Program 3 – The Visible Cell
Data integration and visualisation
32
3.1 Data integration
32
3.1.1 Advanced data management in bioinformatics
32
3.1.2 Classification of genomic data
34
3.1.3 Computational identification of alternatively spliced transcripts in mouse and human, and construction of an alternative splicing database
35
3.2 Visualisation
36
3.2.1 Concept and engineering of The Visible Cell
36
3.2.2 Modelling and visualisation of biological processes in complex spatial environments
38
3.3 Bioinformatics Web Services
39

Key Performance Indicators
40
1. Research Findings
41
Quality of publications
41
Number of publications
41
Number of patents
41
Invitations to address and participate in conferences
42
Invitations to visit leading international research groups and research centres
44
Number and nature of commentaries about the Centre’s achievements
45
Evidence of interdisciplinary linkages within the Centre
46
2. Research Training and Professional Education
46
Recruitment
46
International Exchange Program
46
Number of postgraduates recruited
46
Number of postgraduate completions
47
Number of honours students
47
Number of professional courses
47
Participation in professional courses
47
Number of undergraduate courses in the Centre’s area
47
3. International, national and regional links and networks
47
International visitors to one or more ACB nodes
48
National visitors to one or more ACB nodes
49
National and international workshops
49
Visits to overseas research centres
51
4. End-user links
51
Number and nature of commercialisation activities
51
Number of government, industry and business briefings
52
Number of Centre associates trained or training in technology transfer and commercialisation
52
Number and nature of public awareness programs
52
5. Organisational support
54
Student industrial scholarship
54
Annual cash contributions: collaborating organisations
55
Annual in-kind contributions: collaborating organisations
55
Number of new organisations recruited to or involved in the Centre
56
Level and quality of infrastructure
56
6. National benefit
56
Measures of expansion of Australia’s capability in the priority area(s)
56
Case studies of economic, social, cultural or environmental benefits
56

Appendix 1: Publications and Papers
57
Appendix 2: Financial Statement
64
Front row from left: Dr Jennifer Hallinan, Prof. Geoff McLachlan, Dr Greg Smith (Chair, Advisory Board), Prof. Mark Ragan (Director, ARC Centre in Bioinformatics).
Back row from left: Dr Lindsay Hood (Manager, Visible Cell Project), Dr Sean Grimmond, Prof. John Mattick (Advisory Board), Prof. Kevin Burridge and Dr Rohan Teasdale.
This third annual report presents a view of the ARC Centre in Bioinformatics (ACB), now at its full complement of personnel and research projects.

I am very pleased to report that, in 2005, the ACB demonstrated very solid progress in terms of its research outcomes. For example, these results included 7 very high impact publications and another 74 peer reviewed articles.

The team continues to be a highly collaborative one and involves a set of integrated research activities between the Universities of Queensland, ANU, Deakin and Newcastle, together with a large number of international collaborations. (Please note that a detailed commentary on the Centre’s achievements can be found on page 40 of this report.)

In addition to its national bioinformatics research role, the ACB provides a program of advanced education and training, and, during 2005, it hosted another five-day Winter School in Mathematical and Computational Biology. In 2006, the ACB anticipates continuing its strong educational and training support role in bioinformatics.

While ACB Centre members were all somewhat disappointed not to be selected during the recent ARC Centres of Excellence round, the ACB nonetheless continues to build on the strengths and commitment of its researchers across the numerous disciplines required to undertake effective bioinformatics research. These skills sets include strong bases in mathematics, statistics, computer science, information technology and the molecular biosciences.

As Bioinformatics Australia continues to define its role in 2006, and with important bioinformatics initiatives continuing within the ACB, the group envisages the evolution of a new national bioinformatics strategy as a key outcome for the calendar year. Concurrently, the ACB anticipates an ARC review of the Centre’s research and future plans later in the year.

The ARC Centre for Bioinformatics continues as a strong and developing centre for bioinformatics in Australia. With its many interdisciplinary linkages, both across Australia and internationally, its linkages to Australian and overseas bioinformatics end users, and its expanding relationships across the Australian biotechnology sector, I believe that the ACB is continuing to provide and grow a critical role for the Australian research infrastructure. I do hope that you will agree with this view as you review the ACB’s annual report.

Dr Gregory R. Smith
Chair
March 2006
Most of our projects approached full productivity in 2005, with our investigators publishing 74 peer-reviewed journal articles and conference papers, seven of these in journals with double-digit impact factors (Genome Research, Nature Biotechnology, Nature Reviews Molecular Cell Biology, Proceedings of the National Academy of Sciences USA, and Science). Initial indications are that 2006 will be at least as productive.

ACB’s mission includes advanced education and training, and in July 2005 we again hosted a five-day Winter School in Mathematical and Computational Biology that attracted 169 participants, more than half of them postgraduate students from 34 Australian, New Zealand and other overseas universities and research institutions. Our investigators again contributed to BioInfoSummer at Australian National University, and we co-sponsored the first Bioinformatics Australia workshop, in conjunction with ComBio in Adelaide. We held our 2005 All-Hands Meeting jointly with researchers and students from Bioinformatics Institute New Zealand, and coordinated it with a microarray analysis workshop sponsored by Bioinformatics Australia and Queensland Bioinformatics Consortium.

The year ahead will be eventful, and will potentially be formative for bioinformatics research in Australia. The new Bioinformatics Australia will continue to define its role, and important initiatives are underway involving major collaborative research infrastructure, a new national bioinformatics strategy and coordinating committee, medical and Grid bioinformatics, and research and infrastructure proposals in at least three States. Closer to home, in October the ARC will review our Centre’s research, outcomes and plans for coming years.

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

Professor Mark Ragan
Director, ARC Centre in Bioinformatics
March 2006
Organisational Structure

The organisational chart shows the organisational structure of the ARC Centre in Bioinformatics (ACB) in 2005.

**ARC Centre in Bioinformatics Organisational Chart**
The 2005 Advisory Board meeting was held on Friday 14 October 2005. As of 31 December 2005, the members of Advisory Board were:

1. **Dr Gregory R. Smith** (Chair)
   Director
   SciVentures Investments Pty Ltd
   Melbourne VIC, Australia
   Dr Smith is a co-founder and director of SciVentures Investments Pty Ltd, the manager of the SciVentures Pre-Seed Fund. He also is a director of five start-up companies into which SciVentures has invested. Dr Smith has been a member of the Commonwealth Government’s Industrial Research & Development Board and of its Advisory Council for Intellectual Property. He continues as Board Chair for the NANO Major National Research Facility.

2. **Prof. Ross Coppel**
   Director
   Victorian Bioinformatics Consortium
   and Department of Microbiology
   Monash University
   Melbourne VIC, Australia
   Prof. 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. He is Professor of Microbiology at Monash University, Director of the Victorian Bioinformatics Consortium and co-founder of the biotechnology company Glykoz.

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

4. **Prof. John Mattick AO**
   Director
   Institute for Molecular Bioscience
   The University of Queensland
   Brisbane QLD, Australia
   Prof. 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, and Foundation Director of the Australian Genome Research Facility from 1996-2002.
   Prof. Mattick’s main research interest is the role of noncoding RNA in the evolution and development of humans and other complex organisms. He has been awarded the 1989 Pharmacia-LKB Biotechnology Medal by the Australian Biochemical Society, an Honorary Fellowship of the Royal College of Pathologists of Australasia (2002) and the Centenary Medal by the Australian Government (2003). He was appointed an Officer in the Order of Australia (AO) in 2001 for service to scientific research in the fields of molecular biology, genetics and biotechnology. Professor Mattick serves on the advisory boards of several institutes in Australia and abroad.
Prof. Mark Ragan FLS  
Director, ARC Centre in Bioinformatics  
(ex officio)  
Institute for Molecular Bioscience  
The University of Queensland  
Brisbane QLD, Australia

Prof. Ragan joined IMB in 2000 after 28 years with National Research Council Canada, where he co-founded and developed programs in bioactive compounds, molecular biology, genomics and bioinformatics, including Canadian Bioinformatics Resource.

His research record includes many “firsts” for his Institute and for NRC, including first use of PCR, first Unix workstation, and first 100 Kbp and 1 Mbp of DNA sequence. With Prof. W. Ford Doolittle and Dr Robert Charlebois he received the largest single grant under the first Canadian genome program (CGAT) and sequenced Sulfolobus solfataricus.

He has been President of two international scientific societies, and a Fellow of the Canadian Institute for Advanced Research. He represents Australia on the executive of the International Conference on Bioinformatics and the Association of Asian Societies of Bioinformatics.

Prof. David Siddle  
Deputy Vice-Chancellor (Research)  
The University of Queensland  
Brisbane QLD, Australia

Prof. Siddle became Deputy Vice-Chancellor (Research) from 2002 following his September 2001 appointment as the University’s Pro-Vice-Chancellor (Research). Previously he was Pro-Vice-Chancellor (Research) at the University of Sydney (1997-2001), and Dean, Postgraduate Studies at the University of Queensland (1993-1997). He is responsible for all aspects of the development of the University’s profile in research and research training.

Prof. Siddle is also a Director of four Cooperative Research Centres, the Queensland Parallel Supercomputing Foundation, the Australian Canopy Crane Research Facility, and the Sustainable Minerals Institute.

Prof. Lap-Chee Tsui  
Vice-Chancellor  
Hong Kong University  
Hong Kong

Prof. Tsui received international acclaim in 1989 when he identified the defective gene (Cystic Fibrosis Transmembrane Regulator, CFTR) that causes cystic fibrosis. His discovery is seen as one of the key breakthroughs in the genetics of human disease.

Prof. Tsui has received numerous awards and honours, including Distinguished Scientist of the Medical Research Council of Canada and the Killam Prize, and is a Fellow of the Royal Society of Canada, a Fellow of the Royal Society of London, an Honorary Member of World Innovation Foundation, a Foreign Associate of the National Academy of Sciences USA and an Honorary Fellow of the Royal College of Physicians of London. He has been awarded the Order of Canada (Officer), and the Order of Ontario.

Prof. Marc Wilkins  
Department of Biotechnology and Biomolecular Sciences  
University of New South Wales  
Sydney NSW, Australia

Prof. Wilkins developed the concept of the proteome and coined the term. He has 70+ peer-reviewed research and review publications and holds patents pending. He is a Senior Editor of Proteomics, and has served as a Member of the Australian Government’s Expert Task Force for Bioinformatics, a National Task Force to advise the Australian Government on the current and future bioinformatics requirements. In 1999, Prof. Wilkins co-founded the biotechnology company Proteome Systems, where he served as the Vice President of Bioinformatics (1999-2004) and Head of Proteomics (2004-2005). This company is Australia’s premier proteomics company, employs 60+ staff and is listed on the Australian Stock Exchange. Prof. Wilkins’s research interests are proteomics, bioinformatics and their application to biomedical research.
Members of the ARC Centre in Bioinformatics Research Committee for 2005 were:

**RESEARCH COMMITTEE**

Chair – Prof. Mark Ragan
Advisory Board Member – Prof. John Mattick
Program 1 – Dr Rohan Teasdale
Program 1 – Dr Sean Grimmond
Program 2 – Prof. Kevin Burrage
Program 2 – Dr Jennifer Hallinan 
(Sep 2004 – June 2005)
Program 3 – Dr Lindsay Hood  
(Adjunct member)
Program 3 – Prof. Xiaofang Zhou
Program 3 – Dr Brad Marsh  
(Adjunct member)
Secretary – Ms Lanna Wong

In 2005, the Chair invited Dr Brad Marsh and Dr Lindsay Hood to become adjunct members to the Research Committee to facilitate management of the Visible Cell program.

1 **Dr Brad Marsh**
   Senior Research Fellow/Group Leader
   Beta Cell Structural Biology/Cellular Tomography Group
   Institute for Molecular Bioscience
   The University of Queensland
   Brisbane QLD, Australia

   Dr Marsh is a Group Leader at the Institute for Molecular Bioscience at the University of Queensland. His research is focused on understanding how the crucial metabolic hormone insulin is processed, trafficked and released from the pancreatic beta cell. This work will identify how and where defects occur at the subcellular level, and how they contribute to the onset of the diseases commonly known as type 1 and type 2 diabetes.

2 **Dr Lindsay Hood**
   High Performance Computing Manager
   Institute for Molecular Bioscience
   The University of Queensland
   Brisbane QLD, Australia

   Dr Hood has many years’ experience with high performance computing in academia, government service and the private sector, including with the legendary Thinking Machines Inc. (Cambridge, MA). Prior to joining IMB, he was a member of Compaq’s global life and material sciences team, which was a significant part of the human genome project at the Sanger Centre and at Celera. Dr Hood is responsible for high-performance computing resources at IMB and for ACB, and is the technical lead for the Visible Cell Project.

Dr Marsh is recognized internationally in the diabetes, membrane traffic and 3D EM research communities for his cutting-edge application of large volume EM tomography to visualise mammalian cell structure in 3D at high resolution. His striking images of subcellular architecture in the insulin-secreting pancreatic beta cell have been featured on the covers of *Proceedings of the National Academy of Sciences USA, Proceedings of Microscopy and Microanalysis*, and *Traffic*. His work and methods have been highlighted in *La Recherche, Nature Reviews Molecular Cell Biology* and the National Institutes of Health’s highly esteemed biennial publication *NCRR Highlights*, as well as the latest editions of *Molecular Cell Biology* (Lodish et al.) and *Essential Cell Biology* (Alberts et al.).
**Director**

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

**Chief Investigators**

2. Dr Timothy L. Bailey  
   Advanced Computational Modelling Centre, and Institute for Molecular Bioscience, The University of Queensland  
   
   Dr Bailey’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 developing and supporting web-based algorithms for bioinformatics. He wrote MAST, MEME and MetaMINE, the standard motif-recognition programs in bioinformatics. His MEME website was used by approximately 800 researchers per month in 2005, nearly double the 2004 usage.

3. Prof. Kevin Burrage  
   ARC Federation Fellow  
   Advanced Computational Modelling Centre; Department of Mathematics; School of Information Technology & Electrical Engineering; and Institute for Molecular Bioscience, The University of Queensland  
   
   Prof. Burrage is founding CEO of the Queensland Parallel Supercomputer Foundation (QPSF), the Advanced Computational Modelling Centre (ACMC) at UQ, and the ViSAC visualisation laboratory. He was awarded a prestigious Federation Fellowship by the Australian Research Council in 2003. He has co-authored over 150 papers in the fields of computational science, computational biology, mathematical modelling and complex systems. This oeuvre consists of fundamental work on the numerical solution of both ordinary differential equations and stochastic differential equations, and algorithms for linear systems. He is author of a monograph on parallel and sequential methods for ordinary differential equations.

4. A/Prof. Yi-Ping Phoebe Chen  
   School of Information Technology  
   Deakin University  
   
   Associate Professor (Reader) Yi-Ping Phoebe Chen is Director of Research at the School of Information Technology, Deakin University and head of research groups in bioinformatics and in multimedia. She is Chair of the Steering Committee of Asia-Pacific Bioinformatics Conference, and of Multimedia Modelling. Her research is in bioinformatics, multimedia databases and technology, visual querying, web information systems, machine learning and data mining. A/Prof. Chen has more than 80 publications in bioinformatics, advanced databases and visual query languages.

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

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

Dr Hallinan has academic background in both molecular biology and computer science. Her research centers around databases, data mining and analysis, and complex systems modelling of problems in genomic and proteomic systems biology. She was the 2002 Steinmetz Fellow of the Santa Fe Institute in recognition of her work in complex protein-interaction networks. Dr Hallinan received a UQ Early Career Researcher grant in 2003 to undertake a project on the structure and dynamics of the genetic regulatory network in human cells involving the tumour suppressor gene p53.

In November 2005 Dr Hallinan took up an academic appointment with the Institute for Ageing and Health with a joint appointment in the School of Computing Sciences, University of Newcastle upon Tyne in the UK.

8 Dr Markus Hegland  
Centre for Mathematics and its Applications  
The Australian National University

Dr Hegland is coordinator of Advanced Computation and Modelling at ANU's Centre for Mathematics and its Applications. He is also a member of the Statistical Machine Learning Group of NICTA, an Associate Editor of Computational Statistics and Data Analysis and member of the board of the ANZIAM computational mathematics group. Dr Hegland’s expertise is in high-dimensional problems, in particular the solution of master equations and in numerical techniques for predictive modelling and machine learning.

9 Prof. Geoffrey J. McLachlan  
Department of Mathematics  
The University of Queensland

Prof. McLachlan's research in statistics is in the related fields of classification, cluster and discriminant analyses, 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, he 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.

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

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

11 Prof. Bernard A. Pailthorpe  
School of Physical Sciences  
The University of Queensland

Prof. Pailthorpe was founding Director of Sydney VisLab, and since 2003 has been CEO of Queensland Parallel Supercomputing Foundation, the state partner within the Australian Partnership in Advanced Computing. Until 2000, Prof. Pailthorpe was Director for Visualization at the San Diego Supercomputer Center, and Associate Director for Interaction Environments for the US National Partnership in Advanced Computing Infrastructure. He is building computational and data infrastructure in support of the Centre.
Dr Rohan D. Teasdale  
Institute for Molecular Bioscience  
The University of Queensland

Dr Teasdale holds an R.D. Wright Career Development Award. He heads a multidisciplinary research group using both cellular and computational techniques to investigate how subcellular compartments are generated. He is active in the FANTOM mouse transcriptome annotation consortium, and played a key role in the formation of Bioinformatics Australia.

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

Prof. Zhou is Research Director of the ARC Research Network in Enterprise Information Infrastructure (EII), and Chair of the Steering Committee of the Asia-Pacific WEB Conferences. Until 1999 he was Principal Researcher and Head, CSIRO Spatial Information Systems. Prof. Zhou has more than 80 publications in spatial databases and information systems, Grid computing and bioinformatics.

Dr Isidore Rigoutsos  
IBM Thomas J. Watson Research Center

Dr Rigoutsos is Manager of the Bioinformatics and Pattern Discovery group at IBM’s Thomas J. Watson Research Center, and a lecturer in the Department of Chemical Engineering at the Massachusetts Institute of Technology. He was 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 journals in bioscience and bioinformatics, and on the advisory board for the Master's program in Bioinformatics at Oxford University. For almost 20 years, Dr Rigoutsos has studied invariant methods for representing one-, two- and three-dimensional signals. Since 1996, his efforts have revolved around theoretical and practical aspects of pattern discovery, and the design of pattern-based solutions to problems in computational biology including drug design, gene discovery, and automated protein annotation. His most recent work has focussed on lateral genetic transfer, discovery and design of antimicrobial peptides, computational identification of microRNA targets and novel microRNA precursors, and application of RNA interference to regenerative medicine and therapy. He is editing a two-volume book on systems biology (Oxford University Press).

Dr Prasad Kodali  
IBM Healthcare & Life Sciences

Dr Kodali is currently the Americas Sales Leader for IBM Software Solutions for life sciences. He was previously Manager of the Discovery Solutions development team within IBM Life Sciences, focussed on developing and delivering information solutions to R&D communities.

Dr Kodali received his PhD in computational chemistry from The Pennsylvania State University, and an MSc in organic chemistry from Indian Institute of Technology, Bombay. He has held a variety of technical and management positions spanning research, technical development and product management. His experience ranges from creating domain-specific solutions, to providing data integration solutions to life science informatics companies. His primary areas of interest are life science informatics, molecular dynamics simulations, computer-aided drug discovery, and data integration.
Strategic Partnership

ACB has formed a close research partnership with the Newcastle Bioinformatics Initiative, particularly involving the design and application algorithms for large-scale network and graph optimisation. NBI is headed by Dr Pablo Moscato.

1 Prof. Gene Myers
Departments of Computer Science, and Molecular and Cell Biology
University of California at Berkeley

Prof. Myers is Professor of Computer Science and of Molecular Biology at University of California at Berkeley, and Honorary Professor in the Institute for Molecular Bioscience at The University of Queensland. Prof. Myers has made fundamental contributions to computational molecular biology and bioinformatics, notably the BLAST algorithm (1990) and whole-genome assembly (1999). He designed and led the computer-based assembly of large genome sequences including fruit fly, human and malaria mosquito. In 2004 Prof. Myers received the Max Planck Research Prize for International Cooperation, and in 2005 was named as one of the initial seven Group Leaders at the Janelia Farm Research Campus of Howard Hughes Medical Institute.

2 Dr Pablo Moscato
Department of Computer Science
The University of Newcastle

Dr Moscato is founding director of the Newcastle Bioinformatics Initiative, a partner group of ACB. He has worked in evolutionary computation since 1988, and in metaheuristic methods for combinatorial optimisation since 1985. A former member of the Caltech Concurrent Computation Program, he introduced memetic algorithms in 1989 and has successfully applied these hybrid methodologies to large-scale optimisation problems. Dr Moscato has held academic and research positions in the Argentina, Brazil, the USA and Australia, and is currently at the School of Electrical Engineering and Computer Science, The University of Newcastle (Australia). Dr Moscato is member of the editorial board of Journal of Mathematical Modelling and Algorithms, has served on the program committee of many international conferences in heuristics and optimisation.

Special Advisor

1 Prof. Gene Myers
Departments of Computer Science, and Molecular and Cell Biology
University of California at Berkeley

Prof. Myers is Professor of Computer Science and of Molecular Biology at University of California at Berkeley, and Honorary Professor in the Institute for Molecular Bioscience at The University of Queensland. Prof. Myers has made fundamental contributions to computational molecular biology and bioinformatics, notably the BLAST algorithm (1990) and whole-genome assembly (1999). He designed and led the computer-based assembly of large genome sequences including fruit fly, human and malaria mosquito. In 2004 Prof. Myers received the Max Planck Research Prize for International Cooperation, and in 2005 was named as one of the initial seven Group Leaders at the Janelia Farm Research Campus of Howard Hughes Medical Institute.

2 Dr Pablo Moscato
Department of Computer Science
The University of Newcastle

Dr Moscato is founding director of the Newcastle Bioinformatics Initiative, a partner group of ACB. He has worked in evolutionary computation since 1988, and in metaheuristic methods for combinatorial optimisation since 1985. A former member of the Caltech Concurrent Computation Program, he introduced memetic algorithms in 1989 and has successfully applied these hybrid methodologies to large-scale optimisation problems. Dr Moscato has held academic and research positions in the Argentina, Brazil, the USA and Australia, and is currently at the School of Electrical Engineering and Computer Science, The University of Newcastle (Australia). Dr Moscato is member of the editorial board of Journal of Mathematical Modelling and Algorithms, has served on the program committee of many international conferences in heuristics and optimisation.
**INFORMATION INFRASTRUCTURE**

Mr Matthew Bryant  
System Administrator,  
High-performance computing

Mr Oliver Cairncross  
Database Administrator/Developer

Dr Lynn Fink  
Research Website Administrator

Mr Igor Kromin  
Software Engineer, Visible Cell Project

Ms Mhairi Marshall  
Grid & Database Developer

Mr Ken Steube  
Web Services Developer

**CENTRE MANAGER**

Ms Lanna Wong, MBA

**POSTDOCTORAL FELLOWS / RESEARCH OFFICERS**

Dr Jiuyan An  
Dr Richard Bean  
Dr Robert Beiko  
Dr Francis Clark  
Dr Lynn Fink  
Dr Martin Frith (jointly with RIKEN Genome Sciences Center, Yokohama)  
Dr Donald Gardiner  
Dr Nicholas Hamilton  
Dr Jim Hanan  
Dr Liat Jones  
Dr Pritha Mahata  
Dr Elena Prieto  
Dr Lucia Santoso  
Dr Tianhai Tian  
Dr Simon Wong  
Dr Zheng Yuan  

**RESEARCH ASSISTANTS**

Mr Daniel Bradley  
Mr Geoffrey Faulkner  
Ms Kelly Hanson  
Mr Robert McLeay  
Mr Radosav Pantelic  
Mrs Chikako Ragan  
Mr Dave Tang  
Mr Shane Zhang

**VISITING SCHOLARS**

Prof. Frank Dehne  
Griffith University, and  
Carleton University, Canada

Dr Haijun Wu  
Nanjing University, China

Mr Geir Kjetil Sandve  
Norwegian University of Science and Technology

Prof. Przemyslaw Prusinkiewicz  
University of Calgary, Canada

**MASTERS/PHD STUDENTS**

*Supervision/Associate supervision (principal supervisor in ACB):*

Farah Abdullah  
Supervisor: Prof. Kevin Burrage

Rajith Aturaliya  
Supervisor: Dr Rohan Teasdale

Michael Bode  
Supervisor: Prof. Kevin Burrage

Alhadi Bustamum  
Supervisor: Prof. Kevin Burrage

Simon Carter  
Supervisor: Dr Jennifer Hallinan  
Co-supervisor: Dr Paul Strooper

Cheong Xin Chan  
Supervisor: Prof. Mark Ragan  
Co-supervisor: Dr Robert Beiko

Soong Chang (PhD awarded in 2005)  
Supervisor: Prof. Geoff McLachlan

Yan Chen  
Supervisor: A/Prof. Phoebe Chen

Wagner Emanoel Costa  
Supervisor: Dr Pablo Moscato  
Co-supervisor: Dr Regina Berretta

Melissa Davis  
Supervisor: Dr Rohan Teasdale

Geoffrey Faulkner  
Supervisor: Dr Sean Grimmond

Mohamed Rafi Feroze (Jan – Aug 2005)  
Supervisor: Prof. Xiaofang Zhou  
Supervisor: Prof. Mark Ragan

Al Forrest  
Supervisor: Dr Sean Grimmond

Alex Garcia  
Supervisor: Prof. Mark Ragan  
Co-supervisor: Dr Lindsay Hood

Michael Höhl  
Supervisor: Prof. Mark Ragan  
Co-supervisor: Dr Peter Adams

Mou’ath Adeib Hourani  
Supervisor: Dr Pablo Moscato

Helen Zi Huang  
Supervisor: A/Prof. Xiaofang Zhou  
Supervisor: Prof. Peter Bruza (QUT)

Keith Knapp  
Supervisor: A/Prof. Phoebe Chen

Mario Inostroza-Ponta  
Supervisor: Dr Pablo Moscato  
Co-supervisor: Dr Regina Berretta

Dennis Lee  
Supervisor: Dr Ralf Muhlberger  
Co-supervisors: Dr Matthew Simpson and  
Dr Mark Brown

Stephen Livingstone  
Supervisor: Dr Ralf Muhlberger  
Co-supervisor: Dr Andrew Brown

Andrew Loch  
Supervisor: Dr Ralf Muhlberger  
Co-supervisor: Dr Stephen Viller
Shev MacNamara
Supervisor: Dr Markus Hegland
(untit September 2005)
Supervisor: Prof. Kevin Burrage
(from September 2005)

Katrina Monico
Supervisor: Prof. Geoff McLachlan

Ann Morrison
Supervisor: Dr Ralf Muhlberger
Co-supervisor: Dr Peta Mitchell

Dan Nicolau Jr
Supervisor: Prof. Kevin Burrage

Elena-Prieto Rodriguez
(PhD thesis approved, September 2005)
Supervisor: Dr Pablo Moscato

Tatiana Semenova
(PhD thesis submitted, 2005)
Supervisor: Dr Markus Hegland

Peter Shaw
Supervisor: Prof. Michael Fellows

Chang Jin Shin
Supervisor: Prof. Mark Ragan
Co-supervisor: Dr Simon Wong

Chinnu Subramaniam
Supervisor: A/Prof. Phoebe Chen

Dian Tjondronegoro
(PhD awarded in 2005)
Supervisor: A/Prof. Phoebe Chen

Justin Xi Zhu
Supervisors: Prof. Geoff McLachlan
Co-supervisor: Dr Angus Ng

Co-supervision (principal supervisor not an ACB investigator)
Agnes Boskovitz
Co-supervisor: Dr Markus Hegland
Principal supervisor: Dr Rajeev Goré

Elizabeth Dunn
Co-supervisor: Dr Jim Hanan
Principal supervisor: Dr Christine Beveridge

Pol Haji (PhD awarded in 2005)
Co-supervisor: Dr Markus Hegland
Principal supervisor: Dr Robert S. Anderssen

Stefan Maetschke
Co-supervisor: Prof. Geoff McLachlan
Principal supervisor: Dr Marcus Gallagher

Luke Mathieson
Co-supervisor: Dr Pablo Moscato
Principal supervisor: Dr Ljiljana Brankovic

Duncan Mortimer
Co-supervisor: Prof. Kevin Burrage
Principal supervisor: Dr Geoffrey Goodhill

Radosav Pantelic
Co-supervisor: Dr Nicholas Hamilton
Principal supervisor: Dr Ben Hankamer

Andrew Noske
Co-supervisors: Dr Ralf Muhlberger and Prof. Kevin Burrage
Principal supervisor: Dr Brad Marsh

Michael Pheasant
Co-supervisor: Dr Jennifer Hallinan
(untit November 2005)
Co-supervisor: Prof. Janet Wiles
(from November 2005)
Principal supervisor: Prof. John Mattick

Blaize Rhodes
Co-supervisor: Dr Ralf Muhlberger
Principal supervisor: Prof. Simon Kaplan (QUT)

Stefan Stanley
Co-supervisor: Prof. Mark Ragan
Principal supervisor: Prof. John Mattick

Henk Stolk (PhD thesis submitted, 2005)
Co-supervisor: Dr Jim Hanan
Principal supervisor: Dr Kevin Gates

Stuart Stephen
Co-supervisor: Prof. Mark Ragan
Principal supervisor: Prof. John Mattick

Brendan Tse
Co-supervisor: Dr Robert Beiko
Principal supervisor: Prof. David Hume

James Watson
Co-supervisors: Dr Jim Hanan and Dr Mikael Boden
Principal supervisor: Prof. Janet Wiles

Jemma Wu
Co-supervisor: Dr Markus Hegland
Principal supervisor: Dr John Lloyd

Honours students
Ms Katherine Duczmal
Agent-based virtual insects
Supervisor: Dr Jim Hanan

Mr Xiaofeng (Michael) Liu
Bio-information trust management
Supervisor: Dr Ralf Muhlberger

Mr Tien Boon Puar
Data mining genetic regulatory networks
Supervisor: Dr Jennifer Hallinan

Mr Dave Tang
Lateral genetic transfer within genus
Staphylococcus
Supervisor: Dr Robert Beiko
Co-supervisor: Prof. Mark Ragan

Ms Linda Teng
Identification of novel membrane proteins with specific subcellular localisation from functional genomics data
Supervisor: Dr Rohan Teasdale
**Research Trainees and Volunteers – 2005**

Ms Elizabeth Skippington  
Searching for lateral genetic transfer in the Sargasso Sea  
IMB Undergraduate Research Scholarship  
Supervisor: Dr Robert Beiko

Mr Benjamin Woodcroft  
Predicting inter-protein residue contacts  
IMB Undergraduate Research Scholarship  
Supervisors: Dr Nicholas Hamilton and Dr Thomas Huber

**International & National Interns – 2005**

**International**

Guillaume Angeli  
Polytechnique Universitaire de Marseille (France)  
Supervisor: Dr Pablo Moscato

Denis Baurain  
Université Liège (Belgium)  
Supervisor: Prof. Mark Ragan

Emilie Masse  
Polytechnique Universitaire de Marseille (France)  
Supervisor: Dr Pablo Moscato

Amit Mittal  
Indian Institute of Technology (India)  
Supervisor: Dr Rohan Teasdale

Josefine Sprenger  
University of Applied Sciences, Weihenstephan (Germany)  
Supervisors: Dr Rohan Teasdale and Dr Lynn Fink

**National**

Abirami Ratnakumar  
University of Sydney  
Supervisors: Prof. Mark Ragan and Dr Simon Wong

Jack Wang  
The University of Queensland  
Supervisor: Dr Rohan Teasdale

**Exchange Student – 2005**

Mr Alex Garcia  
January through November 2005 at European Bioinformatics Institute, UK  
Supervisor: Dr Susanna-Assunta Sansone

**Associates**

A/Prof. Ian Atkinson  
Information Technology  
James Cook University

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

Dr Pamela Burrage  
Department of Mathematics  
The University of Queensland

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

Prof. Gillian Bushell  
Dean, Faculty of Science  
Griffith University

Dr Robert Charlebois  
Founder, NeuroGadgets Inc.  
Ottawa

Dr Brian Dalrymple  
CSIRO Livestock Industry

Prof. Frank Dehne  
School of Computing and Information Technology  
Griffith University

Dr Ben Hankamer  
Institute for Molecular Bioscience  
The University of Queensland

Dr Yoshhide Hayashizaki  
RIKEN Genome Sciences Center  
Yokohama

Dr Lindsay Hood  
Institute for Molecular Bioscience  
The University of Queensland

Dr Thomas Huber  
Department of Mathematics  
The University of Queensland

Dr Philip Hugenholz  
Microbial Ecology Program  
DOE Joint Genome Institute, USA

Dr Jane Hunter  
DSTC Pty Ltd and  
The University of Queensland

Dr Lars Jermiin  
The University of Sydney

Dr Jonathan Keith  
Department of Mathematics  
The University of Queensland

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

Dr Brad Marsh  
Institute for Molecular Bioscience  
The University of Queensland

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

Dr Pablo Moscato  
Director, Newcastle Bioinformatics Initiative  
The University of Newcastle

Prof. Shoba Ranganathan  
Biotechnology Research Institute  
Macquarie University
Prof. Allen Rodrigo  
Bioinformatics Institute New Zealand  
Auckland

Dr Christopher Savoie  
Gene Networks International Inc.  
Cambridge, UK

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

Dr Harukazu Suzuki  
RIKEN Genomic Sciences Center  
Yokohama

Dr Mark Wilkinson  
University of British Columbia  
Vancouver

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

Annual All Hands Meeting (18 April 2005) and The Bioinformatics Institute New Zealand Workshop  
(19 April 2005) at Queensland Bioscience Precinct, The University of Queensland. From left: Dr Lucia  
Santoso (Australian National University), A/Prof. Phoebe Chen (Deakin University) and Prof. Allen Rodrigo  
(Bioinformatics Institute, University of Auckland).
The Australian Research Council (ARC) Centre in Bioinformatics is a multi-nodal interdisciplinary Centre that uniquely brings together genomic and post-genomic bioinformatics, complex systems research and experimental phenomics to produce a deep understanding of how information in the genome is transformed into structure and function in the mammalian cell.

The ARC Centre in Bioinformatics is committed to:

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

The ARC Centre in Bioinformatics will:

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

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

Our researchers have access to some of the best computational facilities in Australia, including a 16-processor IBM p690 Regatta (IMB), two 64- and one 16- processor SGI Altix systems (UQ), a 120-processor IBM Opteron cluster (IMB), and terabyte storage. Several of our investigators have been awarded time on the 1600-processor APAC national facility.
The Centre’s research is delivered through programs in experimental phenomics, modelling of cellular networks, and data integration and visualisation. Each of these programs is organised and managed as a set of projects. In the following pages we describe the aims of our 13 projects, identify their personnel, and report our achievements during 2005. In a separate (non-public) document we present detailed activity plans for each project for 2006.

In **Program 1** we apply high-throughput experimental methods to develop and test key hypotheses about the control of developmental processes in mammalian cells, and generate data required to model cellular regulatory networks. We use GFP and epitope tagging, high-throughput reporter screens, microarray technologies, confocal and real-time video microscopy, and computational pipelines to obtain genome-wide views of the production and subcellular localisation of key regulatory protein families in human and mouse. The activity in this program provides the gene and protein networks for construction of computational models, and allows us to validate their predictions against observed phenotype. Phenomic data also underlie our computational models of complex cellular structures, including dynamic spatial relationships among vesicles, lipid rafts and tubules during endocytosis. In 2005 we initiated new research based on confocal and real-time video microscopy (Project 1.3) that will extend further continuity across our Centre’s three research Programs.

The objective of **Program 2** is to model and simulate selected networks of gene regulation and protein interaction in mammalian cells, using powerful analytical tools of computational mathematics and complex systems research. To achieve this we must develop efficient algorithms, apply advanced methods in statistics and machine learning, develop expertise in network representation and analysis, properly identify orthologs, and use of high-performance computing effectively and efficiently. In 2005 we made good progress toward linking our Centre’s methods-centric computational mathematics with hypothesis-driven biology on one hand, and with data manipulation, visualisation and querying on the other.

**Program 3** will yield an innovative prototype data, modelling and visualisation environment, the **Visible Cell**, that will seamlessly embed molecular structures, networks and data from quantitative simulations into a virtual 3-D mammalian cell reconstructed from high-resolution tomograms. Understanding how the molecular and regulatory networks in mammalian cells interact in space and time to yield complex phenotype requires us to synthesise large quantities of diverse, multi-scale data including images, molecular structures and mathematical simulations. In other areas of science similarly characterised by very large, complex, multi-scale data (e.g. atmospheric science, physical oceanography), integrative analyses are all but impossible without advanced visualisation.

Genomic biology presents an even greater challenge, as our data are semantically rich and highly interrelated. By directly utilising the most highly evolved pattern-recognition system on Earth – the human visual cortex – our Visible Cell will facilitate hypothesis-formulation and hypothesis-testing with the large, complex, semantically rich datasets of genomic and post-genomic bioscience. By providing a common implementation platform, this project will unify our Centre at the practical, hands-on level. We also appreciate the challenges of making our data, models and visualisations discoverable, searchable and readily integrated across the nascent Grid, and to this end in 2005 we began to convert some of our web-based tools and data to Web Services (Project 3.3).

It is not easy to deliver a deeply integrative research programme across disciplines as methodologically and sociologically diverse as mathematics, information technology and genomics. In 2005 we continued to focus our limited resources on projects that build toward integrative research outcomes, supporting collaborations, student projects, shared datasets, use of common standards and protocols, and our Visible Cell project.
Program 1 – Experimental Phenomics

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

Investigator:

Dr Sean Grimmond
The University of Queensland

In collaboration with:

Prof. Y. Hayashizaki
RIKEN Genomic Sciences Center, Japan

Prof. David Hume
The University of Queensland

Dr Harukazu Suzuki
RIKEN Genomic Sciences Center, Japan

Dr Matt Sweet
The University of Queensland

Dr Christine Wells
Griffith University

Researchers:

Research Assistants:

Mr Geoffrey Faulkner
The University of Queensland

Mr Al Forrest
The University of Queensland

Ms Brooke Gardiner
The University of Queensland

Aims:

(a) Functionally annotate key classes of gene products in the mouse, particularly phosphoregulators and transcription factors. Many of their transcripts potentially encode novel dominant negative and other cis-regulating gene products that are likely to play major roles in mammalian development, differentiation and disease.

(b) Globally monitor the expression of all transcripts from each phosphoregulator locus and define cell-specific variants.

(c) Globally monitor the activity of the non-coding transcripts and alternative promoter usage for every phosphoregulator locus monitor across a panel of cell lines and in cell differentiation models.

Achievements (2005):

(a) As anticipated in the project Aims and in our 2005 Activity Plan, over the past year we focussed work in this project on phosphoregulators. To emphasise our biological objective (understanding transcriptional complexity of the phosphorylation system in mammalian cells), and to reflect a reduced emphasis on reverse transfection technology, the project was re-named as shown above.

(b) In order to study the mammalian phosphorylation system as a network, we conducted computational efforts to identify, classify, annotate and localise of all protein kinases in the mouse at a subcellular and histological level (http://phosphoreg.imb.uq.edu.au). A manuscript was accepted for publication (Forrest et al., BMC Bioinformatics 2006).

(c) As part of FANTOM3 we reviewed the transcriptional output of all loci of the phosphoregulator network (Carninci et al., Science 2005). The 680 loci generate more than 2000 different transcripts and more than 1400 different ORFs. A computational pipeline was developed to summarise diversity in domain content and subcellular localisation motifs for all transcripts and predictions made on the functional implications of each variant (http://variant.imb.uq.edu.au). These studies identified a wealth of putative dominant negative variants, with more than 30% of the receptor kinase and phosphatase loci generating transcripts encoding secreted versions of the receptors, and approximately 50% generating tethered decay or catalytically dead receptor proteins (Forrest et al., Genome Biology 2006). We used public array and MPSS data to validate the existence and transcriptional activities of these variants where possible.

(d) Given the relatively poor performance of all major methodologies for profiling alternative transcription, we undertook to design a screening approach to model systems using splicing arrays. We implemented a computational pipeline that identifies diagnostic probes for each variant in the mouse transcriptome, and in collaboration with Dr Christine Wells used subsets of these probes to study alternative transcription in phosphoregulators and for the Tlr4 signalling pathway (Wells et al., Genome Biology 2006).

(e) As part of FANTOM3, during 2005 we assessed promoter usage in mammals, and integrated this information with gene expression. We found that, on average, each mammalian gene possesses more than two promoters, providing opportunity for novel layers of control of regulation for many genes; a manuscript was submitted to Nature Genetics (Carninci et al.).

(f) We also examined the content of the mouse transcriptome to define transcripts that are likely to be non-coding (Ravasi et al., Genome Research 2005) and those transcripts that are likely to encode small ORFs rather than being truly non-coding (Frith et al., PLoS Genetics 2006).
1.2 Computational cellular biology: combining computational prediction and experimental validation

**Investigators:**
Dr Rohan Teasdale  
The University of Queensland
Dr Timothy Bailey  
The University of Queensland

**In collaboration with:**
Ms Seetha Karunaratne  
The University of Queensland
Dr Harukazu Suzuki  
RIKEN Genomic Sciences Center, Japan
Prof. Y. Hayashizaki  
RIKEN Genomic Sciences Center, Japan
Dr Piero Carninci  
RIKEN Genomic Sciences Center, Japan

**Researchers:**

**Research Officers:**
Dr Zheng Yuan  
The University of Queensland
Dr Lynn Fink  
The University of Queensland
Mr Donald Gardiner  
The University of Queensland
Dr Francis Clark  
The University of Queensland, jointly with APAC Grid Bioinformatics program

**PhD students:**
Mr Raj Aturaliya  
The University of Queensland
Ms Melissa Davis  
The University of Queensland

**Honours student:**
Ms Linda Teng  
The University of Queensland

**Aims:**

(a) Establish a publicly accessible database for data on subcellular localisation of proteins.
(b) Refine the computational prediction pipelines for membrane organisation.
(c) Define the set of membrane and secreted proteins within the murine transcriptome.
(d) Apply pattern discovery and other informatic methods to recognise and systematise our understanding of intracellular targeting signals, and to validate predicted subcellular localisations.

**Achievements (2005):**

(a) We extended the LOCATE subcellular localisation database to include data describing alternately spliced transcripts and effects of alternate splicing on membrane organisation of gene products, and developed an XML schema for these data to facilitate accessibility and distribution through http and Web Services protocols. A manuscript documenting this database will appear in the 2006 Nucleic Acids Research database issue (Fink, Teasdale).

(b) We developed new algorithms for predicting the lipid-accessible surface area within transmembrane domains (manuscript accepted for publication in Journal of Proteome Research), and initiated a small-world network analysis of properties associated with transmembrane proteins (Yuan, Teasdale).

(c) We initiated work on new algorithms for prediction of secondary structure and other structural features of proteins (Yuan, Bailey).

(d) During 2005, the subcellular localisations of the majority of transcriptional regulators encoded in the mouse genome were determined in collaboration with Ms Seetha Karunaratne (IMB, UQ). Based on current gene ontologies, colleagues at RIKEN GSC inferred 1600 proteins to be part of this network. Protein-protein interactions within this set that result in regulation of transcription levels will be determined at RIKEN GSC using high-throughput reporter screens. We applied protocols established in the Teasdale laboratory (PCR-based generation of expression constructs, cell transfection and protein detection) to establish the subcellular localisation of these 1600 proteins, and developed an automatic image analysis protocol to determine relative distribution individual proteins between the nucleus or cytoplasm (see Project 1.3).

(e) Using the membrane organisation prediction pipeline developed by Ms Melissa Davis, we identified 615 genes for which alternative splicing results in differences in protein subcellular localization (Davis, Clark).
1.3 Generation of the experimental data for the modeling of cellular compartments and the development of organelle biogenesis phenotype screens

Investigators:
Dr Rohan Teasdale
The University of Queensland

Prof. Kevin Burrage
The University of Queensland

In collaboration with:
Dr Brad Marsh
The University of Queensland

Prof. Jenny Stow
The University of Queensland

Researchers:
Principal Research Fellow:
Dr Jim Hanan
The University of Queensland, jointly with ACMC, ACCS and the ACILR

Research Associates:
Dr Nick Hamilton
The University of Queensland

Dr Lynn Fink
The University of Queensland

Research Assistant:
Ms Kelly Hanson
The University of Queensland

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

(a) Develop a suite of reagents to mark individual membrane subcellular compartments and organelles.

(b) Express organelle markers within a range of cell types.

(c) Generate stable cell lines expressing GFP-tagged organelle markers.

(d) Establish automatic image analysis and capture technologies (with Dr Nick Hamilton).

(e) Generate custom data in response to requirements of cellular modellers (Dr Nick Hamilton and Dr Jim Hanan).

Achievements (2005):

(a) We identified a set of candidate reporter proteins localised to 13 distinct subcellular regions in the mammalian cell. Mammalian expression vectors for each reporter protein were generated and evaluated by transient expression in HeLa cells.

(b) We developed and implemented the Automated Subcellular Phenotype Classification system (ASPIC) software with the purpose of automatically selecting cells from experimental images and classifying the subcellular expression of tagged proteins using machine learning techniques. During 2005 we implemented a three-class (cytoplasm, nuclear, nuclear/cytoplasm) system and demonstrated that it can yield classification accuracy as great as 95%. The system also provides statistics on the area and intensity of expression.
2.1 Hybrid algorithms for large-scale network and graph optimisation problems in bioinformatics

**AIMS:**

(a) Identify a common core of generic optimisation techniques that are highly "reusable" or "successful" among the different problem areas of interest in the ARC Centre, and implement the algorithmic solutions in a modular framework that will enable parallel, distributed and grid computing.

(b) Develop a systematic approach towards the hybridisation of fixed-parameter tractable (FPT) algorithmics, other stochastic search heuristics and metaheuristics.

(c) Address open research issues of interest to the entire community of the ARC Centre for Bioinformatics, in particular by collaborating in the area of algorithm design, analysis and implementation of large-scale graph and network optimisation problems.

**Achievements (2005):**

(a) A collaboration involving Dr Pablo Moscato (Newcastle Bioinformatics Initiative), Prof. Frank Dehne (Griffith) and Prof. Michael Langston (Oak Ridge National Laboratory and University of Tennessee) has grown from interactions at the 2004 Winter School. Dr Moscato and Dr Langston are collaborating on the analysis of microarray datasets involving a novel algorithm (NGI) to compute \((\alpha,\beta)-k\)-feature sets for genetic signature identification of tumour subtypes, computation of clique intersection graphs (ORNL), and visualisation of an optimised layout display of annotated networks (NGI). This has produced a new scheme for characterising and annotating highly correlated groups of genes in disease subtypes.

(b) In 2005 we recruited Dr Pritha Mahata from Uppsal University (Sweden) to expand algorithm development for use in other ACB projects and to contribute to the development of new hybrid FPT/heuristic algorithms. Dr Mahata and Dr Moscato found the MaxCut algorithm for hierarchical clustering to give inadequate performance with biological datasets, and with A/Prof. Carlos Cotta (University of Málaga, Spain) developed an approach based on a new objective function; this problem has been named \(\text{arithmetic/harmonic cut}\). This new formulation shows promise with a variety of challenging datasets including an 84-species mitochondrial DNA dataset, a 78-sample breast cancer gene expression dataset, and distance and time-series data.

(c) Dr Mahata, Dr Moscato and PhD student Mr Wagner Costa extended the above work on methodology for molecular classification of breast cancers using microarray data by developing new algorithms for the MinWeight Ultrametric Tree problem, and by finding optimal consensus between trees (collaboration with A/Prof. Carlos Cotta and PhD student Mr Mario Inostroza-Ponta). We observed an important correlation between identified subgroups and clinical annotation, pointing to specific groups of genes that appear to correlate strongly with prognosis.

(d) We continued work on natural parameterised problems that arise in motif detection (with Dr Tim Bailey) and ortholog identification (with Prof. Mark Ragan). Attempts to devise useful FPT algorithms for these applications were ongoing through 2005.
(e) During 2005 the main focus of Prof. Michael Fellows was in exploring graph-optimisation problems related to node or edge deletion in clustering. The parallel supercomputer implementation (Dehne, Langston) of the crown rule vertex cover FPT kernelisation algorithm developed by Prof. Fellows and collaborators in 2003-2004 has created a versatile tool for a wide variety of clustering problems in bioinformatics. Vertex cover is essentially minimum vertex deletion to create a single clique. During 2005 we sought to devise similarly useful FPT algorithms for variations of vertex cover, including minimum edge deletion to create a union of some number of cliques (edge edit to cliques) and minimum vertex deletion to create a union of some number of cliques (vertex edit to cliques), and for similar problems involving near-cliques. We obtained a major new result on a closely related problem, minimum vertex deletion to render a graph acyclic (feedback vertex set) and presented it at COCOON 2005. Implementations are underway.

(f) The PhD thesis of Ms Elena Prieto was nominated as best 2005 PhD dissertation in Computer Science at The University of Newcastle, and was one of two runners-up for the 2005 Australasian Distinguished Doctoral Dissertation Award of the Australasian Computer Science Societies.
2.2 Computational discovery of gene-regulatory motifs and motif clusters

**Investigator:**
Dr Timothy Bailey  
The University of Queensland

**In collaboration with:**
Prof. John Mattick  
The University of Queensland
Dr William Noble  
The University of Washington  
Prof. Mark Ragan  
The University of Queensland  
Dr Isidore Rigoutsos  
IBM Thomas J. Watson Research Center  
Prof. Michael Zuker  
Rensselaer Polytechnic Institute

**Researchers:**

**Research Associates:**  
Dr Martin Frith  
The University of Queensland, and RIKEN Genome Sciences Center  
Dr Richard Davis  
The University of Queensland  
Prof. Mark Ragan  
The University of Queensland  
Dr Isidore Rigoutsos  
IBM Thomas J. Watson Research Center  
Prof. Michael Zuker  
Rensselaer Polytechnic Institute

**PhD students:**  
Mr Michael Höhl  
The University of Queensland  
Mr Stefan Stanley  
The University of Queensland

**Visiting researcher:**  
Ms Sonika Tyagi  
Jarawalahal Nehru University

**Aims:**

(a) Understand genetic regulatory networks through the computational discovery of sequence motifs responsible for transcription of protein-coding genes.

(b) Design algorithms and use machine-learning techniques, including hidden Markov models, for the discovery of regulatory motifs and clusters of regulatory motifs.

(c) Develop software and web tools implementing these algorithms and techniques.

(d) Apply pattern-discovery algorithms and software tools to the annotation of genomic regions and proteins, the characterisation and comparison of prokaryotic and eukaryotic genomes, and the genome-scale study of biological processes including the regulation of genes and gene networks.

**Achievements (2005):**

(a) We tested several ways of improving the sensitivity and selectivity of transcription factor binding site (TFBS) prediction algorithms, and incorporated them into our Motiph algorithm (Bailey, Noble, Tyagi) which scans multiple alignments of regulatory regions of orthologous genes for known TFBS patterns. We created “gold standard” reference datasets that contain known TFBSs in regulatory regions of the human genome, and alignments of these regions with orthologous regions in other species (Tyagi).

(b) We developed a motif-discovery algorithm that can learn from negative as well as positive examples. This extends the MEME algorithm, and provides a more-sensitive method for characterising TFBSs as well as other biological sequence signals (Bailey).

(c) We took part in a collaborative assessment of the ability of current algorithms to discover TFBSs from genomic sequence data alone. The comparison was published in *Nature Biotechnology* (Bailey, Frith).

(d) In collaboration with Prof. John Mattick and PhD student Mr Stefan Stanley, we developed GONOME, an algorithm for detecting over- and under-represented sequence features associated with different cellular processes. A webtool was developed and linked to the ACB research website, and a manuscript was submitted to *BMC Bioinformatics*. GONOME has many applications, one of which is *ab initio* discovery of TFBSs.

(e) We developed a statistically based approach to identify “pseudomRNAs”, sequences in the mouse transcriptome that appear to encode proteins but do not (Frith, Bailey). A manuscript describing these “phantoms of the transcriptome” was submitted to *PLoS Genetics*.

(f) We developed and compared statistical approaches to discriminate protein-coding from non-protein-coding mRNAs, and submitted a manuscript for publication (Frith, Bailey).

(g) Short transcripts that apparently encode proteins are remarkably abundant in the mouse transcriptome. We analysed these transcripts using statistical approaches (Frith, Bailey), and submitted a manuscript to *PLoS Genetics*.

(h) In collaboration with Prof. John Mattick, we applied statistical approaches to examine the effect of polyadenylation signals and splicing on 3'-end read-through, hence on transcriptome architecture (Frith, Bailey), and submitted a manuscript for publication.
Under a Materials Transfer Agreement with Rensselaer Polytechnic Institute, we applied a novel energy-based algorithm (Zuker) to the identification of potential miRNA target sites in orthologous regions of the human, mouse and rat genomes (C. Ragan, M. Ragan). We implemented code on the 120-processor IMB Opteron cluster, analysed the genomic distribution and conservation of predicted target sites, identified potential targets involved in a major developmental pathway, and in collaboration with Dr Sean Grimmond worked toward an experimental test of an initial set of top predictions.
2.3 New statistical approaches in bioinformatics

Investigators:
Prof. Geoff McLachlan
The University of Queensland
Dr Markus Hegland
The Australian National University

In collaboration with:
Dr Christophe Ambroise
(Universite de Technologie de Compiègne)
Prof. Kim-Anh Do
M.D. Anderson Cancer Center, University of Texas
Dr Richard Kerr
MMI Genomics
Dr Angus Ng
The University of Queensland
Dr Thriyambakam Krishnan
Systat Bangalore, India

Researchers:

Research Associates:
Dr Richard Bean
The University of Queensland
Dr Liat Jones
The University of Queensland

PhD student:
Mr Justin Xi Zhu
The University of Queensland

Aims:

(a) Develop statistical and computational techniques to analyse cellular transcription.
(b) Classify samples at genomic scale, particularly based on data arising from expression microarray technologies.

Achievements (2005):

In 2005 we focussed on the development of new statistical methodology for a range of problems including detection of differentially expressed genes, the associated issue of multiple tests, and new sources of selection bias in supervised classification of tissue samples where expression levels of thousands of genes are sampled from a limited number of tissues.

(a) We continued development of a simple mixture model approach to the detection of differentially expressed genes in tissue samples in two or more different experimental classes, and reported results in Australian Journal of Experimental Agriculture.
(b) We presented results on stochastic models of gene regulation and expression at the European Conference on Mathematical and Theoretical Biology.
(c) Results on use of the EM algorithm in quantitative trait locus (QTL) detection were published in Genetics Selection Evolution.
(d) We continued work on the second edition of the book with Thriyambakam Krishnan on the EM algorithm and its extensions.

(f) We continued work on problems associated with the selection bias that arises when a classifier is formed on the basis of a small subset of genes selected in some optimal way from a very large number of genes. We published results on the case when the data set lists the expression levels of only the top-ranked, not all, genes.

(g) We further refined the EMMIX-GENE procedure for clustering tissue samples for unstructured gene-expression data, and commenced work on the problem with structured data.
(h) Work is ongoing on the use of gene-expression data in the construction of genetic regulatory networks.
2.4 MATHEMATICAL MODELLING OF GENE REGULATORY NETWORKS

INVESTIGATORS:
Prof. Kevin Burrage
The University of Queensland
Dr Markus Hegland
The Australian National University

In collaboration with:
Dr Pamela Burrage
The University of Queensland
Dr Margherita Carletti
Università di Urbino, Italy
Dr Grant Lythe
University of Leeds, UK

RESEARCHERS:

Principal Research Fellow:
Dr Jim Hanan
The University of Queensland, jointly with ACMC, ACCS and ACILR

Research Associates:
Dr André Leier
Advanced Computational Modelling Centre, The University of Queensland
Dr Nicholas Hamilton
The University of Queensland
Dr Lucia Santoso
The Australian National University
Dr Jiangning Song
Advanced Computational Modelling Centre, UQ

PhD students:
Mr Shev McNamara
The Australian National University, and The University of Queensland
Mr Dan Nicklau Jr
The University of Queensland

AIMS:

(a) Develop quantitative mathematical models to analyse biological systems and networks, in particular genetic regulatory networks, using stochastic differential equations and other formalisms, and implement them in computer code.

(b) Test these models using computational simulation with both simulated and empirical data, and collaborate in their experimental validation where appropriate.

(c) Collaborate and build synergies with other ARC research centres in mathematical modelling of biological systems at the cell and molecular levels.

Achievements (2005):

(a) During 2005, we extended our models of genetic regulatory networks to investigate the behaviour of a toggle switch interfaced with quorum sensing using a new stochastic model. This seems to predict system effects more accurately than do deterministic-based models.

(b) We further investigated switching effects in genetic regulatory networks using the well-characterised lambda phage system.

(c) We began to incorporate temporal delays into our models of regulation of transcription and translation.

(d) We developed new simulation methods that substantially speed up the stochastic simulation algorithm, allowing us to investigate much larger stochastic chemical and molecular reaction systems.

(e) We investigated the use of sparse grids in conjunction with parallel computing implementations for solving the chemical master equation that arises when modelling stochastic chemical reaction dynamics.
2.5 Stochastic simulators for diffusive cellular processes

Investigator:
Prof. Kevin Burrage
The University of Queensland

In collaboration with:
Dr John Hancock
The University of Queensland

Researchers:

Research Associates:
Dr André Leier
Advanced Computational Modelling Centre, The University of Queensland

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

PhD students:
Mr Dan Nicolau Jr
The University of Queensland

Mr Shev McNamara
The Australian National University, and The University of Queensland

Aims:
Many cellular processes involve situations where the number interacting molecular species is small or modest, and/or the media is spatially complex leading to long mixing timescales. In such cases, continuity assumptions (e.g., rate constants, diffusion coefficients) may lead to approximations that are not useful, i.e., fail to abstract actual behaviour. In this project we will develop codes for the (initially Monte Carlo) simulation of diffusive cellular processes under such situations. Specifically we plan to:

(a) Develop prototype stochastic simulators based around specific research questions and/or areas of interest within the Centre.

(b) Develop or modify generalised codes that can be rapidly configured for specific simulation requirements.

(c) Use these tools in the development of higher-level formalisms, and incorporate them into the simulators as they are determined. This, in turn, will allow us to increase the size and complexity of the biological processes that we are able to simulate.

Achievements (2005):

(a) During 2005 we refined our voxel-based model for movements of molecules in two-dimensional model membranes. This Monte Carlo model treats molecules as discrete entities capable of taking small but finite steps in a randomly chosen direction at each time increment. We showed that microdomains can readily operate as protein concentrators or isolators.

(b) This stochastic random walk model of biomolecule diffusion in a cellular membrane was used to investigate the fundamental causes of anomalous diffusion in two-dimensional biological media such as cell membranes. We investigated the roles of fixed obstacles, “picket fence” compartmentalisation leading to “hop” diffusion, and capture by (or exclusion from) lipid rafts.

(c) We developed a new model for simulating anomalous diffusion based on an extension of the stochastic simulation algorithm to crowded environments. Using extensive computational simulations, we were able to relate waiting time to the crowdedness of the environment. Experimental validation was carried out in collaboration with Prof. John Hancock (Institute for Molecular Bioscience).
2.6 Complex systems modelling of regulatory and protein-protein interaction networks

**Investigators:**
Dr Jennifer Hallinan
The University of Queensland

In collaboration with:
Prof. Janet Wiles
The University of Queensland
Prof. John Mattick
The University of Queensland

**Researchers:**

**Research Assistants:**
Mr Daniel Bradley
The University of Queensland
Ms Amanda Barnett
The University of Queensland

**PhD students:**
Mr Nic Geard
The University of Queensland
Mr James Watson
The University of Queensland
Mr Kai Willadsen
The University of Queensland
Mr Simon Carter
The University of Queensland

**Honours student:**
Mr Tien Boon Puar
The University of Queensland

**Aims:**

(a) Develop the Artificial Genome Algorithm as an abstract representation for the modelling of genetic regulatory networks (GRNs) in mammalian cells.

(b) Apply insights gained from computational modelling to GRNs based on biological data, particularly those involving RNA control networks and cancer.

(c) Quantify structural features of GRNs relevant to their behaviour and dynamics in living systems, by application of techniques of graph theory and analytical mathematics.

**Achievements (2005):**

(a) Dr Jennifer Hallinan joined the Institute for Ageing and Health, University of Newcastle upon Tyne (UK) in November 2005. Prof. Janet Wiles is now leading this project. Prof. Wiles was CIA on the ARC Discovery Project grant relinquished into ACB in 2004 that formed the basis of this project.

(b) Development of the network analysis tool suite is nearing completion, and a manuscript describing the tool set and illustrating its use in the analysis of GRNs is in preparation.

(c) Using the network analysis tool suite, we generated a suite of large networks with and without an RNA-type control layer and analysed their dynamics. A manuscript describing the results has been submitted for publication.

(d) During 2005 we extended previous research into the relationship between network topology and dynamics by developing algorithms for the detection of network motifs and feedback loops, and explored the effect of different network motifs on the dynamics of regulatory networks, particularly the shape of the network state space. Dr Jennifer Hallinan reported on this work at the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology in November 2005, and her paper received the conference best paper award.

(e) We produced a database of a highly curated literature information on interactions involving the p53 cell-cycle checkpoint / tumor suppressor protein, and submitted a manuscript on the analysis of this network.
2.7 Ortholog Mapping of Prokaryotic and Eukaryotic Genomes

Investigator:
Prof. Mark Ragan
The University of Queensland

In collaboration with:
Dr Robert Charlebois
NeuroGadgets Inc., Ottawa
Dr Jonathan Keith
The University of Queensland

Researchers:

Research Associates:
Dr Robert Beiko
The University of Queensland
Dr Simon Wong
The University of Queensland

Research Assistant:
Mr Timothy Harlow
The University of Queensland

PhD students:
Mr Cheong Xin Chan
The University of Queensland
Mr Chang Jin Shin
The University of Queensland
Mr Michael Höhl
The University of Queensland

Honours students/trainees:
Ms Abhirami Ratnakumar
The University of Sydney

ACB staff:
Mr Oliver Cairncross
ARC Centre in Bioinformatics

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

(a) Consolidate the development of an automated high-throughput bioinformatic pipeline that identifies protein family clusters over multiple genomes, generates optimal multiple sequence alignments, infers statistically sound phylogenetic trees, and rigorously compares subtrees to identify instances and patterns of topological congruence among trees. These instances of incongruence are prima facie cases of lateral genetic transfer (LGT).

(b) Continue to apply this bioinformatic pipeline to identify orthologous and paralogous gene families, ortholog-characteristic patterns, and prima facie instances of LGT among prokaryotic genomes.

(c) Characterise these instances of LGT statistically, and using advanced database methods correlate them with genomic, physiological, metabolic, ecological, phyletic and other features to develop and test specified hypotheses about the frequency, mechanisms, patterns and biological significance of LGT among prokaryotes.

(d) Continue to develop and implement additional capabilities into this pipeline, notably algorithms and tools that allow the automated recognition and analysis of arbitrarily or self-defined sub- and super-genic regions.

(e) Develop a Grid-aware portal to manage this pipeline, including genome and protein sequences, clusters, alignments, trees, comparisons, analytical results and metadata.

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

Achievements (2005):

(a) During 2005, we achieved project aims (a) and (b) by completing the development of a semi-automated high-throughput bioinformatic pipeline; applying it to the analysis of 42971 proteins in 144 completely sequenced prokaryotic genomes; inferring 22432 Bayesian phylogenetic trees; comparing the 95194 bipartitions supported at BPP ≥ 0.95 against an MRP supertree; and computing minimal edit distances and minimal edit paths that resolve 12721 topologically discordant bipartitions. The main results were published (PNAS 102:14332-37, 2005; rated a “Must Read” in Faculty of 1000).

(b) Three papers describing algorithmic and technical issues relating to this inference pipeline (optimisation, model violation, alignment validation) appeared in print during 2005. At year’s end, one further paper was in press (estimation of edit distance and edit paths), and one manuscript (convergence and mixing in MCMC) and an invited book chapter (technical methods) were under review.
(c) We successfully applied database methods to correlate inferred instances of LGT with selected biological features to test hypotheses about the mechanisms and significance of LGT among prokaryotes. The first manuscript will be submitted during Q1 2006, and we anticipate that other manuscripts will follow.

(d) Comparison of our rigorous phylogenetic results with predictions of surrogate LGT methods yielded a paper that will appear in the January 2006 issue of Trends in Microbiology. Application of our pipeline to five Staphylococcus genomes yielded a high-quality honours thesis for Mr Dave Tang; following further refinement in progress at year’s end, we anticipate submitting a manuscript for publication during Q2 2006.

(e) Following critical comparison of existing methods, we developed a novel alternative method to define orthologous regions among eukaryotic genomes, implemented it in software, and initiated an analysis of four complete animal genomes using the 1600-processor APAC National Facility.

(f) The novel method referred to immediate above is directly extensible to partial genomic, EST and large-scale environmental data. We brought several large metagenomic datasets in-house and a student carried out initial analyses, but we are focusing our effort on the four-genome comparison.

(g) In collaboration with Dr Robert Charlebois, we developed software for statistical modelling and computational simulation of genome evolution including LGT. At year’s end a manuscript was under review, and the software is freely available via the ACB research website (http://bioinformatics.org.au/evolsim).

(h) Following recruitment of new PhD student Mr Chang Jin Shin, the planned sub-project on epistatic selection in mammalian genomes was expanded to a comparison of genomic and cellular determinants of conservation. To this end we initiated construction of a database linking positional, protein-protein interaction, expression and subcellular localisation data for three mammalian genomes. Records are cross-indexed by UniProt number, and assigned Gene Ontology descriptors. The database will be sufficiently developed by Q3 2006 to support cross-domain queries.

(i) We defined and constructed both synthetic and empirical test datasets for benchmarking the application of alignment-free comparative methods, and compared our method with other alignment-free as well as conventional alignment-based approaches. A manuscript was submitted in Q4 2005.

(j) We concluded the benchmarking of approaches for detection of recombination in genomic sequences, with separate focus on the extent of recombination and on recombination breakpoints. In Q2 we submitted a manuscript dealing with optimal approaches to studying genetic recombination. At year’s end, this manuscript remained in initial review. A second manuscript is nearing completion and will be submitted by mid-2006.
PROGRAM 3 –
THE VISIBLE CELL –
DATA INTEGRATION AND
VISUALISATION

3.1 Data integration:
3.1.1 Advanced data management
in bioinformatics

INVESTIGATORS:
Dr Ralf Muhlberger
The University of Queensland

Prof. Mark Ragan
The University of Queensland

Prof. Bernard Pailthorpe
The University of Queensland

Prof. Xiaofang Zhou
The University of Queensland

In collaboration with:
Prof. Kaye Basford
Australian Centre for Plant
Functional Genomics

Ms Kimberly Begley
Griffith University, and APAC Grid
Bioinformatics Program

Prof. Peter Bruza
Distributed Systems Technology CRC

Dr Lindsay Hood
The University of Queensland

Dr Dawei Song
Distributed Systems Technology CRC

RESEARCHERS:
Research Associate:
Dr Lynn Fink
The University of Queensland

PhD students:
Mr Alex Garcia
The University of Queensland

Ms Helen Zi Huang
The University of Queensland

Mr Mohamad Rafi
The University of Queensland

ACB staff:
Mr Oliver Cairncross
ARC Centre in Bioinformatics

Ms Mhairi Marshall
ARC Centre in Bioinformatics

Mr Ken Steube
ARC Centre in Bioinformatics and APAC
Grid Bioinformatics Program

AIMS:
(a) Develop techniques and implement
databases that deal powerfully with
modern genome-phenome data,
including molecular sequences,
structures, interactions and networks,
results of computational modelling
and simulation, images, data from
array-based experimental research,
automatic annotations and metadata.

(b) Identify, implement and integrate
technologies that make it possible
to deliver the above data into a
common research and visualisation
environment (The Visible Cell) in an
efficient, time-critical manner. These
include generic and domain-specific
technologies for data management
and integration, mass storage,
indexing and retrieval, as well as XML
and specialised mark-up languages
(e.g. CellML, MathML) based on XML.

(c) Develop fast database indexing
methods to support efficient
biological query processing with very
large amount of complex biological
data, in particular focusing on protein
surface docking problems.

(d) Grid-enable the Centre’s research
database and selected data.

ACHIEVEMENTS (2005):
(a) During 2005 we continued
developing, populating and
establishing updating protocols for
molecular, cellular, ontological and
systems databases including the
native XML version of SwissProt, and
the genomic cluster data from Project
2.7. However, we accorded this
sub-project a lower priority in view of
other developments (points b and d,
below).
(b) To strengthen coordination with the APAC Grid Bioinformatics programme, Dr Hong Liang Hiew (Murdoch University), representing programme Director Prof. Matthew Bellgard, visited ACB in September to meet with Prof. Mark Ragan and Project 3.1.1 research staff. It was agreed that ACB will make the ENSEMBL databases for human, mouse and all other vertebrates available through data Grid technology (Storage Resource Broker), and that we will in addition serve the ENSEMBL web interface, perhaps via Web Services (Project 3.3). During Q3 and Q4 we mirrored the ENSEMBL data, and implemented some but not all update scripts. The ENSEMBL web interface has proven more problematic because it relies on old software versions that IMB no longer supports. We developed a contingency plan, and at year’s end began to implement the ENSEMBL web interface at Griffith University.

(c) Version 3.1 of Storage Resource Broker failed to compile on the 64-bit equipment we use for external Grid applications. Throughout Q4 we worked with the SRB development team at San Diego Supercomputer Center, but were unable to resolve the problem. IMB will make a 32-bit machine available to us early in Q1 2006 as part of its in-kind contribution.

(d) Prof. Mark Ragan (ACB) and Dr Anthony Maeder (e-Health Research Centre) organised and submitted a proposal to Queensland Innovation Fund to establish an advanced bioinformatics facility that would leverage much of ACB’s data infrastructure R&D team. The partners in this proposal are Griffith University, Queensland Department of Primary Industries & Fisheries, Queensland Parallel Supercomputing Facility, Queensland University of Technology, and The University of Queensland. The proposal was coordinated with the APAN Bio-Mirror, APAC and other relevant parties.

If the proposal is successful, the new facility would provide not only public bioinformatic software and data, but also commercial software such as SRS (LION Bioscience).

(e) Release of a more-easily mirrored version of Protein Data Bank (PDB), originally expected during 2005, has been postponed into 2006.

(f) In Q4 2005 we became the third group in the world, and the first outside the USA, to obtain the first release version of Cell-Centered Database (CCDB) from San Diego Supercomputer Center. We successfully implemented its main schema module, and extended it to store mammalian cell-image volume data under our own schema. However, the application front-end proved not to be fully 64-bit compatible. At year’s end we continued to work on a daily basis with the CCDB development team. Thus CCDB is in use, but as yet cannot easily be accessed by users (e.g. biologists).

(g) Prof. Xiaofang Zhou led the development of a new method for generating and indexing protein unit 3-D structures. “Seed motifs” are represented as structures in high-dimensional space. Up to now, representation of protein structures has largely been based on individual atoms. This new approach supports unit structures of different granularities (from atoms and bowties to complex motifs) and has more-flexible structure similarity definitions that can represent e.g. global similarity between two structures with constraints along each dimension. Application to PDB data demonstrated efficient dimensionality reduction and similarity query processing.

(h) The sub-project on interoperability of models in cell biology was developed sufficiently to be presented in an open seminar during Q3, but has since been put on hold while the student is on leave of absence.

(i) During 2005 Prof. Kevin Burrage visited the UK and strengthened ties with researchers in the UK e-Science Project. ACB consolidated links with the APAC Grid Bioinformatics programme (b above). ACB researchers were involved in a grant proposal, led by Prof. Shoba Ranganathan (Macquarie University), join the European Union BioSapiens project, but the Australian component was not funded by DEST.

(j) Prof. Bernard Pailthorpe consolidated data storage initiatives in Australia, and coordinated Storage Resource Broker (SRB) development and support in QPSF and APAC. This provides 30 TB per annum data support to ACB and IMB. He was an invited participant in a Joint US National Science Foundation – US National Archives and Records Administration planning group for preserving scientific archives.

(k) Prof. Bernard Pailthorpe, Dr Lindsay Hood and Dr Brad Marsh successfully linked the data generation and initial manipulation pipeline for the Visible Cell project with terabyte mass storage at QPSF, and at year’s end the terabyte data store was in routine use by this project.
3.1.2 Classification of genomic data

**Investigators:**
A/Prof. Yi-Ping Phoebe Chen  
Deakin University  
Prof. Kevin Burrage  
The University of Queensland  
Prof. Xiaofang Zhou  
The University of Queensland

**Researchers:**

**Research Associates:**
Dr Jiyuan An  
Deakin University  
Dr Lucia Santoso  
The Australian National University

**PhD student:**
Ms Yan Chen  
Deakin University

**AIMS:**

(a) Develop algorithms and implementations for matching DNA sequences using approaches based on time-warping distance, to deal with large genomic datasets.

(b) Interrelate and analyse disparate data types (sequences, structures, expression data) by classifying and clustering genomic data using data mining and machine learning technologies.

(c) Link interactive methods with data-mining techniques to contribute to the Visible Cell environment.

**Achievements (2005):**

(a) We continued research on pattern recognition, similarity searching and visualisation for genomic sequence data. We published a novel technique for keyword extraction for text categorisation. The algorithm is based on enumeration of combination of dimensions. To avoid the vast number of dimensional combinations, we introduced pruning to create $k$-dimensional bins whose size can be reduced drastically, and the algorithm run much more efficiently, compared with methods such as decision tree and CN2. The patterns produced by our algorithm are also more readily interpretable.

(b) We described the first algorithm that finds complexes from all possible selectors to decide a positive cover. It guarantees that each rule is the best one in the training examples. We plan to analyse its pruning power in extremely high-dimensional data e.g. from gene expression microarrays.

(c) We analysed the preferences of amino acid residues in native protein structures, and based on the atomic distance distribution of all atom types published a new method to predict side chain conformations. To counter the exponential expansion of the number of side chain rotamer candidates, we employed a "beam search" technique to limit the number of candidates with only a little loss of accuracy. Most correct answers are in the candidate set even for small $N$ (100-1000).

(d) We published results using a new indexing method for high-dimensional molecular datasets. This involves a new index structure based on recursive partitioning space. To break the curse of dimensionality, high-dimensional data points are transformed to 1-dimensional values. In this way, classical index structures such as the B+-tree can be adapted. By partitioning the space recursively, our approach overcomes the restriction of two-dimensional pyramids in the pyramid-tree. More pyramids are partitioned, and the selection of key features is improved.
3.1.3 Computational identification of alternatively spliced transcripts in mouse and human, and construction of an alternative splicing database

**AIMS:**

(a) Contribute to the identification and analysis of genetic regulatory and protein-protein interaction networks in mammalian cells by computational identification of alternatively spliced transcripts and alternative-splicing signals.

(b) Contribute in a focused way to specific biological research projects within the ARC Centre in Bioinformatics, e.g., characterisation of the mouse secretome.

(c) Make these data available via an online database of information on alternative splicing in mammalian cells.

**Achievements (2005):**

This project was completed during 2005.

(a) We examined in detail the impact of alternative splicing on the organisation of membrane proteins, specifically orientation with respect to the membrane. Application of the membrane organisation annotation pipeline MemO (generated in Project 1.2) to the mouse FANTOM3 isoform protein sequence set revealed that within the 8032 transcriptional units (TUs) with multiple protein isoforms, 573 show variation in use of signal peptides, 1527 variation in use of transmembrane domains, and 615 generated protein isoforms from distinct membrane organisation classes. The differential use of endoplasmic reticulum signal peptides and transmembrane domains is thus common within the variable protein output of transcriptional units, and generation of protein isoforms targeted to multiple subcellular locations represents a major functional consequence of transcript variation in mouse. This work has been accepted for publication in *PLoS Genetics.*

(b) We integrated visualisation of the underlying mechanisms of alternative splicing within the mouse FANTOM3 isoform protein sequence set into the LOCATE subcellular localisation database (Project 1.2).
3.2 VISUALISATION

3.2.1 Concept and engineering of The Visible Cell

INVESTIGATORS:

Prof. Bernard Pailthorpe
The University of Queensland

Dr Ralf Muhlberge
The University of Queensland

Prof. Mark Ragan
The University of Queensland

Prof. Xiaofang Zhou
The University of Queensland

Researchers:

Principal Research Fellow:
Dr Jim Hanan
The University of Queensland, jointly with ACMC, ACCS and ACILR

Research Associate:
Dr Nicholas Hamilton
The University of Queensland

ACB staff:
Mr Matthew Bryant
ARC Centre in Bioinformatics

Mr Oliver Cairncross
ARC Centre in Bioinformatics

Mr Igor Kromin
ARC Centre in Bioinformatics

Mr Robert McLeay
ARC Centre in Bioinformatics

Mr Ken Steube
ARC Centre in Bioinformatics and APAC Grid Bioinformatics Program

In collaboration with:

Dr Ben Hankamer
The University of Queensland

Dr Lindsay Hood
The University of Queensland

Dr Jane Hunter
Distributed Systems Technology CRC

Dr Brad Marsh
The University of Queensland

Aims:

(a) Refine our conceptualisation of a three-dimensional visualisation environment for exploring the biology of the mammalian cell that seamlessly integrates empirically determined molecular structures and interactions, quantitative simulations based on mathematical and complex systems models, and high-resolution cell tomograms and electron micrographs. This environment, The Visible Cell, will support user interaction with the 3-D cellular models, facilitating hypothesis-generation and hypothesis-testing in genome-phenome biology including gene regulation, cellular growth and development, metabolism, physiology, transport, cell-cell interaction, disease, and cell death.

(b) Based on technical specifications developed earlier, undertake modular development of a prototype Visible Cell.

(c) Explore in detail whether the Cell Illustrator environment can provide the software basis for a pathway modelling and visualisation module within our Visible Cell.

(d) Deepen collaboration with the Data Integration project (3.1, above) and other activity in ACB with the goal of basing The Visible Cell environment over federated databases that will support hypothesis-generation and hypothesis-testing by the efficient delivery of semantically rich information, including user-specified metadata.

(e) Collaborate with experimental phenomics researchers in the use of this environment in the generation and testing of biological hypotheses in mammalian cell biology.
Achievements (2005):

(a) During 2005 we successfully recruited three highly qualified technical staff into the Visible Cell and immediately associated projects, reporting to Project Manager Dr Lindsay Hood: Mr Igor Kromin (from UniQuest), Ms Mhairi Marshall (from Sanger Centre, UK), and Mr Ken Steube (from San Diego Supercomputer Center, USA).

(b) We continued to refine, and began to operationalise, technical specifications for modular development and prototyping of the Visible Cell. In particular, we began to establish a three-dimensional electronic representation of the mammalian beta cell, derived from high-resolution cell tomography, as the working matrix for the Visible Cell environment. We developed and implemented methods to transfer, store, retrieve, sample, display, rotate, manipulate and modify these large (100 GB – 1 TB) images.

(c) We evaluated four software technology frameworks (ADS, Amira, OpenDX and VTK) and selected VTK version 5 as best-suited for our identified technical specifications and user needs.

(d) We established a configuration management process that includes project wikis, forums, versioning control, and task and bug tracking. Code is tested each night for compilability, and all information is automatically backed-up. By year’s end almost 20 projects in ACB and IMB were using this configuration management process.

(e) With the assistance of IMBcom, we signed an agreement with Gene Networks International Inc. (Tokyo) giving us five licenses of GNI’s Cell Illustrator software, as well as confidential documentation and access to their software development team at FQS Poland (a Fujitsu partner). A substantially revised Cell Illustrator version 2.0 was released in Q3 2005, which we installed and learned to use.

(f) We put on three Cell Illustrator information and training workshops during 2005, one (as part of the 2005 Winter School) given by the GNI principals, Prof. Satoru Miyano (Tokyo) and Dr Christopher Savoie (Cambridge, UK). Following on from these workshops, summer intern Mr Robert McLeay worked with three IMB research groups (Prof. David Hume, Prof. Peter Koopman and Prof. Mike Waters) to capture into Cell Illustrator their expert knowledge of specific mammalian cell signalling pathways.

(g) Prototype visualisations of models of biological processes are reported under the corresponding projects (2.4 and 3.2.2). Several components of the Visible Cell were sufficiently functional to be presented (via AccessGrid) in the inaugural eIMBL Symposium on 17 November 2005.
3.2.2 Modelling and visualisation of biological processes in complex spatial environments

**INVESTIGATORS:**

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

**In collaboration with:**

Prof. John Hancock  
The University of Queensland  
Prof. Jenny Stow  
The University of Queensland

**RESEARCHERS:**

**Principal Research Fellow:**
Dr Jim Hanan  
The University of Queensland, jointly with ACMC, ACCS & ACILR

**Research Associate:**
Dr Nicholas Hamilton  
The University of Queensland

**PhD students:**
Mr Shev McNamara  
Australian National University, and  
The University of Queensland
Mr Dan Nicolau Jr  
The University of Queensland

**AIMS:**

(a) Develop active collaborations among mathematicians, modellers and experimental biologists enabling the realistic quantitative description of selected structures, spatial relationships and dynamic processes in mammalian cells.

(b) Develop methodology and local expertise in the modelling of biological processes, with particular focus on lipid rafts and anomalous diffusion.

(c) Work toward collaborative experimental validation and refinement of these models.

**Achievements (2005):**

(a) We further developed a mathematical model and 2-D (projected 3-D) visualisation of membrane-derived vesicles, tubules and endosomes during endocytosis, based on measurement of time-resolved confocal microscopic imaging of mammalian cells provided by Mr Markus Kerr and Dr Rohan Teasdale. This model explains many features observed in the tubule system, including length, branching and membrane recycling, and predicts important parameters including concentration, spatial distribution and time courses of receptors, bound and unbound ligands, lipid rafts and water. Publications emphasising both modelling and experimental aspects of these results have been submitted for publication. The models are available via our Visible Cell wiki.

(b) We developed a model of macropinosome interactions for endocytic maturation in mammalian cells.

(c) We developed a tool for visualising protein domains and sites, in collaboration with Dr Nick Hamilton and Dr Lynn Fink. This tool is used in the multi-pass trans-membrane subcellular localisation predictor (part of the LOCATE database) available through the ACB research website.

(d) We developed a web-based server for predicting contacts between residues in a protein.

(e) We developed a stochastic spatial model of the plasma membrane to determine how microdomains affect protein dynamics. We showed that microdomains can readily operate as protein concentrators or isolators, but there appear to be significant constraints on size and mobility if microdomains are also required to function as reaction chambers that facilitate nano-scale protein-protein interactions.

(f) We developed a stochastic random walk model of biomolecule diffusion on a cellular membrane and applied it to investigate the fundamental causes of anomalous diffusion in two-dimensional biological media such as cell membranes. Three interaction types were considered: collisions with fixed obstacles, “picket fence” compartmentalisation leading to “hop” diffusion, and capture by (or exclusion from) lipid rafts.

(g) We developed a prototype modelling system for intracellular signalling and trafficking based on L-systems, and presented it in a poster at the annual All-Hands Meeting.

(h) In a joint project with the ARC Centre in Complex Systems, we also developed prototype models of intracellular signalling and trafficking based on Petri net approaches, and demonstrated them to ACB and IMB researchers in small group tutorials. A variety of systems were prototyped, including growth factor signal cascades and Toll-like receptors in macrophages. We extended this work to multicellular systems, implementing a model of vulval development in Caenorhabditis elegans using the Cell Illustrator software. The macrophage systems have been targeted for further development in collaboration with Prof. David Hume (IMB) using the Cell Illustrator modelling system based on Petri nets. Preliminary models at a network description level were developed in a joint ACB-ACCS project.
3.3 Bioinformatics Web Services

In Investigator:
Prof. Mark Ragan
The University of Queensland

In collaboration with:
Dr Lindsay Hood
The University of Queensland

Researchers:
Research Associate:
Dr Lynn Fink
The University of Queensland

ACB staff:
Mr Ken Steube
ARC Centre in Bioinformatics and APAC Grid Bioinformatics Program

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

Achievements (2005):
(a) We secured 50% matching funds from APAC via the Queensland Parallel Supercomputing Foundation for a position, and recruited Mr Ken Steube from San Diego Supercomputer Center (USA). He arrived in September 2005.
(b) Owing to Mr Ken Steube’s experience in software engineering and debugging, much of the initial activity was focused on resolving 64-bit compatibility problems with Storage Resource Broker and the Cell-Centered Database (Project 3.1.1).
(c) During Q4 2005, we assisted the Australian Centre for Plant Functional Genomics with their installation of BioMOBY. The developer of BioMOBY, Dr Mark Wilkinson, visited ACB in November 2005.
(d) To assess user needs in a complex distributed project, we participated in a workshop in Sydney in November 2005 co-sponsored by the ARC Research Network in Enterprise Information Infrastructure, the Australian Centre for Plant Functional Genomics, the CRC for Molecular Plant Breeding, National ICT Australia, and the ARC Centre in Bioinformatics.
(e) During Q4 2005, we implemented a prototype web service for high-resolution electron tomography data in the Visible Cell project. This web service acts as an intermediary between the initial data store (Project 3.1.1) and output options.
(f) During Q4 2005, we also implemented web services for protein cluster data from Project 2.7. This service accepts a cluster ID and returns properties including cluster name, phyla involved, recognition threshold, and statistical values cluster.
ACB’s Key Performance Indicators (KPIs) encompass the following areas:

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

Highlights for 2005

ACB hosted and sponsored the 2005 Winter School in Mathematical and Computational Biology. Our co-sponsors were the ARC Research Network in Complex System (COSNet), the Queensland Parallel Supercomputing Foundation (QPSF), Silicon Graphics, Inc. (SGI) and Ceanet Pty Ltd. The Winter School is designed for advanced undergraduate and postgraduate students, postdoctoral researchers and others working in mathematics, statistics, computer science, information technology, complex systems analysis, and biological, chemical and medical sciences and engineering. The 2005 Winter School was highly successful, attracting 169 registrants.

2005 was ACB’s best year for research publications so far. Our investigators co-authored articles in Genome Research, Nature Biotechnology, Nature Reviews Molecular Cell Biology, Proceedings of the National Academy of Sciences USA (two) and Science (two). Five papers from ACB appeared in major bioinformatics journals in 2005 and others are under review.

The Visible Cell project reached its full complement of research personnel fundable under ACB. Software engineering and collaboration in the project is taking place within a formal configuration management system under industrial best practice.

ACB co-sponsored the inaugural meeting of Bioinformatics Australia (Adelaide, 25 September 2005) as well as BioInfoSummer 2005 (Canberra, 28 November - 2 December 2005). Investigators A/Prof. Phoebe Chen and Dr Rohan Teasdale were elected to the inaugural Bioinformatics Australia Committee.

ACB funded Professor Philip Bourne, keynote speaker at Bioinformatics Australia’s inaugural workshop “Bioinformatics enabling the Life Sciences”, to visit the ACB node in Brisbane in September 2005, and meet with our investigators and students.

Three ACB PhD students participated in the IMBcom BioBusiness retreat on the Sunshine Coast in September 2005.

ARC Federation Fellow Prof. Kevin Burrage concluded his tenure as Oliver Smithies Fellow at Balliol College, Oxford.

ACB Advisory Board member Prof. John Mattick was awarded a prestigious ARC Federation Fellowship to examine and model the RNA-based control circuitry of the vertebrate cell. John stepped down as Director of Institute for Molecular Bioscience on 31 December 2005 to devote full time to this very exciting research.

ACB Special Advisor Professor Gene Myers was selected as one of seven founding Group Leaders at the Janelia Farm Research Campus of Howard Hughes Medical Institute.

ACB co-organised the first Australia-Taiwan Bioinformatics Workshop in Hsinchu, Taiwan (6-9 September 2005). Representing ACB and Australia were A/Prof. Phoebe Chen, Dr Martin Frith, Dr Pablo Moscato and Prof. Mark Ragan.

ACB co-sponsored a Plant Genome Informatics workshop in Sydney (14-15 November 2005) with National ICT Australia (NICTA), Australian Centre for Plant Functional Genomics (ACPFG), Molecular Plant Breeding CRC (MPBCRC), and the ARC Research Network in Enterprise Information Infrastructure (EII). Four ACB researchers participated.
1. Research Findings

Quality of publications

In 2005, 23 of our C1 papers (66%) appeared in journals with 2005 impact factor 3.0 or above, 16 (46%) in journals with impact factor 5.0 or above, and 7 (20%) in journals with impact factor 10.0 or above. Two C1 papers (6%) appeared in new journals that do not yet have impact factors.

Number of publications

The table below categorises the published output by ACB investigators and their research groups that appeared with a 2005 publication date. We do not include papers that at the end of 2005 were published electronically in advance of 2006 print publication, or remained in press, under review, submitted or in preparation. For details, please refer to Appendix 1: Publications and Papers.

<table>
<thead>
<tr>
<th>Category</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3 Books – edited</td>
<td>1</td>
</tr>
<tr>
<td>B Book chapter</td>
<td>15</td>
</tr>
<tr>
<td>C1 Journal articles –</td>
<td>35</td>
</tr>
<tr>
<td>articles in scholarly</td>
<td></td>
</tr>
<tr>
<td>refereed journal</td>
<td></td>
</tr>
<tr>
<td>C2 Journal articles –</td>
<td>1</td>
</tr>
<tr>
<td>other contribution to</td>
<td></td>
</tr>
<tr>
<td>refereed journal</td>
<td></td>
</tr>
<tr>
<td>E1 Conference – full</td>
<td>39</td>
</tr>
<tr>
<td>written paper – refereed</td>
<td></td>
</tr>
<tr>
<td>proceedings</td>
<td></td>
</tr>
<tr>
<td>E2 Conference – full</td>
<td>4</td>
</tr>
<tr>
<td>written paper –</td>
<td></td>
</tr>
<tr>
<td>non-refereed proceedings</td>
<td></td>
</tr>
<tr>
<td>E3 Conference – extract</td>
<td>2</td>
</tr>
<tr>
<td>of paper</td>
<td></td>
</tr>
<tr>
<td>E4 Conference – edited</td>
<td>1</td>
</tr>
<tr>
<td>volume of conference</td>
<td></td>
</tr>
<tr>
<td>proceedings</td>
<td></td>
</tr>
<tr>
<td>G Computer software</td>
<td>9</td>
</tr>
<tr>
<td>Other academic outputs –</td>
<td>22</td>
</tr>
<tr>
<td>posters</td>
<td></td>
</tr>
<tr>
<td>Theses</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>138</td>
</tr>
</tbody>
</table>

Number of patents

No patents were awarded to our investigators during 2005.

IMBcom Pty Ltd, a technology management corporation owned by The University of Queensland, conducted BioBusiness training sessions in April (introductory workshop) and August (three-day retreat) 2005. The introductory workshop provided first-year students an overview of the concepts of bio-business. The second was a three-day retreat for third-year students, covering issues in commercialisation as well as career options. Several months prior to the retreat, students were assigned to groups and prepared, with the help of an IMBcom mentor, a business plan for presentation at the retreat. The two BioBusiness sessions form a compulsory component of the postgraduate programme for IMB students.

ACB has arranged with IMBcom to sponsor the participation of its students from other universities, and from other faculties within UQ. ACB students attending the second session in 2005 were:

- Cheong Xin Chan
  Institute for Molecular Bioscience, UQ
- Melissa Davis
  Institute for Molecular Bioscience, UQ
- Justin Zhu
  Mathematics, UQ

Melissa Davis was co-recipient of the award for Best Venture Capital Pitch.
INVITATIONS TO ADDRESS AND PARTICIPATE IN CONFERENCES

Chief Investigators
In 2005, ACB investigators accepted 89 invitations to speak at, or otherwise contribute significantly to, national and international scientific conferences. Invitations to ACB’s Winter School and All-Hands Meeting are not included (see pp. 49 and 50). We list invitations for associates Dr Brad Marsh and Dr Pablo Moscat, but do not include them in this sum. Our postdoctoral researchers and students accepted a further 10 invitations.

Dr Timothy Bailey
- Program Committee, Thirteenth International Conference on Intelligent Systems for Molecular Biology (ISMB 2005), Detroit (June)

Prof. Kevin Burrage
- 2005 International Conference on Scientific Computation and Differential Equations (SCICADE’05), Nagoya (May)
- Sixth European Conference on Mathematical and Theoretical Biology (ECMTB), Dresden (July)
- Universitas 21 Meeting in Computational Biology, Nottingham (July)
- Guest Lecturer, School in Computational Cell Biology, Urbino School, Italy (November)

A/Prof. Phoebe Chen
- Program Co-Chair, Third Asia-Pacific Bioinformatics Conference (APBC2005), Singapore (January)
- Program Chair, Eleventh International Multimedia Modelling Conference (MMM2005), Melbourne, Australia (January)
- Joint Australia-Taiwan Workshop in Bioinformatics, Sydney (February)
- ARC Centre in Bioinformatics Annual Meeting, Brisbane, Australia (April)
- 21st International Conference on Data Engineering (ICDE2005), Tokyo (April)
- International Workshop on Biomedical Data Engineering (BMDE2005), Tokyo (April)
- 2005 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS ’05), Las Vegas (June)

- ARC Research Network in Enterprise Information Infrastructure (EII), Sunshine Coast (July)
- First Australia-Taiwan Bioinformatics Symposium, Taipei (September)
- Knowledge-Based Intelligent Information and Engineering Systems: 9th International Conference (KES 2005), Melbourne (September).
- ACM Multimedia 2005, Singapore (November)
- 18th Australian Joint Conference on Artificial Intelligence, Sydney (December)

Prof. Michael Fellow
- Program Committee, 28th Australasian Computer Science Conference, Newcastle (January-February)
- Plenary lecture, Algorithms and Complexity in Durham 2005, Durham UK (July)
- Plenary lecture, Third Dagstuhl Workshop on Parameterized Complexity, Schloss Dagstuhl, Germany (July)

Dr Sean Grimmer
- US NIH Stem Cell Genome Anatomy Project Programming Jamboree, Washington DC (January)
- Discovery Science and Biotechnology Conference, Melbourne (May)
- US NIH Stem Cell Genome Anatomy Project Bioinformatics Workshop, Philadelphia (June)
- QIMR Institute Seminar, Brisbane (August)
- QIMR Genetics and Epidemiology Division Seminar, Brisbane (August)
- Peter Macallum Cancer Institute seminar, Sydney (October)

Prof. Geoffrey McLachlan
- ANU Mathematical Sciences Institute Workshop, Canberra (February)
- 3rd Asia-Pacific Bioinformatics Conference, Singapore (January)
- Workshop on Statistics for Functional Genomicists and Functional Genomics for Statisticians, Florence, Italy (May)
- Workshop on Oncogenetics, Florence, Italy (May)
- Queensland Branch meeting of the Statistical Society of Australia, Brisbane (May)
- Joint Annual Meeting of the Interface and Classification Society of North America, St Louis (June)
- Keynote address, Bioinformatics Symposium, HELP University College, Kuala Lumpur (July)
- Keynote address, Sixth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL2005), Brisbane (July)
- European Conference on Mathematical and Theoretical Biology, Dresden (July)
- International Conference on Perspectives in Modern Statistical Inference III, Mikulov, Czech Republic (July)
- 16th Conference on Advancement of Animal Breeding and Genetics (AAABG), Noosa (September)

Prof. Jennifer Hallinan
- IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, San Diego (November)

Dr Markus Hegland
- Director, AMSI Summer School in Mathematics, Canberra, Australia (January)
- NICTA Machine Learning Summer School, Canberra, Australia (January)
- Organiser, Workshop on High-dimensional Approximation, Canberra, Australia (February)
- UC Statistics Workshop 4-5 April 2005, Canberra, Australia (April)
- Organiser, Symposium on Optimisation and Data Analysis, Canberra, Australia (September)
- BioInfoSummer, Canberra (December)

- Breaking Complexity: Nonlinear/Adaptive Approximation in High Dimensions, Bad Honnef, Germany (December)
- Fast numerical solution of partial differential equations, Utrecht, Netherlands (December)
- (http://www.math.uu.nl/FastSolvers)

- ANU Mathematical Sciences Institute Workshop, Canberra (February)
- 3rd Asia-Pacific Bioinformatics Conference, Singapore (January)
- Workshop on Statistics for Functional Genomicists and Functional Genomics for Statisticians, Florence, Italy (May)
- Workshop on Oncogenetics, Florence, Italy (May)
- Queensland Branch meeting of the Statistical Society of Australia, Brisbane (May)
- Joint Annual Meeting of the Interface and Classification Society of North America, St Louis (June)
- Keynote address, Bioinformatics Symposium, HELP University College, Kuala Lumpur (July)
- Keynote address, Sixth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL2005), Brisbane (July)
- European Conference on Mathematical and Theoretical Biology, Dresden (July)
- International Conference on Perspectives in Modern Statistical Inference III, Mikulov, Czech Republic (July)
- 16th Conference on Advancement of Animal Breeding and Genetics (AAABG), Noosa (September)

Dr Brad Marsh
- Principal Organiser and Instructor, New Zealand
- Microscopy Conference, University of Otago, Dunedin (February)
- Biophysical Society 49th Annual Meeting and symposium “Advances in High-Resolution Cellular Electron Tomography”, Long Beach, California (February)
- Microscopy and Microanalysis 2005 and Symposium “Tomography in Biological & Materials Sciences”, Honolulu (July)
- ComBio (Australian Society for Biochemistry and Molecular Biology Annual Meeting) and symposium “Cellular Architecture and Biology”, Adelaide (September)
- 45th American Society for Cell Biology Annual Meeting, San Francisco (December)
Dr Pablo Moscato
- All-Hands Meeting of ARC Centre in Bioinformatics, Brisbane (April)
- Microarray Data Analysis Workshop, Brisbane (April)
- 27th Annual Scientific Meeting of the Australia-New Zealand Breast Cancer Trials Group, Perth (July)
- 6th Metaheuristics International Conference (MIC 2005), Vienna (August)
- Program Committee, IV Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, Granada (September)
- First Australia-Taiwan Bioinformatics Symposium Workshop, Hsinchu (September)

Prof. Bernard Pailthorpe
- LAK and CORAL REEF 2005 Workshop, San Diego (March)
- MMS Network Workshop, Melbourne (March)
- APAC Partner’s Forum in Canberra (May)
- APAC’05 Conference, Gold Coast (September)
- Session co-chair via HDTV, Calit2’2s inaugural event Grid05, San Diego (September)
- Symposium on Optimisation and Data Analysis, Canberra (September)
- OECD Global Science Forum on e-Research, Sydney (September)
- APAC-APSR Research Data Collections Workshop, Canberra (October)
- Supercomputing SC’05, Seattle (November)

Prof. Mark Ragan
- ARC Centre for Complex Systems Genetic Regulatory Networks Workshop, Brisbane (January)
- International Society for Evolutionary Protistology (ISEP XVI), Melbourne (February)
- National Bioinformatics Strategy Meeting, Canberra (February)
- Bio21 “Aspects of Informatics in Biology & Medicine” Symposium, Melbourne (February)
- Blue Sky Forum “Convergence of IT and Life Sciences”, Brisbane (March)
- ARC Centre in Bioinformatics Annual All-Hands Meeting and Bioinformatics Institute New Zealand Workshop, Brisbane (April)
- Australian Partnership for Advanced Computing Workshop, Brisbane (June)
- ARC Research Network in Enterprise Information Infrastructure (EII) Advisory Board, Steering Committee and General Member meetings, Sunshine Coast (July)

Dr Xiaofang Zhou
- 16th Australasian Database Conference (ADC 2005), Newcastle (January)
- Third Asia Pacific Bioinformatics Conference (APBC 2005), Singapore (January)
- First IEEE International Workshop on Managing Data for Emerging Multimedia Applications (EMMA 2005), Tokyo (April)
- 15th International Workshop on Research Issues on Data Engineering: Stream Data Mining and Applications (RIDE-SDMA 2005), Tokyo (April)
- Database Systems for Advanced Applications (DASFAA 2005), Beijing (April)
- Asia-Pacific Web Conference 2005 (APWeb2005), Shanghai (April)
- ACM SIGMOD 2005 International Conference on Management Data, Baltimore (June)

Dr Isidore Rigoutsos
- Program Committee, Third Conference on Email and Anti-Spam (CEAS 2005), Mountain View, California (July)

Dr Rohan Teasdale
- Joint Taiwan-Australian Bioinformatics Workshop, Sydney (February)
- Organiser and Chair, Bioinformatics Australia Inaugural Workshop, Adelaide (September)
- Renal Regeneration Consortium Retreat, Brisbane (October)

Prof. Xiaofang Zhou
- 16th Australasian Database Conference (ADC 2005), Newcastle (January)
- Third Asia Pacific Bioinformatics Conference (APBC 2005), Singapore (January)
- First IEEE International Workshop on Managing Data for Emerging Multimedia Applications (EMMA 2005), Tokyo (April)
- 15th International Workshop on Research Issues on Data Engineering: Stream Data Mining and Applications (RIDE-SDMA 2005), Tokyo (April)
- Database Systems for Advanced Applications (DASFAA 2005), Beijing (April)
- Asia-Pacific Web Conference 2005 (APWeb2005), Shanghai (April)
- ACM SIGMOD 2005 International Conference on Management Data, Baltimore (June)

Dr Robert Beiko
- Canadian Institute for Advanced Research Evolutionary Biology Program, Parksville, British Columbia (September)
- BioInfoSummer, Canberra (November)

Dr Martin Frith
- First Australia-Taiwan Bioinformatics Symposium Workshop, Hsinchu (September)

Dr Nicholas Hamilton
- Workshop on Geometry and Applications of Discrete Mathematics, Institute for Geometry and its Applications, Adelaide (February)
- Keynote address, 30th Australasian Conference in Combinatorial Mathematics and Combinatorial Computing, Brisbane (December)
Dr Jim Hanan  
- 17th International Botanical Congress (XVI IBC), Vienna, Austria (July)  
- ARC Centre in Complex Systems Patterns Workshop, Brisbane (July)  
- ARC Centre of Excellence for Integrative Legume Research Hypothesis-driven Modelling Workshop, Brisbane (November)

PhD student  
Mr Alex Garcia  
- Generation Challenge Program Platform Engineering Workshop, International Rice Research Institute, Manila, Philippines (May)  
- Protégé-8, Madrid, Spain (July)

Invitations to visit leading international research groups and research centres

Chief Investigators  
ACB investigators were invited to visit a number of leading international research groups and research centres in 2005.

Dr Timothy Bailey  
- University of Washington, Seattle, USA, invited presentation (October)

Prof. Kevin Burrage  
- During his fellowship at Oxford, Professor Burrage was invited to speak at 12 European universities including Cambridge, Leeds, Sheffield, Warwick, Strathclyde, Reading, Amsterdam, Braunschweig and Geneva  
- University of Auckland, Auckland, NZ, invited presentation (June)  
- Ruakura Agricultural Institute, Hamilton, NZ, invited presentation (June)

A/Prof. Phoebe Chen  
- Tsukuba University, Tokyo, Japan, invited presentation (April)  
- Brain Research Center, National Tsing Hua University, Hsinchu, Taiwan (September)  
- National Center for High-Performance Computing, Hsinchu, Taiwan (September)  
- National Cheng Kung University, Tainan, Taiwan (September)  
- National Health Research Institute, Taipei, Taiwan (September)  
- Institute of Chemistry, Academia Sinica Taiwan, Taipei (September)  
- University of California San Diego (October)

Prof. Mark Ragan  
- Brain Research Center, National Tsing Hua University, Hsinchu, Taiwan (September)  
- National Center for High-Performance Computing, Hsinchu, Taiwan (September)  
- National Cheng Kung University, Tainan, Taiwan (September)  
- National Health Research Institute, Taipei, Taiwan (September)  
- Institute of Chemistry, Academia Sinica Taiwan, Taipei (September)  
- University of Malaysia, Kuala Lumpur, Malaysia, invited lecture and external assessor of M.Sc. Biotechnology programme (September)  
- Bioinformatics Centre, University of Malaya, Kuala Lumpur (September)

Prof. Xiaofang Zhou  
- National University of Singapore, Singapore (January)  
- Nanyang Technological University, Singapore (January)  
- University of Tokyo (April)  
- Harbin Institute of Technology, China (May)  
- Wuhan University, China (August)

Postdoctoral Fellows / Research Officers

Dr Robert Beiko  
- U.S. Department of Energy Joint Genome Institute, invited presentation, Walnut Creek, California (September)  
- University of California, Berkeley (September)  
- University of British Columbia (September)  
- University of Sydney (November)

Dr Martin Frith  
- Brain Research Center, National Tsing Hua University, Hsinchu, Taiwan (September)  
- National Center for High-Performance Computing, Hsinchu, Taiwan (September)  
- National Cheng Kung University, Tainan, Taiwan (September)  
- National Health Research Institute, Taipei, Taiwan (September)  
- Institute of Chemistry, Academia Sinica Taiwan, Taipei (September)  
- Genome Institute of Singapore, invited presentation (December)

Dr Jim Hanan  
- Hort Research New Zealand, Mt Albert Research Centre, Auckland (March)  
- Hort Research New Zealand, Palmerston North Research Centre, Palmerston North (March)  
- University of Minnesota, USA (July)  
- INRA, Institut Jean-Pierre Bourgin, Versailles (July)

PhD student  
Mr Alex Garcia  
- University of Manchester, UK (June)  
- Universidad Santiago de Chile, Santiago de Chile (July)  
- Centro Internacional de Agricultura Tropical, Cali, Colombia (July)  
- Agencia de Cooperacion Espanola, Cartagena de Indias, Columbia (July)
NUMBER AND NATURE OF COMMENTARIES ABOUT THE CENTRE’S ACHIEVEMENTS

The paper by Dr Robert Beiko, Mr Timothy Harlow and Professor Mark Ragan on lateral genetic transfer in the evolution of genomes of prokaryotes (Proceedings of the National Academy of Science USA 102:14332-37, 2005) was featured as a “Must Read” in Faculty of 1000 (http://www.f1000biology.com/article/16176988/evaluation).

Prof. Mark Ragan was interviewed by Mr Robyn Williams, host of Australian Broadcasting Corporation Radio National’s Science Show, on 19 November 2005 on his group’s work on lateral genetic transfer (http://www.abc.net.au/rn/science/ss/stories/s1508801.htm). This work was also featured in number of media items during 2005 including:

- http://www.medicalnewstoday.com/newsid=33457
- http://www.biologynews.net/archives/2005/05/11/map_of_life_on_earth_could_be_used_on_mars.html
- http://www.ivy-rose.co.uk/Health/show_dil.php?id=521
- http://www.bioresearchonline.com/content/news/article.asp?docid=d878e64c-36e1-4b
- http://www.news-medical.net/?id=14478

Dr Jennifer Hallinan was featured in a news feature in Nature 438 (1 December 2005) regarding her weblog (blog) on cancer research.

In January and February 2005, Dr Pablo Moscato was interviewed by several radio stations regarding implications of his work on classification of tumour samples based on gene expression profiles.
Evidence of interdisciplinary linkages within the Centre

Our vision for ACB embraces the interdisciplinary nature of our investigators, research programs and activities. Our investigators, postdoctoral researchers and staff hold degrees in bioinformatics, computer science, information technology, mathematics, molecular biosciences, physics and statistics. Of our 16 projects and sub-projects active during 2005, 13 (81%) include personnel with advanced degrees in different disciplines. Of our 43 papers and posters during 2005 with at least two Centre investigators or students as co-authors, 15 (35%) of our co-authors hold advanced degrees in different disciplines. Eight of 14 (57%) of our co-authored C1 journal publications were interdisciplinary in this sense.

Two of our investigators are also investigators in the ARC Centre for Complex Systems (ACCS), and one of our researchers (Dr Jim Hanan) is jointly appointed by the ACCS, the Advanced Computational Modelling Centre at The University of Queensland, and the ARC Centre of Excellence for Integrative Legume Research.

During 2005 we fostered interdisciplinary and inter-institutional linkages by hosting the Winter School in Mathematical and Computational Biology, and by co-sponsoring BioInfoSummer. We also co-sponsored a workshop on informatics for plant genomics together with the ARC Research Network in Enterprise Information Infrastructure, the Australian Center for Plant Functional Genomics, the CRC for Molecular Plant Breeding, and National ICT Australia. Our four-person delegation to the first Australia-Taiwan Bioinformatics Workshop included two bio- and two information scientists.

2. RESEARCH TRAINING AND PROFESSIONAL EDUCATION

Recruitment

ACB recruited the following in 2005:

- Web Services Research and Development Officer Mr Ken Steube from San Diego Supercomputer Center, a 50% co-appointment with Queensland Parallel Supercomputing Foundation (QPSF), into Project 3.3 (Bioinformatics web services development).
- Research Associate Ms Mhairi Marshall from the Sanger Centre (UK), a co-appointment with IMB (25%), in integration of genomic databases using Storage Resource Broker.
- Research Associate Mr Geoffrey Faulkner, into Project 1.1 (Computational and microarray-based analysis of transcriptional complexity of mammalian cells).
- Software Engineer Mr Igor Kromin, into Project 3.2.1 (Concept and engineering specification of the Visible Cell).
- Research Associate Ms Kelly Hanson, into Project 1.3 (Generation of the experimental data for the modelling of cellular compartments and the development of organelle biogenesis phenotype screens).
- Research Associate Dr Pritha Prieto based at The University of Newcastle into Project 2.1 (Hybrid algorithms for large-scale network and graph optimisation problems in bioinformatics).
- Research Associate Mr Robert McLeay, a joint appointment with the ARC Centre for Complex System (ACCS) (50%), on a term basis into Project 3.2.1 (Concept and engineering specification of The Visible Cell) using Cell Illustrator to develop the more abstract modelling and cellular pathway elements for the Visible Cell project.
- Research Associate Mr Robert McLeay, a joint appointment with the ARC Centre for Complex System (ACCS) (50%), on a term basis into Project 3.2.1 (Concept and engineering specification of The Visible Cell) using Cell Illustrator to develop the more abstract modelling and cellular pathway elements for the Visible Cell project.

International Exchange Program

ACB student Alex Garcia (see page 53) spent twelve-month internship at the European Bioinformatics Institute, on the Wellcome Trust Genome Campus at Hinxton, near Cambridge UK.

Number of postgraduates recruited

In 2005, ABC investigators recruited sixteen Masters/PhD students and eight postdoctoral researchers (Research Officers and Postdoctoral Fellows) into our Programs.
**Number of Postgraduate Completions**

One postgraduate student supervised or co-supervised by ACB investigators was awarded the PhD degree in 2005, and three further PhD thesis were submitted and approved.

**Number of Honours Students**

ACB investigators supervised five honours students during 2005.

**Number of Professional Courses**

ACB investigators did not organise any professional courses during 2005.

**Participation in Professional Courses**

ACB investigators did not participate in any professional courses during 2005.

**Number of Undergraduate Courses in the Centre’s Area**

ACB investigators and researchers participated in the following for-credit courses in 2005:

- **Dr Timothy Bailey**
  - BIOL3014 Advanced Bioinformatics (coordinator)

- **A/Prof. Phoebe Chen**
  - Bachelor of Information Technology (Multimedia Technology) Deakin University (coordinator)

- **Dr Jennifer Hallinan**
  - COMP4001 Introduction to Complex Systems (coordinator)

- **Dr Sean Grimmond**
  - BIOL3004 Genomics and Bioinformatics (lecturer)
  - BIOT7009 Emerging Technologies (lecturer)
  - 2009BBS Genomics and Bioinformatics Griffith University (guest lecturer)

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

- **Dr Rohan Teasdale**
  - BIOL3004 Genomics and Bioinformatics (lecturer)

- **Prof. Xiaofang Zhou**
  - INFS4205/7205 Spatial and Multimedia Databases (lecturer)

- **Dr Jim Hanan**
  - MATH3104 Mathematical Biology (guest lecturer)
  - COMP4001/7001 Introduction to Complex Systems (guest lecturer)

- **Ms Melissa Davis**
  - BIOC8000 Advanced Genomics and Bioinformatics (supervision of post-graduate student for Linda Teng (BSc Hon))
  - BIOL3004 Introduction to Bioinformatics (tutor)

**3. International, National and Regional Links and Networks**

The ARC Centre in Bioinformatics has strong research linkages with many international and national institutions and commercial entities including:

- European Bioinformatics Institute
- Gene Networks International Inc.
- IBM Healthcare and Life Sciences
- IBM Thomas J. Watson Research Center
- King Mongkut's University of Technology Thonburi (KMUTT)
- Korea Advanced Institute for Science and Technology (KAIST)
- National University of Singapore
- Oxford University
- RIKEN Genomic Sciences Centre
- San Diego Supercomputer Center
- University of Auckland
- University of California at Berkeley
- University of California at Santa Cruz
- University of Tokyo
- University of Washington
- US Department of Energy Joint Genome Institute (JGI)
- US National Center for Supercomputing Applications

The Newcastle Bioinformatics Institute (NBI) has established cooperation involving the University of Tennessee, the U.S. Department of Energy's Oak Ridge National Laboratories, and the University of Udine (Italy) and in late 2005 initiated a collaboration with the Functional Genomics group at IBM (Yorktown Heights, New York).

ACB’s 2005 All-Hands Meeting was conducted jointly with Bioinformatics Institute New Zealand (Brisbane, 18–19 April 2005). Thirteen NZ researchers participated. ACB and Bioinformatics Institute will extend their collaboration (e.g. co-funded visits, collaborative teaching, joint symposia and workshops) and consolidate research cooperation by developing joint supervision arrangements and research projects for postgraduate students.
ACB co-organised the first Australia-Taiwan Bioinformatics Workshop in Hsinchu, 6-9 September 2005. ACB representatives were Prof Mark Ragan, A/Prof Phoebe Chen, Dr Pablo Moscato and Dr Martin Frith. The purpose of the visit was to establish research collaborations, and identify grant application opportunities from the respective governments.

ACB funded Professor Philip Bourne, the keynote speaker at the inaugural workshop of Bioinformatics Australia, Adelaide (25 September 2005) to visit Brisbane and present a seminar on 29 September 2005.

ACB co-sponsored a Plant Genome Informatics workshop in Sydney (14-15 November 2005) with National ICT Australia (NICTA), Australian Centre for Plant Functional Genomics, Molecular Plant Breeding CRC, and the ARC Research Network in Enterprise Information Infrastructure. The purpose of the workshop was to bring together leading Australian and international researchers in ICT and plant functional genomics to address critical problems in biosciences enterprise information infrastructure.

Prof. Mark Ragan was a member of the judging committee for the 2005 the Australian Museum Eureka Prize for Bioinformatics Research, sponsored by the NSW Ministry for Science and Medical Research, awarded on 9 August 2005 in Sydney.

As the External Examiner to the Master of Biotechnology programme of Universiti Malaya, Prof. Mark Ragan travelled to Kuala Lumpur to review the program, meet with students and academic staff, and present a seminar (14-24 September 2005). He also met with professors and bioinformatics research staff at Universiti Kebangsaan Malaysia (Kuala Lumpur).

Prof. Mark Ragan participated via Access Grid in the opening ceremony and symposium of eIMBL (electronic International Molecular Biology Laboratory), based in Seoul, Korea on 17 November 2005. eIMBL is a new innovative web-based molecular biological laboratory network linking leading scientists, institutions, entrepreneurs and investors. eIMBL operates as a non-profit organization under the aegis of the Asia-Pacific International Molecular Biology Network (A-IMBN). eIMBL intends to enhance biotechnology capability, facilitate the flow of information and scientists, and increase cooperative activities among APEC countries.

Prof. Mark Ragan met with a delegation from Malaysia visiting IMB in May 2005: Prof. Dato’ Dr Mohd Salleh Mohd Yasin (Vice-Chancellor), Prof. Nor Muhammad Mahadi (Microbiology), Prof. Rahmah Mohamed (Molecular Biology), Mr Mohd Abd Rashid Mohd Fadzil (Bursar), A/Prof. Siti Naishah Hambali (Legal Advisor, Universiti Kebangsaan Malaysia) and Dr Sharr Anzi Harmin (Director, National Biotechnology Directorate, Ministry of Science, Technology and Innovation). Collaboration in research and training in genomics and bioinformatics was discussed.

INTERNATIONAL VISITORS TO ONE OR MORE ACB NODES

Dr Denis Baurain
Université Liège, Belgium

Prof Philip E. Bourne
Professor of Pharmacology, University of California, San Diego and Head, Protein Data Bank (PDB)

Prof. Thomas Cavalier-Smith FRS
Department of Zoology, Oxford University

Prof. Warren Ewens
Department of Biology, University of Pennsylvania

Delegation from Universiti Kebangsaan Malaysia (see above)

Prof. David Galas
Chancellor and CSO, Keck Graduate Institute, California

Dr Grant Lythe
School of Mathematics, University of Leeds, UK

National Tsing Hua University, Taiwan (6 September 2005). From left: Prof. Jung-Hsien Chiang (National Cheng-Kung University), Dr Pablo Moscato, Prof. Jenn-Kang Hwang (Institute of Bioinformatics, National Chiao Tung University), Prof. Ping-Chiang Lyu (Institute of Bioinformatics and Structural Biology, National Tsing Hua University), Prof. Mark Ragan, Dr Hidde De Jong (Institut National de Recherche en Informatique et en Automatique, France), Dr Timothy P. Galitski (Institute for Systems Biology, USA) and A/Prof. Phoebe Chen.
Dr Philip Morrow  
School of Computing and Information Engineering, University of Ulster

Dr Sharr Anzi Harmin  
Director, National Biotechnology Directorate, Ministry of Science, Technology and Innovation, Malaysia

Dr Philip Hugenholtz  
Group Leader, Microbial Ecology Program, US Department of Energy Joint Genome Institute, Walnut Creek, California

Prof. Hiroaki Kitano  
Director, Sony Computer Science Laboratories, Inc., Project Director of Kitano Symbiotic Systems Project, ERATO-SORST, Japan Science and Technology Agency, President of The Systems Biology Institute, and Adjunct Professor, Keio University, Japan

Prof. Sang Yup Lee  
LG Chem Chair of Chemical Engineering, Head of Metabolic Engineering National Research Laboratory, Director of BioProcess Engineering Research Center, and Director, Bioinformatics Research Center, Korean Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

Dr Romeo Rizzi  
Department of Informatics and Telecommunications, University of Trento, Italy

Prof. Osvaldo Rosso  
Institute of Computation, University of Buenos Aires

A/Prof. Andrey Rzhetsky  
Columbia Genome Center and Department of Medical Informatics Bioinformatics Unit, Columbia University, New York

Mr Geir Kjetil Sandve  
Norwegian University of Science and Technology, Trondheim

Dr Wojtek Szmer  
Manager, Software Development, FQS Poland

A/Prof. Masao Nagasaki  
Institute of Medical Science, Human Genome Center, University of Tokyo

Prof. Mark Wilkinson  
Department of Medical Genetics, University of British Columbia, and iCAPTURE Centre, St. Paul’s Hospital, Vancouver, BC, Canada

Mr Philip Yeo  
Chairman, Agency for Science Technology and Research (A*Star), Singapore

National visitors to one or more ACB nodes

Dr Michael Breakspear  
School of Psychiatry, University of New South Wales, and The Brain Dynamics Centre, Westmead Hospital

Dr Stephen Goldon  
Chief Research Strategist, AgResearch New Zealand

Dr Lars Jermiin  
The University of Sydney

Prof. Tony Larkum  
The University of Sydney

Mr Hiew Hong Liang  
IT Principal Scientist, Centre for Bioinformatics and Biological Computing, and School of Information Technology, Murdoch University

Dr Jesus A. Lopez  
Coordinator, Bioinformatics Interest Group, Department of Mathematics and Computing, University of Southern Queensland

Prof. Mandym V. Srinivasan  
ARC Federation Fellow, Research School of Biological Sciences, The Australian National University

Prof. Sue Wilson  
Co-Director, Centre for Bioinformatics Science, The Australian National University

A/Prof. Michael Wise  
Biochemistry and Molecular Biology, The University of Western Australia

National and international workshops

In 2005, ACB hosted and/or co-sponsored:

- Annual All-Hands Meeting (18 April 2005)
- ACB – Bioinformatics Institute New Zealand joint workshop (19 April 2005)
- Workshop on Large Scale Gene Expression Analysis (20 April 2005)
- Winter School in Mathematical and Computational Biology (11-15 July 2005)
- The inaugural meeting of Bioinformatics Australia (Adelaide 25 September 2005)
- IMBicom BioBusiness workshops (August/September 2005)
- Plant Genome Informatics Workshops, with National ICT Australia, Australian Centre for Plant Functional Genomics, Molecular Plant Breeding CRC, and ARC Research Network for Enterprise Information Infrastructure (14-15 November 2005)

**ACB Annual All-Hands Meeting and ACB – Bioinformatics Institute New Zealand Joint Workshop (18-19 April 2005, Brisbane)**

These two back-to-back meetings brought together more than 75 Investigators, postdoctoral researchers, research assistants, students (PhD, Masters, honours, exchange), ACB staff and colleagues from The University of Queensland, Bioinformatics Institute New Zealand, The University of Auckland, AgResearch New Zealand, CSIRO, Monash University, DSTC Ltd, Griffith University, Nanjing University, University of Southern Queensland, The University of Newcastle, and Deakin University. The 13-member group from Bioinformatics Institute New Zealand was headed by Director Prof. Allen Rodrigo.
Workshop on Large Scale Gene Expression Analysis (20 April 2005)

More than 100 participants attended this workshop, which was sponsored by Queensland Bioinformatics Consortium in conjunction with Bioinformatics Australia. Speakers were:

- Dr Richard Brandon
  Genetraks Pty Ltd
- Dr Conrad Burden
  The Australian National University
- Dr Michael Elashoff
  Private consultant based in the USA
- Dr Sean Grimmond
  Institute for Molecular Bioscience, The University of Queensland
- Dr Robert Henke
  Millennium Science Pty Ltd
- Dr Tim Littlejohn
  IBM Australia
- Dr David Mitchell
  CSIRO Mathematical and Information Sciences
- Dr Pablo Moscato
  The University of Newcastle
- Dr Glenn Stone
  CSIRO Mathematical and Information Sciences
- Dr Gordon Smythe
  Walter & Eliza Hall Institute of Medical Research
- Dr Mervyn Thomas
  Emphron Informatics Inc.

Winter School in Mathematical and Computational Biology (11-15 July 2005, Brisbane)

169 participants attended, about 65% from Queensland universities and the rest from major universities, other research institutes across Australia, CSIRO, and universities and research institutes in Belgium, Japan, New Zealand, Taiwan and USA.

From Australia:
- Australian Centre for Plant Functional Genomics
- The Australian National University
- Bio21 Molecular Science, Melbourne
- Ceanet Pty Ltd
- CSIRO Livestock Industries
- CSIRO Mathematical and Information Sciences
- CSIRO Sustainable Ecosystems
- Deakin University
- Griffith University
- James Cook University
- National ICT Australia
- Queensland Institute of Medical Research
- Queensland University of Technology
- RMIT University
- Silicon Graphics, Inc. (SGI)
- Swinburne Institute of Technology
- The Walter and Eliza Hall Institute of Medical Research
- The University of Adelaide
- The University of Melbourne
- The University of New England
- The University of New South Wales
- The University of Queensland
- The University of Sydney
- The University of Western Australia

From overseas:
- Columbia University, USA
- Gene Networks International, Japan and UK
- Joint Genome Institute, US Department of Energy
- National Tsing Hua University, Taiwan
- Novartis Institute for Tropical Disease, Singapore
- Université Liège, Belgium
- University of Auckland, New Zealand
- University of Pennsylvania, USA
- University of Tokyo, Japan
- Wellington Institute of Technology, New Zealand

Each day of our Winter School focused on a different theme:

Day 1
Complex biological systems & networks
- Prof. Janet Wiles
  School of Information Technology & Electrical Engineering
  The University of Queensland
- Dr Jennifer Hallinan
  Institute for Molecular Bioscience, and ARC Centre in Bioinformatics
  The University of Queensland
- Prof. John Mattick
  Institute for Molecular Bioscience
  The University of Queensland
- Dr Christopher Savoie, CEO
  Gene Networks International
  Cambridge, UK
- Dr Michael Gagen
  Institute for Molecular Bioscience
  The University of Queensland

Day 2
Molecular evolution and phylogenetics
- Dr Lars Jermiin
  School of Biological Sciences
  University of Sydney
- Prof. Mark Ragan
  Institute for Molecular Bioscience, and ARC Centre in Bioinformatics
  The University of Queensland
- Dr Rob Beiko
  Institute for Molecular Bioscience, and ARC Centre in Bioinformatics
  The University of Queensland

ACB hosted and sponsored a five-day Winter School at Queensland Bioscience Precinct. The ARC Research Network in Complex System (COSNet), Queensland Parallel Supercomputing Foundation (QPSF), Silicon Graphics, Inc. (SGI) and Ceanet Pty Ltd were co-sponsors.
Day 5
Computational neurobiology

Prof. Perry Bartlett
ARC Federation Fellow
Queensland Brain Institute
The University of Queensland

Dr Pankaj Sah
Queensland Brain Institute
The University of Queensland

Dr Geoff Goodhill
Queensland Brain Institute
The University of Queensland

Dr Michael Breakspear
School of Psychiatry,
University of New South Wales
The Brain Dynamics Centre,
Westmead Hospital

Prof. Mandyam V. Srinivasan
ARC Federation Fellow
Research School of Biological Sciences
The Australian National University

BioInfoSummer
(28 November - 2 December 2005, Canberra)

ACB again co-sponsored BioInfoSummer at The Australian National University. Prof. Geoff McLachlan, Prof. Mark Ragan, and Dr Robert Beiko were invited lecturers.

2006

In 2006, ACB plans to host or co-sponsor:

- ACB Chief Investigators’ Retreat (2 February 2006)
- Annual All-Hands Meeting (date to be coordinated with the ARC review)
- Winter School in Mathematical and Computational Biology (26-30 June 2006)
Dr Teasdale negotiated a licence agreement with BioMed Central Biology Image Library in regard of image data from the LOCATE database (http://locate.imb.uq.edu.au, also available via ACB’s research website).

ACB sponsored three PhD students to participate in IMBcom’s three-day BioBusiness retreat, where they carry out projects, receive training in technology transfer and commercialisation, and meet businesspersons, entrepreneurs and legal representatives. Ms Melissa Davis shared the award for Best Venture Capital Pitch with Mr Leith Fremlin, Mr Daniel Sangermani, and Mr Andy Wu (IMB).

**NUMBER OF GOVERNMENT, INDUSTRY AND BUSINESS BRIEFINGS**

**A/Prof. Yi-Ping Phoebe Chen**
- Victorian Partnership for Advanced Computing, Melbourne (February)
- Elected member, inaugural Bioinformatics Australia Committee (September)

**Prof. Bernard Pailthorpe**
- Meeting of the ACcESS MNRF and Dr Peter Andrews, Queensland Chief Scientist, Brisbane (July)
- “Science in Parliament”, Brisbane (August)

**Prof. Mark Ragan**
- National Bioinformatics Strategy, Canberra (February)
- Premier of Queensland’s launch of the Smart Queensland: Smart State Strategy 2005-2015 (April)
- Dr Stephen Golden, Chief Research Strategist, AgResearch New Zealand (May)
- 2005 Premier of Queensland’s SMART Awards Presentation (August)
- 2005 Australian Museum Eureka Prizes Awards (August)
- Ms Shamila Mercier, Mercorp Consulting Pty Ltd (Queensland ICT R&D Strategy Review) (October - December)
- Dr Janet Pagan, Head and Mr Peter Mere, Policy Analyst, DCITA ICT Strategy & Analysis (November)

**Dr Rohan Teasdale**
- Member, National Health & Medical Research Council Working Group on development of the Medical Bioinformatics, Genomics, and Proteomics Program
- Elected Member, inaugural Bioinformatics Australia Committee (September).

**NUMBER OF CENTRE ASSOCIATES TRAINED OR TRAINING IN TECHNOLOGY TRANSFER AND COMMERCIALISATION**

Three PhD students from ACB’s nodes took part in the BioBusiness 3-day retreat hosted by IMBcom in 2005 (see page 41).

**NUMBER AND NATURE OF PUBLIC AWARENESS PROGRAMS**

ACB continues to expand our research website (http://bioinformatics.org.au) and maintains an active, up-to-date list of Australian and overseas bioinformatics events, conferences and job opportunities. Our website received more than 10,000 unique hits during 2005, from a monthly average of 350 unique addresses. Slightly more than half of visitors to our website came from the .au (Australia) top-level domain.

Our website also hosts a growing number of software and data resources in bioinformatics and computational biology, including Tim Bailey’s highly popular MEME, MAST and Meta-MEME webtools, and tools and databases from Francis Clark (alternative splicing), Rohan Teasdale (mouse and human LOCATE databases, transcriptional regulators database, DomainDraw software), Sean Grimmond (PhosphoregDB, VariantDB) and Mark Ragan (EEP, GANN, WOOF and EvoSimulator). We provide hyperlinks to software from other Australian bioinformatics research groups at ANU, Monash, UNSW, UQ and WEHI.

The GPIPE workflow environment (Garcia et al., *BMC Bioinformatics* 6:87, 2005) is now in use at several institutions internationally including Institut Pasteur (Paris) and Centro Internacional de Agricultura Tropical (Cali), where new features are being added and a ZOPE migration is in progress. A web version will be supported through mid-2006 at IMB (http://it-web1.imb.uq.edu.au/Pise/5.a/gpipe.html) while migration to CIAT is in progress; executables can be downloaded from ftp://ftp.pasteur.fr/pub/GenSoft/unix/misc/Pise/G-Pipe.tar.gz.

Dr Rohan Teasdale co-organised Bioinformatics Australia (http://www.ausbiotech.org/bioinformatics.asp), and he and A/Prof. Phoebe Chen were elected to its inaugural steering committee (September 2005). Bioinformatics Australia (BA) was created in 2004 by the merger of AusBiotech’s Bioinformatics Special Interest Group (ABSIG) and the Australian Association for Bioinformatics (AAB). Bioinformatics Australia was
created under the auspices of AusBiotech to steer and advise bioinformatics in Australia, and to provide leadership in bioinformatics research, education and commercialisation. In its first full year of operation, Bioinformatics Australia co-sponsored several events including the microarray data analysis workshop associated with our 2005 Annual All-Hands Meeting (20 April 2005, Brisbane) and the First Bioinformatics Australia workshop held in conjunction with ComBio (25 September, Adelaide).

ACB hosted and co-sponsored the second Winter School in Mathematics and Computational Biology (11-15 July 2005), attended by 169 participants from 34 universities and institutions in seven countries. Large posters with ACB’s logo and website address were mailed to all bioscience, medical, mathematics, statistics, computer science and information technology faculties and departments in Australia and New Zealand, and sent by email to nearly 400 individual researchers.

We distributed about 250 printed copies of the ARC Centre in Bioinformatics 2004 Annual Report (not including electronic copies through our website), more than 200 copies of our Student Scholarship prospectus, and more than 100 copies of a sponsorship brochure aimed at industry. About 100 of the Annual Reports were distributed by post to university departments, institutes research centres in Australia and New Zealand, while almost all of the other printed items were distributed at conferences, events and briefings in Australia, Japan, Malaysia, New Zealand, Singapore and the UK.

My assignment, during my year at EBI, was to be the knowledge engineer in the development of ontologies for functional genomics, more specifically for environmental genomics, nutritional genomics (nutrigenomics) and toxicological genomics (toxicogenomics). Our first problem was that no formal methodology existed to guide ontology development. After studying the problem, I began to apply conceptual mapping theory and cognitive support principles for GUI design from the cognitive sciences, and introduced groupware technology to support our knowledge engineering. Conceptual maps provided a tool to elicit knowledge from domain experts, and to start designing a methodology for ontology development. An important issue about methodologies is their relationship with the software that is meant to support them, and this led me to try to design a groupware technology that can provide cognitive support for developing ontologies in decentralised environments such as the bioscience community.

From both personal and professional angles my time at EBI was, without doubt, a joyful experience. It was year in which I could not only study, but also see the impact and relevance of my work. It gave me an opportunity to think about the immense work ahead, and how little we really know about “things” and their relationships. My personal appreciation of people, and work in science, changed radically after this experience, and allowed me to understand better the importance of other people’s points of view, and how their personal histories influence these outlooks. When working with knowledge we tend to consider humans as “informants”, “domain experts”, and/or “users”, imposing a space between the knowledge engineer and the others. But really we don’t work with knowledge: we work with people. We therefore need to understand people beyond their work, from a broader human perspective, to gain the best insights into their know-how.

ACB student Alex Garcia’s twelve-month internship at the European Bioinformatics Institute, on the Wellcome Trust Genome Campus at Hinxton, near Cambridge UK

One year at the European Bioinformatics Institute, working in an environment where academic needs meet industrial requirements, is a challenging opportunity for any PhD student. For me, EBI provided an open door to understanding some of the human factors, community policies and behaviours associated with the design of technology for bioinformatics research. In the same way that toy companies involve kids in the design of new toys, working with researchers makes one reflect on the role of the scientist in the development of his/her own technology. What do biologists actually think about bio-ontologies and the role ontologies may have in their information systems? How can we use conceptual design in the bio domain, and what is the relationship between declarative and functional models?

These questions are not new for computer scientists, but have not been fully answered in the bio domain. The aim of functional genomics is to understand biological systems from a pragmatically holistic perspective, so information sciences clearly must play a critical role. But other disciplines are important too, including the cognitive sciences and philosophy. If we are to integrate information to facilitate the semi-automatic discovery of knowledge, interdisciplinary work is needed. It was precisely this interdisciplinarity that my time at EBI allowed me to see from different perspective.
5. Organisational support

Student industrial scholarship

In 2005, we worked with The University of Queensland’s Development office to generate a list of international biotechnology companies, particularly in India and Taiwan, that ACB might approach to sponsor Student Industry Scholarships. This process led to the First Australia-Taiwan Bioinformatics Workshop in Hsinchu (6-9 September 2005), co-sponsored by ACB and the Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu. For details of this workshop, please refer to International, National and Regional Links and Networks, page 47.

First Australia-Taiwan Bioinformatics Symposium, National Tsing Hua University, Hsinchu (6-9 September 2005).
### Annual Cash Contributions: Collaborating Organisations

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

* Possible adjustments during Years 2-5 not shown. ** In addition to the cash support identified here, the Centre will benefit from half the IGS generated by the core ARC grant funds at UQ that is distributed via IMB. *** Does not include an IBM SUR (Shared University Research) grant valued at A$3 million.

### Annual In-Kind Contributions: Collaborating Organisations

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

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

**TOTAL (ANNUAL CASH + ANNUAL IN-KIND CONTRIBUTIONS)**: 2,577,556, 2,276,756, 2,296,538, 1,999,424, 2,018,872
Number of new organisations recruited to or involved in the Centre

During 2005, ACB further developed our strategic partnership with the Newcastle Bioinformatics Initiative (The University of Newcastle).

One position within ACB was co-funded by the ARC Centre for Complex Systems and the ARC Centre of Excellence in Integrative Legume Research, and another by the Grid Bioinformatics Program of Australian Partnership in Advanced Computing.

Level and quality of infrastructure

The Centre is headquartered on Level 6–West of the Queensland Bioscience Precinct (QBP) at the St Lucia campus of The University of Queensland. Level 6–West is a purpose-built space for bioinformatics, with space for more than 75 researchers and students.

Refer also to page 17 Capabilities – Core technologies and infrastructure.

6. National benefit

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

During 2005, ACB continued involvement in the preparation of Australia’s first National Bioinformatics Strategy. Director Prof. Mark Ragan, and Advisory Board members Prof. Ross Coppel, Prof. Simon Easteal and Prof. Marc Wilkins were members of the NBS Bioinformatics Expert Task Force. The National Bioinformatics strategy was released in November 2005:


Prof. Mark Ragan was invited to brief Dr Janet Pagan, Head and Mr Peter Mere, Policy Analyst, Department of Communications, Information Technology and the Arts about ICT aspects of ACB’s research programme (November 2005).

In December 2005 Prof. Mark Ragan submitted a Queensland Innovation Fund proposal to provide advanced bioinformatics data and services for the Queensland biotechnology, health and IT sectors, and as appropriate to build into a broader national role. Dr Anthony Maeder (e-Health Research Centre, a joint venture between Queensland Government and CSIRO) was co-proponent. Support for the bid was secured nationally from Australian Partnership for Advanced Computing, and within Queensland from a number of academic and research institutions. For more details, see page 51, Section 4 (End-user links).

Case studies of economic, social, cultural or environmental benefits

The criterion for case studies of economic, social, cultural or environmental benefits was not brought forward into the KPIs for ARC Centres.
Appendix 1: Publications and Papers

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

A. BOOKS

A3. EDITED


B. BOOK CHAPTERS


C. JOURNAL ARTICLES

C1. JOURNAL ARTICLES IN SCHOLARLY REFEREED PUBLICATIONS


E2. CONFERENCE – FULL WRITTEN PAPER – NON-REFEREED PROCEEDINGS


E3. CONFERENCE – EXTRACT OF PAPER


E4i. CONFERENCE – EDITED VOLUME OF CONFERENCE PROCEEDINGS


E4. CONFERENCE – UNPUBLISHED PRESENTATIONS

This listing does not include presentations at the 2005 Winter School in Mathematics and Computational Biology (see page 50), or at Centre’s 2005 Annual All-Hands Meeting.


G. COMPUTER SOFTWARE


OTHER ACADEMIC OUTPUT

POSTERS


THESES


Acknowledgements:

Images shown on the cover and page 25 were provided courtesy of Ms Seetha Karunaratne, Senior Research Assistant of Dr Rohan Teasdale’s group, and page 31 by Dr Rohan Teasdale, Institute for Molecular Bioscience, The University of Queensland.
## APPENDIX 2: FINANCIAL STATEMENT

### INCOME
- Balance carried forward 2004: $905,169
- ARC Centre Grant - 2005: $827,728
- Host Institution Support: $555,000

**Total Income**: $2,287,897

### EXPENDITURE

#### Personnel
- Centre Manager: $6,293
- Research Website Administrator: $71,632
- Database Administrator/Developer: $63,819
- Computer Systems Officer: $39,027
- Web Services Developer: $25,014
- Web & Database Developer: $28,822
- Software Engineer: $28,576
- PDFs/Research Officers (10): $459,309
- Research Assistants (5): $101,068
- Programmers (3): $40,225

**Total Personnel**: $913,785

#### Scholarship
- $92,917

#### Funds to other nodes
- $376,122

#### Student Programming/Research Experience/Workshops
- $14,106

#### Strategic Visitors/Sponsorship Program/Director Strategic Discretionary
- Sponsorship - Winter School: $20,000
- Sponsorship - BioInfoSummer (ANU): $3,580
- Sponsorship - Plant Genome Informatics Workshop (EII): $532

**Total Strategic Visitors/Sponsorship Program/Director Strategic Discretionary**: $24,112

#### Equipment
- $33,144

#### Accommodation
- $7,539

#### Travel
- Local travel: $4,531
- Overseas travel: $7,150
- **Total Travel**: $11,681

#### Maintenance
- Conferences/meetings: $5,081
- Computer consumables/maintenance: $207
- Laboratory consumables: $36,285
- Printing and stationery: $14,846
- Postage and freight: $471
- Recruitment: $1,567
- Relocation: $585
- Software licence: $10,118
- Telephone and fax: $290
- Travel subsistence: $123
- Transport: $3,482

**Total Maintenance**: $73,056

#### Other
- Other general expenses and cost recovery etc.: -$15,637

**Total Expenditure**: $1,530,824

#### Balance
- $757,073