice and humans
  - A/Prof. Niels Volkmann, Burnham Institute for Medical Research (USA)
    Semi-automated pre- and post-processing approaches to segmentation of cellular tomographic data

- Prof. John Mattick
  - Professor Mark Mehler, Institute for Brain Disorders and Neural Regeneration, Albert Einstein College of Medicine of Yeshiva University (USA)
    Noncoding RNA expression, function and editing in brain development and cognition
  - Dr Bodo Lange, Max-Planck Institute for Molecular Genetics (Germany)
    Noncoding RNAs in centrosomes
  - Dr Charles Perou, University of North Carolina (USA)
    Noncoding RNA expression and function in breast cancer
  - Dr Jeff Rosen, Baylor College of Medicine, Houston (USA)
    Noncoding RNA expression and function in breast development
  - Dr David Spector, Cold Spring Harbor Laboratories (USA)
    Noncoding RNAs in muscle development and paraspeckle function
  - Prof. Eric Westhof, University of Strasbourg (France)
    Structure-based genome alignments and estimation of conservation, deep sequencing and analysis of brain small RNAs

- A/Prof. Phoebe Chen
  - Prof. Limsoon Wong, National University of Singapore & Dr. J. Li, Nanyang Technological University (Singapore)
    Development of knowledge discovery for high-dimensional biomedical data
4. End-user links

Commercialisation activities

ACB was directly involved with a number of commercial partners in 2008:

- Prof. Kevin Burrage and Prof. Mark Ragan collaborated with Pfizer Research Technology Center (USA) in the BioMANTA project (see our 2007 Annual Report, page 41).
- Dr Nick Hamilton met with Dr Stephan Kuppig (Carl Zeiss AG, Germany) to discuss commercialisation potential of his iCluster and OBCOL software (September 2008).
- Prof. John Mattick worked with Invitrogen Life Technologies (USA) in the development and commercialisation of mouse and human noncoding RNA arrays (“NCode”).
- Prof. Mark Ragan and colleagues collaborated with commercial partners in the QosCosGRID project funded by the EU Sixth Framework Programme.

Government, industry and business briefings

The events below are in addition to the numerous meetings, teleconferences and interactions with colleagues in government, industry and business implicit in the above collaborations and links.

Prof. Mark Ragan

- Australian delegation organised by the Australian Academy of Science on behalf of the Department of Innovation, Industry, Science and Research (DIISR) – Australia-Brazil Workshop on Biotechnology Innovations for Agriculture, Brasilia (May)
- Executive briefing and presentation Pfizer, Brisbane (May)
- Facilitator, National Technology Roadmap for Bioinformatics, Ministry of Science Technology & Innovation Malaysia, Kuala Lumpur (August)
- Forum on Innovation and commercialisation in genome-scale bioinformatics, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)

Prof. Shoba Ranganathan

- Forum on the development of genome-scale bioinformatics in China and India, 19th International Conference on Genome Informatics (GIW-2008), Gold Coast (December)

Trained/ing in technology transfer and commercialisation

Dr Melissa Davis attended the “Engineering your IP” hosted by IP Australia and Institute of Engineers Australia in June 2008.

Public awareness programs

Dr Melissa Davis and Dr Karin Kassahn participated in the “Talking Scientists” public outreach programme sponsored by Queensland Government (September 2008), as well as preparatory workshops.
5. Organisational support

New organisations recruited to or involved in the Centre

Three new organisations began formal involvement with the ARC Centre of Excellence in Bioinformatics in 2008: Macquarie University, The University of Auckland, and The University of Tennessee.

Infrastructure

World-class expertise

Bioinformatics is a key research technology, particularly in enabling advances in life sciences and biotechnology. ACB researchers, students and collaborators have access to top Australian experts in many areas of bioinformatics, 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 more than 600 research-dedicated compute cores and nearly 100 TB research-dedicated storage at IMB, newly upgraded SGI Altix systems at UQ/QCIF, a new cluster at Newcastle Bioinformatics Initiative, and (via merit allocation) the 2200-processor national computing facility (NIC/NCRIS). We are a partner in the European Union Sixth Framework project Quasi-Opportunistic Supercomputing for Complex Systems in GRID Environments (www.qoscospgrid.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. Many of these have been partnered-into QFaB, which is further developing this infrastructure with commercial bioinformatics software (e.g. SRS) and data sources. QFaB has also adopted some of the ARCHER e-research tools (www.archer.edu.au) including Pione. We have access to storage at ANU and UQ currently at 100s–TB scale, and being developed toward PB scale.

Microscopy

The $1.4M ACRF/IMB Dynamic Imaging Facility and the High-content Screening facility were established with active involvement of CI Rohan Teasdale. The former includes two state-of-the-art Zeiss Meta S10 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, with two SOLID systems in place during 2008.

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 has signed an agreement to become the Australasian reference site and distributor for GeneGo, and will host a local mirror.
### Key Performance Indicators

#### Annual Cash Contributions from Other Organisations 2008–2010

<table>
<thead>
<tr>
<th>Organisation</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARC*</td>
<td>1,144,719</td>
<td>1,167,613</td>
<td>1,190,966</td>
</tr>
<tr>
<td>The University of Queensland**</td>
<td>525,000</td>
<td>525,000</td>
<td>525,000</td>
</tr>
<tr>
<td>Australian National University</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Deakin University</td>
<td>30,000</td>
<td>30,000</td>
<td>30,000</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>70,000</td>
<td>70,000</td>
<td>70,000</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>80,000</td>
<td>80,000</td>
<td>80,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,859,719</strong></td>
<td><strong>1,882,613</strong></td>
<td><strong>1,905,966</strong></td>
</tr>
</tbody>
</table>

* The numbers shown assume indexing at 2% per year. **In addition to the cash support identified here, the Centre will benefit from IGS generated by the core ARC grant funds at UQ distributed via IMB.

#### Annual In-Kind Contributions: Collaborating Organisations 2008–2010

<table>
<thead>
<tr>
<th>Organisation</th>
<th>2008</th>
<th>2009*</th>
<th>2010*</th>
</tr>
</thead>
<tbody>
<tr>
<td>The University of Queensland</td>
<td>551,499</td>
<td>579,074</td>
<td>608,027</td>
</tr>
<tr>
<td>Equipment</td>
<td>20,000</td>
<td>20,000</td>
<td>20,000</td>
</tr>
<tr>
<td>Other</td>
<td>212,105</td>
<td>73,000</td>
<td>73,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>783,604</strong></td>
<td><strong>672,074</strong></td>
<td><strong>701,027</strong></td>
</tr>
<tr>
<td>Australian National University</td>
<td>22,988</td>
<td>24,138</td>
<td>25,345</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>8,800</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>31,788</strong></td>
<td><strong>24,138</strong></td>
<td><strong>25,345</strong></td>
</tr>
<tr>
<td>Deakin University</td>
<td>22,052</td>
<td>23,155</td>
<td>24,312</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>12,750</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>34,802</strong></td>
<td><strong>23,155</strong></td>
<td><strong>24,312</strong></td>
</tr>
<tr>
<td>Macquarie University</td>
<td>34,420</td>
<td>36,141</td>
<td>37,948</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>48,618</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>83,038</strong></td>
<td><strong>46,141</strong></td>
<td><strong>47,948</strong></td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>74,796</td>
<td>78,536</td>
<td>82,463</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>21,738</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>96,534</strong></td>
<td><strong>78,536</strong></td>
<td><strong>82,463</strong></td>
</tr>
<tr>
<td>University of Auckland</td>
<td>6,985</td>
<td>7,334</td>
<td>7,700</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>6,985</strong></td>
<td><strong>7,334</strong></td>
<td><strong>7,700</strong></td>
</tr>
<tr>
<td>University of Tennessee</td>
<td>32,416</td>
<td>33,389</td>
<td>34,391</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>32,416</strong></td>
<td><strong>33,389</strong></td>
<td><strong>34,391</strong></td>
</tr>
<tr>
<td>IBM</td>
<td>20,000</td>
<td>20,000</td>
<td>20,000</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>133,353</td>
<td>133,353</td>
<td>133,353</td>
</tr>
<tr>
<td>Other – Tivoli Application</td>
<td>30,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>183,353</strong></td>
<td><strong>153,353</strong></td>
<td><strong>153,353</strong></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>1,252,521</strong></td>
<td><strong>1,038,119</strong></td>
<td><strong>1,076,539</strong></td>
</tr>
</tbody>
</table>

* The numbers shown assume indexing at 5% per year for the personnel component only, in addition to probable salary increments where known.

**TOTAL (Annual Cash + Annual In-Kind Contributions)**  
3,112,240 | 2,920,732 | 2,982,505
6. National benefit

Measures of expansion of Australia’s capability in the priority area(s)

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). The review appeared in *PLoS Computational Biology* on 29 August 2008.

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. In mid-2008 National Computational Infrastructure (NCI), successor under NCRIS to Australian Partnership in Advanced Computing (APAC), called for expressions of interest in developing and hosting Specialised Facilities; it is a measure of the expansion of Australia’s capability that at least two expressions of interest were submitted for an NCI SF in bioinformatics.

**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) and the Australian Research Council (ARC) have committed support for an initiative under Australia’s associate membership. The profile of bioinformatics within EMBL Australia has been high, led by an initiative to mirror substantial components of the European Bioinformatics Institute (EBI) data facility in Australia. During 2008 Mark Ragan represented UQ on the EMBL Australia Working Group. Dr Graham Cameron, Deputy Director of EBI, visited Australia (including ACB) in September-October 2008 and was a keynote speaker at the e-Research Australasia conference in Melbourne. Dr Ewan Birney, a senior scientist at EBI and key driver of major bioinformatics environments including Ensembl, will spend almost one month in Australia in early 2009 under the sponsorship of EMBL Australia; his visit will raise of the profile of bioinformatics as an area of potentially enhanced research collaboration between Australia and Europe.

**Australian Bioinformatics Network and Bioinformatics Australia**

Following a recommendation in the National Bioinformatics Strategy (2005), in mid-2008 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 was 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. 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 Committee, while Mark Ragan was elected Vice-President (Tim Littlejohn is President). Bioinformatics Australia hosted the 2008 International Conference on Genome Informatics (GIW-2008), with the 2008 annual BA conference combined into 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 start-up by seconding certain infrastructure staff on a term basis; in return QFAB is managing the corresponding components of our infrastructure requirements, originally focusing on the Visible Cell® project but during 2008 transitioning to support of ACB’s Research Program 2 (Phenotype-informed discovery of networks and systems). 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 is using Plone, a highly capable portal developed in the ARCHER e-research project, to provide data to partners in Adelaide, Brisbane, Melbourne, Perth and Sydney as part of a large project on nuclear receptor signalling funded by the National Breast Cancer Foundation. The latter program is likewise highly synergistic with our Research Program 2, and with parallel work in research groups at the University of Newcastle and the University of Queensland.

QosCosGRID

During 2008 ACB became a key node in the EU Sixth Framework Information Society Technology project “Quasi-opportunistic supercomputing for complex systems in Grid environments”. QosCosGRID Technical Manager Dr Krzysztof (Chris) Kurowski was based within the UQ node of ACB during 2008 and managed the project from Australia. An early outcome is that a small cluster in IMB is now part of the QosCosGRID computing network, and research within ACB’s Program 3 (spatial models of diffusion in lipid membranes) was demonstrated at the 2008 EU review in Brussels. Chris built on earlier links with Prof. David Abramson (Monash University), and joint proposals have been submitted within national research and infrastructure funding contexts. New links among the AAF, AARNet, APAC/NCI, NCRIS Platforms for Collaboration, and national e-research capability 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 do not list papers that at the end of December 2008 were available only electronically in advance of print publication, or were in press, under review, submitted or in preparation.

A. BOOKS

A1. BOOK – AUTHORED RESEARCH


A3. BOOK – EDITED


B. BOOK CHAPTERS


C. JOURNAL ARTICLES

C1. Journal articles in scholarly refereed publications


Appendix 1: Publications


Teirigos, A. & Rigoutsos, I. (2008) Human and mouse introns are linked to the same processes and functions through each genome most frequent non-conserved motifs. Nucleic Acids Research 36:3484-3493.


C2. Journal article – other contribution to refereed journal


C3. Journal article – non-referred article

C5. Unpublished reports (including commercial consultancies)


E. CONFERENCE PROCEEDINGS


Appendix 1: Publications


Appendix 1: Publications


Appendix 1: Publications

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


This listing does not include presentations at the 2008 Winter School in Mathematics and Computational Biology.


E4i.Conference – edited volume of conference proceedings


G. COMPUTER SOFTWARE


OTHER ACADEMIC OUTPUT

Demos

Newsletter

Posters

Appendix 1: Publications


Appendix 1: Publications


**Postgraduate theses**


GIW-2008 Roundtable Workshops

Round table 1
Monday 1 December 2008

Forum on innovation and commercialisation in genome-scale bioinformatics

Chair/Moderator:
Dr Mervyn Thomas, CEO, Emphron Informatics Pty Ltd, Australia

Expert Panel:
Mr Jeremy Barker, CEO, Queensland Facility for Advanced Bioinformatics (QFAB)
Dr Jill Gray, Manager, Technology & Commercialisation, Department of Tourism, Regional Development & Innovation, Queensland Government
Professor Ming Li, Founder, Bioinformatics Solutions Ltd (Canada)
Professor Limsoon Wong, Founder, GeneticXchange (USA) and Molecular Connections (India)

Bioinformatics grew out of the early genome-sequencing projects, and has now developed into an enabling platform essential for life science research. Advances in technology make it possible to deeper questions, but also present challenges arising from the large volumes of data that must be captured, managed, integrated, analysed and visualised. Addressing these challenges demand innovative bioinformatics and information technology solutions. Organisations are tackling these challenges in a variety of ways, and have applied a range of commercialisation models with varying degrees of short- and long-term success. This forum on innovation and commercialization in bioinformatics brought together representatives from industry, government and academia with experience in the commercialisation of bioinformatics to share their thoughts and experience, and to answer questions from conference delegates.

The panellists first shared with GIW delegates their own diverse experiences with bioinformatics commercialisation and innovation. Mervyn, Limsoon and Ming described their experiences with commercialisation of bioinformatics research and services and various start-up companies, highlighting commercial strategies that were or were not successful. Jill then presented Queensland Government policies supporting innovation and commercialisation, and Jeremy concluded the opening addresses with a description of his key findings from a Churchill Fellowship-supported international study tour of bioinformatics services provision, and the QFAB experience.

A common theme from these presentations involved the commercial challenges faced in the commercial sector for bioinformatics products, including the nature of the product (e.g. software, data or hardware) or service (e.g. data curation or provision of skills-based solutions) to be commercialised. A strength highlighted by the discussions was the sustainability provided by commercial bioinformatics solutions: whereas open-source or in-house bioinformatics can be derailed by loss of key personnel, commercial services can guarantee continuity and quality. It was also noted, however, that open-source solutions can be adequate for some intended uses, and companies have successfully commercialised software in this arena.

The panellists presented a variety of models for bioinformatics commercialisation which had met varying degrees of success. Limsoon contrasted pricing strategies actually used by two different companies: in one, a full product was packaged with a single, expensive price, whereas the second was incremental, with an entry-level price for a basic product and many small increases for additional features; the latter was more successful. Mervyn highlighted the importance of offering cutting-edge products and services in a field such as bioinformatics that changes so rapidly. A skill or technique may be “cutting-edge today, but in six months a commodity”; it is, therefore, imperative for providers to stay at the crest of the wave in order to survive. Ming characterised his company’s strategy as providing cutting-edge services to the more-discriming sector of the market.

Discussions initiated by questions from the floor covered difficulties in assessing the value of research and knowledge-based products to justify government investment in innovation and commercialisation; the importance of preserving commercial value through intellectual property management; managing the potentially competing interests of science and commercialisation; the difficulties in determining an appropriate business model for bioinformatics; and the important role of governments in linking R&D with domestic and international markets. The animated discussion around these and other topics of interest reflect the level of interest in innovation and commercialisation in bioinformatics across a broad sector of the Australian and international community.

The overarching conclusion was that each commercialisation strategy demands a careful assessment of the presented opportunity; no single solution is the absolute answer in every instance, and opportunities can appear, evolve and disappear on a dynamic basis. Astute business judgment determines the blend of nature of the products and/or services that can successfully be commercialised in each case.

The workshop was very well-attended, with no shortage of questions and comments from the audience. Informal discussion continued well after the close of the session.
Round table 2
Tuesday 2 December 2008
Forum on the development of genome-scale bioinformatics in China and India
Chair/Moderator:
Professor Shoba Ranganathan, Macquarie University (Australia)
Participants:
Professor Ashok Kolaskar (National Knowledge Commission and University of Pune, India)
Professor Ming Li (University of Waterloo, Canada)
Professor Yi-xue Li (Shanghai Jiaotong University and Academia Sinica, China)
Professor Shoba Ranganathan (Macquarie University, Australia)

Genome-scale biology is generating unprecedented quantities of data relevant to most fundamental and applied areas of the life sciences including agriculture, fisheries and forestry, ecological sciences and biodiversity, human health and disease, and molecular and cellular biotechnology. Existing and new technologies for automated DNA sequencing, gene expression analysis, proteomics, systems analysis, imaging and high-content screening are widely available, adding locally produced data sources to the wealth of information available on-line from international data facilities such as EBI or NCBI. The quantity and complexity of these data requires researchers, institutions and countries to apply information technologies of corresponding scale and sophistication. Bioinformatics encompasses the management and analysis of these data, and development and provision of software tools, and immediately proximate areas of mathematics, statistics and algorithmics.

In Asia, the economically developing countries China and India have been deeply involved in bioinformatics for several decades. With the emergence of these countries as economic superpowers, this is an opportune moment to discuss where these countries are currently poised; what challenges they face; what strategies they have developed or are developing in research, development, infrastructure, education and training for bioinformatics; what the prognosis might be for their future development in this important area; what opportunities are available for international coordination and collaboration; and not least, what broader lessons can be learnt from their development and success in genome informatics.

“Rising bioinformatic stars”

China
The article “Bioinformatics in China: a personal perspective” by Liping Wei and Jun Yu (PLoS Computational Biology 4(4): e1000020, April 2008) is absolutely essential reading in this context. The authors present statistics on the growth of publications, and information on online databases, web tools and software developed and maintained in China and on bioinformatics training programs offered by institutions in China.

Specific issues discussed in this workshop included:
• The recruitment of overseas Chinese back to China
• Diversity and flexibility of arrangements, e.g. appointment as guest professors
• Wide range of salaries, start-up packages across institutions
• The possibility of “reverse culture shock”
• R&D investment plans
• Multiple national Ministries and bodies are involved
• Ministry of Science & Technology: 2001-2010 US$3.8B in biological & biomedical research
• Phase I (2001-05): main fields are bioengineering, gene manipulation, Bioinformation technologies, biomedical technologies
• Phase II (2006-10): main fields are industrial technologies, gene manipulation, biomedical technologies, bioinformation & biocomputing technologies(*)
• New national programs (November 2008):
  • New drugs (US$ 1B) – note that China doesn’t have a large pharmaceutical industry, although international pharmaceutical companies have recently located R&D facilities in China and graduates are finding jobs there.
  • Infectious disease (US$ 800M)
  • 3%-5% of these funds set aside for bioinformatics
  • Funds must be spent in next two years
• Heavy investment nationally in data production (genome sequencing, metagenomics, microbial genomics, functional genomics, proteomics, metabolomics).

Job prospects for bioinformatics graduates in China are good. Some graduates go overseas, e.g. for further training. Many others are finding jobs within China.
India

Much information is available at the following two websites:

http://www.britisnet.gov.in
(Biotechnology Information System Network)

http://bioinfo.ernet.in
(Bioinformatics Centre at University of Pune)

Bioinformatics in India started in 1981
Bioinformatics task force – 1983

• Recommends a distributed network approach: BioGrid
• Triggered internet connectivity for scientific research across India
• More recently, development of a three-tiered system:
  • Centres of Excellence (6 nationally)
  • Distributed information centres (10) and sub-centres (51)
  • Infrastructure facilities (76)

Main components of national strategy (for further detail, see below):

• Advanced research
• Development of human resources
• Academia-industry interface
• International linkages
• Advanced Research
  • About 1050 bioinformatics research papers, plus >3000 that use bioinformatics
• Development of human resources
  • Goal is to train >2000 bioinformaticians per year at all levels
• Central government support for 7 undergraduate, 28 postgraduate, 12 certificate / diploma, 9 PhD and 9 other training programs; many other programs funded by private institutions. Quality, especially of the latter, is variable. Hence BINC, Bioinformatics National Certification Examination. The BINC is open to anyone, including self-taught candidates, but is (intentionally) very difficult and the pass rate is “very, very low”. The BINC is implemented in three steps: short-answer questions, long-answer questions, and a practical; >50% correct answers is needed to proceed to each successive step. Website: http://bioinfo.ernet.in/BINC

• Discussions are underway with Malaysia about adopting the BINC there.
• BioGrid as a “virtual classroom”
• Market is (so far) imperfect in matching graduates to jobs. There is difficulty in finding qualified teachers. Some graduates have trouble finding jobs, although these are mostly from the programs of questionable quality. There was much discussion in the workshop about how to match graduates to jobs lies; Professor Kolaskar advised that the principal approach must be through improvement in quality.
• Academia-industry interface
  • Since 1982, central government support of the bioinformatics industry has been about US$ 25M in total, of which >$5M is being spent this year.
  • There is a very active ICT industry in India, including large and small companies in bioinformatics (e.g. Molecular Connections, mentioned in the Innovation and Commercialisation forum).
  • Note that there is a large drug industry in India. This has attracted bioinformatics and software companies from overseas, e.g. Accelrys, to invest in India.
  • International linkages
    • The main partners in bioinformatics at this point are Japan, Australia, and others in the Asia-Pacific region; and Europe, particularly Italy.

IBIN – Indian Bioresource Information Network

• In three regional languages
• Pictorial, common-language focus to facilitate broad public impact
• Development of databases in highly relevant: rice, tuberculosis

The Indian government and institutions have noted China’s success in attracting overseas Chinese back to China, and are considering how to emulate this vis-à-vis overseas Indians, whether for permanent relocation or on a part-time basis. One possibility being discussed is that overseas Indians might be made eligible to apply for government research grants.
## Appendix 2: Financial statement

### INCOME

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance carry forward</td>
<td>440,763</td>
</tr>
<tr>
<td>ARC</td>
<td>1,144,719</td>
</tr>
<tr>
<td>Host institution support</td>
<td>525,000</td>
</tr>
<tr>
<td><strong>TOTAL INCOME</strong></td>
<td><strong>2,110,482</strong></td>
</tr>
</tbody>
</table>

### EXPENDITURE

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaries</td>
<td>1,215,962</td>
</tr>
<tr>
<td>Scholarships</td>
<td>20,325</td>
</tr>
<tr>
<td>Student programming/research experience/workshops</td>
<td></td>
</tr>
<tr>
<td>Interns</td>
<td>6,320</td>
</tr>
<tr>
<td>Workshops/conferences</td>
<td>23,571</td>
</tr>
<tr>
<td>Strategic visitors/sponsorship program/director discretionary</td>
<td>15,047</td>
</tr>
<tr>
<td>Equipment</td>
<td>14,016</td>
</tr>
<tr>
<td>Accommodation</td>
<td>2,730</td>
</tr>
<tr>
<td>Travel</td>
<td>1,682</td>
</tr>
<tr>
<td>Maintenance/consumables</td>
<td></td>
</tr>
<tr>
<td>Annual report</td>
<td>14,575</td>
</tr>
<tr>
<td>Computer consumables/maintenance</td>
<td>738</td>
</tr>
<tr>
<td>Laboratory consumables</td>
<td>12,268</td>
</tr>
<tr>
<td>Photography/slides/posters</td>
<td>315</td>
</tr>
<tr>
<td>Postage &amp; freight</td>
<td>1,076</td>
</tr>
<tr>
<td>Stationery</td>
<td>304</td>
</tr>
<tr>
<td>Transport - local &amp; overseas</td>
<td>2,787</td>
</tr>
<tr>
<td>Meetings - Advisory Board, Research Programs, Annual All Hands</td>
<td>2,206</td>
</tr>
<tr>
<td>Printing - books/journals/reprints</td>
<td>4,422</td>
</tr>
<tr>
<td>Relocation</td>
<td>3,483</td>
</tr>
<tr>
<td>Recruitment</td>
<td>207</td>
</tr>
<tr>
<td>Telephone &amp; fax</td>
<td>251</td>
</tr>
<tr>
<td>Registration: trademark, domain name</td>
<td>512</td>
</tr>
<tr>
<td>Other (bank charges, payroll tax, FBT, cost recoveries/journal adjustments etc.)</td>
<td>-17,083</td>
</tr>
<tr>
<td><strong>TOTAL EXPENDITURE</strong></td>
<td><strong>1,325,944</strong></td>
</tr>
</tbody>
</table>

### BALANCE

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance</td>
<td>784,539</td>
</tr>
</tbody>
</table>
Acknowledgements:
Cover image courtesy of Mr Oliver Cairncross, ARC Centre of Excellence in Bioinformatics, Dr Brad Marsh and Mr Andrew Noske, Institute for Molecular Bioscience, The University of Queensland.

Other images courtesy of Ms Denis Bauer, Mr Fabian Buske and Dr Timothy Bailey (page 25); Dr Nicholas Hamilton (page 21), and Dr Brad Marsh and Mr Andrew Noske (page 23), Institute for Molecular Bioscience, The University of Queensland.
