Division of Genomics & Computational Biology

RESEARCH FOCUS

This program includes the ARC Centre of Excellence in Bioinformatics and the Queensland Facility for Advanced Bioinformatics. It intersects with the Department of Mathematics and the School of Information Technology and Electrical Engineering. It focuses on understanding the genetic programming of humans, specifically, comparative mammalian and vertebrate functional genomics, rnomics, and computational modelling of genetic and cellular regulatory networks (i.e. the Visible Cell® project).

Research Group Leaders

Tim Bailey
Kevin Burrage
Sean Grimmond
Nick Hamilton
John Mattick
Mark Ragan
Rohan Teasdale

The IMB is a highly collaborative environment where researchers from different fields combine to contribute to strategic research programs investigating the basis of growth and development at the genetic, molecular, cellular and organ levels. Only by understanding the complex molecular and cellular events that occur throughout a normal human life can scientists understand abnormalities responsible for many common human diseases and to find treatments for them.
My research develops and applies computational methods to extract knowledge and understanding of biological processes from the huge quantities of raw data made possible by automated biology. The current focus of my research is on understanding how the cell regulates the expression of genes. My approach is to develop computer algorithms for discovering patterns in high-throughput data related to control of gene expression, and to build models of regulation based on those patterns. Knowing how gene expression is regulated is essential to understand cellular processes such as reproduction and metabolization. It will also enhance our understanding of development and pathologies in higher organisms, and may also lead to advancements in biotechnology.

This year my group studied combining multiple types of high-throughput data for predicting the places in the genome where transcription factor proteins bind to DNA to control gene transcription. We extended our "evolutionary motif model" of binding sites to allow for the loss of sites during evolution, and conducted a survey of the relative power of various conservation-based binding site prediction methods. We also published the first study showing that using another type of data—epigenetic information enables accurate tissue-specific prediction accuracy. We extended our work on an initial motif discovery to include the use of non-sequence data such as "binding intensity" data from chromatin immunoprecipitation experiments. This year we also refined our computational model of the regulation of gene transcription, showed that it can predict gene expression in novel species, and studied algorithms for "training" models of expression. We have integrated these algorithms into the popular MEME Suite of motif-based sequence analysis tools, of which I am a principal author.

In the coming year, we will apply the tools we developed this year to the study of the regulation of gene expression in developing blood and neural cells. We will also continue to develop novel algorithms for pattern discovery and modeling, especially algorithms that combine additional types of non-sequence data for discovering the targets of transcription factors and for identifying interactions among factors. We will also initiate a project to analyze the role of DNA-RNA triplex formation in gene expression.

**KEY PUBLICATIONS**


This group works on developing simulations and visualization methodologies for understanding the behaviour of complex cellular processes, both on the plasma membrane, in the cytosol and at the genetic regulatory level. The simulation models take into account stochastic effects, while the visualization focuses on two- or three-dimensional displays. In microscopic systems formed by living cells, the smallest molecules of interest can result in dynamic behaviour that is discrete and stochastic rather than continuous and deterministic. Our research introduces new classes of discrete stochastic methods that more accurately and effectively reflect the underlying cellular models. We are also focusing on some new methods for both large-scale kinetics and spatial models that more faithfully capture complex kinetics and transport processes within the cell.

**RESEARCH PROJECTS**

- Developing new Monte-Carlo simulation techniques in conjunction with the group of John Hancock and researchers at Oxford University (Dr. Nicola Jr.) that allow us to model the behaviour of lipid rafts and to investigate the effects of anomalous diffusion and the linking of kinetics on the plasma membrane with cascading reactions such as MAPK.
- Developing the effects of transcriptional and translational delays in genetic regulatory systems.
- Building mathematical models from imaging data, with the Teasdale, Hamilton and Parton labs.
- Developing spatial models that capture complex chemical kinetics within the cell.

**KEY PUBLICATIONS**


This group works on developing simulation and visualization methodologies for understanding the behaviour of complex cellular processes, both on the plasma membrane, in the cytosol and at the genetic regulatory level. The simulation models take into account stochastic effects, while the visualization focuses on two- or three-dimensional displays.
Expression Genomics

The central focus of the IMB’s expression genomics lab is to globally survey genomic, transcriptomic and epigenomic information and then use this data to define the underlying molecular networks controlling key biological processes (such as cell division and differentiation) and pathological states (breast and pancreatic cancer). These systems-wide studies give us the opportunity to identify both the key genes driving specific phenotypes and also the chance to recognise the different layers of control guiding biological states. It also provides a strong foundation from which to study novel genome biology (such as the role of miRNAs, non-coding RNAs, retrotransposons, RNA editing etc.). As the capturing of “omic” data is a key component of our research, we are actively pursuing the use of microscopy-based profiling, automated in situ hybridisation screening and next-generation sequencing technologies for these studies.

For more information on our research and details of the research projects listed below, please see our webpage at


RECENT PUBLICATIONS


High throughput screens for applications such as drug and genomic discovery are leading to massive image sets in need of new methods of analysis. Further, live cells may now be imaged in 3D over time with the interactions and dynamics of multiple proteins observed at high resolution. The core of my group’s research is to develop the methodologies and tools needed to enable the full benefit of these rich new data sources to be realised.

Recent research has focused on automated classification, clustering and visualisation of high-throughput microscopy imaging. Towards this, the Automated Subcellular Phenotype Classifier (ASPiC) was developed by combining novel image statistics created in the group with machine learning methodologies to enable rapid classification of high throughput imaging with near-perfect accuracy. The approach will enable whole-proteome imaging to be analysed in days rather than months. Building on this, the iCluster methodology currently being developed allows the clustering, differentiation and visualisation of high throughput image sets to enable sense to be made of the vast sets being generated. A recent highlight has been the creation of a more sensitive statistical test to enable the automated detection of subtle differences between treated and untreated cells.

Towards the analysis of 3D and 4D bio-imaging, the group has been developing two streams of research. The first is in quantification, to extract the key parameters that describe the systems being observed. In this area we have developed the Object Based Colocalisation (OBCoL) system to segment and quantify individual structures from 3D and 4D whole-cell imaging. This approach has enabled the detailed analysis of spatial distribution of proteins on individual subcellular structures and their true diversity to be seen for the first time.

The second is in building mathematical models of the subcellular systems observed based on the quantification methodologies of first stream. For instance, dynamic geometric models based on live cell imaging have provided surprisingly detailed information and insights into the systems observed and have been used to predict biologically relevant and experimentally verifiable quantities such as pH change and solute concentration. Other areas of interest include modelling of recruitment and expulsion of proteins from membrane surfaces.

The group is strongly multidisciplinary and collaborative, with a focus on delivering methodologies and tools to be used by researchers. This year has seen the public release of iCluster and OBCoL, both available under open-source license via Institute-hosted websites.

KEY PUBLICATIONS


Modelling, visualisation and classification of live cell imaging

High throughput bio-imaging visualisation.

Automated classification of subcellular imaging.

LAB MEMBERS

Senior Research Officer: Dr Paul Leo
Research Officers: Dr Briska Gardiner, Dr Nicole O’connor, Dr Gabriel Kolo, Dr Nicola Wadell, Dr Logan Walker, Dr Ehsan Nourbkash
Senior Research Assistants: Graeme Bathol, Anita Starbucks
Research Assistants: Milana Gongora, Shivaqsi Wani
PhD Students: Geoff Faulkner, Melissa Brown
Masters Students: Rathi Thagaparan, Alay Pharmar
Honours Student: Alan Robartson

RECENT PROJECTS
• Studying mammalian transcriptomes at single nuclear resolution
• Predicting the function of mRNA-mRNA networks
• Defining the complete repertoire of genetic damage driving development and progression of breast cancer in a mouse model
• Studying tempsns-spacial transcriptome dynamics at histological resolution

LAB MEMBERS

Research Officers: Dr John Belward, Dr Fawang Liu
PhD Student: Ahmed Arefin
Co-supervised PhD Students: Athadi Bustamam, Mitchell Stanton-Cook, Josefine Sprenger
Rnmos: RNA in mammalian evolution and development

We are exploring the thesis that the genetic programming of higher organisms has been fundamentally misunderstood for the past 50 years, because of the assumption that most genetic information is transcribed by proteins. It is now clear, despite the fact that only a small fraction encodes proteins, that the majority of the genomes of mammals and other complex organisms is transcribed in a developmentally-regulated manner, and that most complex genetic phenomena are RNA-directed. Working in conjunction with collaborators in the United States, Europe and Japan, we are working to characterise and understand the functions of the mammalian transcriptomes, and to validate the prediction that most genetic information in mammals is conveyed by RNAs that control differentiation and development. This includes the identification of small RNAs that regulate gene expression at various levels, including transcription, and to determine the expression patterns and function of the tens of thousands of longer noncoding RNAs that are dynamically expressed during differentiation in mammalian cells, including embryonic stem cells. Among our recent findings we have shown that it is possible, not likely, that most of the mammalian genome is under evolutionary selection, and demonstrated that the majority of long noncoding RNAs are expressed in the brain, many in precise cellular and subcellular locations, some of which are novel. We use advanced computational, visual and experimental methods, integrating in silico, in vitro and in vivo approaches. The outcomes of our research will be to expand our understanding of human evolution, genetics and development, with important practical implications in medicine, genetic engineering and programming of self-assembling systems.

Research Projects:
1. Bioinformatic prediction and experimental validation of new classes of small RNAs in animals
2. Analysis of the dynamic expression of long noncoding RNAs during the differentiation of embryonic stem cells, neural stem cells, muscle, macrophages, T-cells and developing tissues such as the male and female genital ridge, as well as the alteration of noncoding RNA expression in pathologial states such as cancer
3. Analysis of the subcellular location of noncoding RNAs to expand knowledge of existing cellular compartments and discover new ones
4. Targeted functional analysis of selected non-coding RNAs involved in developmental processes and neurogenesis
5. Analysis of the conservation patterns of noncoding regions in the mammalian genome and alignment on the basis of RNA structural rules
6. Deep sequencing of the small and large RNA transcriptomes in embryonic stem cells, and various tissues in mouse and human, as well as RNAs associated with chromatin modification complexes, transcription factors, RNA editing enzymes and DNA-RNA triplex structures in chromatin

KEY PUBLICATIONS


Computational genomics

We use advanced computational and data management methods to investigate similarities and differences among genomes and the gene products they encode. Our goal is to make rigorous quantitative inferences, at both global and fine scales, about how genomes, gene and protein families, regulatory networks and cellular functions have evolved and diversified. We are particularly interested in scalable approaches, including those based on Semantic Web technologies, approaches that let us interrogate diverse data types including molecular sequences and structures, signalling pathways, regulatory and molecular interaction networks, gene expression patterns, subcellular localisation and cellular function.

Genomes have diversified, both structurally and functionally, from shared ancestral states. We develop methods and analyze phylogenetic pipelines to reconstruct the trees of descent (phylogeny/evolution) and to study processes of change through time (evolutionary genomes) in bacterial pathogens, teloskasts and mammals. Within the Nuclear Receptors in Breast Cancer consortium we are responsible for expression inference and network information. We also collaborate in the Visible Cell® research project.

For more information on our group and our research projects, please see: www.imb.uq.edu.au/index.html?tpage=11671

Research Projects:
1. Automatic inference of vertical and lateral gene transmission, genetic recombination breakpoints, and molecular interaction networks in pathogenic bacteria
3. Computer discovery of novel miRNA targets in mammalian genomes
4. Integration and querying of molecular network and cellular structure information, and querying-over these data, using Semantic Web technologies
5. Software and data infrastructure for the Visible Cell®

KEY PUBLICATIONS


The endosomal/lysosomal system of mammalian cells is a highly dynamic organelle, and the trafficking pathways within the endosomal system are fundamental for a wide variety of key cellular processes. My group is developing cellular and computational approaches to identify novel mammalian proteins associated with the endosomal system.

The regulated movement of membrane receptors and ligands between the cell surface and intracellular compartments is vital to many cellular operations, including communication between cells and their environment. A major current focus of the group is the characterisation of the mammalian retromer complex. We have implicated this complex, using real-time microscopy and molecular interaction techniques, in the sorting of numerous membrane receptors, including EGFR, within the endosomal system.

Macropinocytosis is a regulated form of endocytosis that mediates the non-selective uptake of extracellular material. Macropinocytosis is highly relevant to many aspects of both normal cell function and disease with particular importance in tumour progression and metastasis and in many infectious diseases. Our recent work has focused on characterising macropinosome maturation by the recruitment of sorting nexins. We have determined that the regulation of phosphoinositides is central to macropinocytosis and leads to the recruitment of key effector proteins including the PtdIns(3)P-binding PX-domain family of proteins. This emerging protein family performs a range of critical biochemical actions within the mammalian endosome and we are keenly interested in the roles these proteins play. Currently we are undertaking a systems biology approach to examine the distinct stages of macropinocytosis.

Numerous infectious pathogens exploit specific endocytic pathways to invade the host. Characterisation of pathogen entry pathways is essential for understanding infectious diseases but has also proven to be a powerful tool for gaining insight into normal cellular processes. We are currently investigating the molecular details of these pathways and how they are modulated in response to infection with Salmonella, a leading cause of human gastroenteritis.

RESEARCH PROJECTS
- Host-pathogen interactions during Salmonella infection
- Maintaining and updating LOCATE: A Protein Subcellular Localisation Database - http://locate.imb.uq.edu.au
- Developing computational approaches to analyse image and real-time microscopy data
- Studying endosome dynamics, macropinocytosis and retromer
- Systems biology of the mammalian endosome

KEY PUBLICATIONS