**InterMine and HumanMine – enhancing the biomedical relevance of model organisms and enabling other data mining projects.**

Funding body: Wellcome Trust

Principal Investigator: G. Micklem

InterMine is a data warehouse system developed by the Micklem group and already in use for a number of model organism databases, including for yeast, fruit fly, nematode worm (as part of modENCODE), mouse and rat. As such, InterMine is a good system to use for developing cross-species inter-operability, and indeed this is the aim of the NIH-funded <InterMOD project –link to grant>.

This grant covers the continuing development of the InterMine framework, as well as the development of HumanMine - an extensive *Homo sapiens* genetics, genomics and proteomics resource built using the InterMine framework. The existence of HumanMine will both provide a new resource for human genetic research, and facilitate easier cross-organism research by creating improved links with model organism data.

**An infrastructure for platform technology in synthetic biology**

Funding body: EPSRC

Received by: Consortium of Imperial College, University of Cambridge, University of Edinburgh, Newcastle University, led by R. Kitney (Imperial College)

Exploitation of standard genetic parts from a range of organisms in order to design new systems is the basis of the field of synthetic biology. This is a very promising area of research, with the potential for both furthering our understanding of basic biology, and creating genetically engineered biological systems that have practical applications. This grant is aimed at doing a systematic analysis of bacterial genomes for useful genetic components, and creating a data warehouse resource containing this information, available for synthetic biology researchers. The role of the Micklem group within this collaboration is to analyse data from bacterial genomes, and set up SynBioMine - a data warehouse targeted specifically at the synthetic biology research community.

**Toxoplasma: Parasitic subversion of host functions**

Funding body: KAUST

Principal Investigator: J. Ajioka (U. Cambridge) & A. Pain (KAUST)

Co-investigators: V. Bajic (KAUST), S. Ali (KAUST), K. Lilley (U. Cambridge), G. Micklem (U. Cambridge)

Infection by members of the parasitic protozoan phylum Apicomplexa, such as *Plasmodium* (malaria) and *Toxoplasma* are major causes of morbidity and mortality worldwide. The problem is further compounded by the efficacy of drug treatment waning due to the emergence of drug resistance in the parasites in questions.

The purpose of this grant is to enable the application of genomic and proteomic methods to further our understanding of host-parasite interactions during *Toxoplasma* infection. Specifically, we aim to study the manner in which the parasite controls gene expression for growth and development, and how the parasite disrupts host function through protein secretion. Insights into these mechanisms have the potential to be used for therapy development.

**InterMOD: integrated data and tools to support model organism research**

Funding body: NIH/NHGRI

Principal Investigator: G. Micklem

Co-investigators: M. Cherry (Stanford University), J. Richardson

       (Jackson Laboratory), L. Stein (OICR, Toronto), S. Twigger

       (Medical College of Wisconsin), M. Westerfield (U. Oregon)

InterMOD is an international consortium composed of the InterMine team and five model organism databases: budding yeast, nematode worm, rat, mouse and zebrafish. The aim of this project is to improve the infrastructure available for cross-species research, in order to facilitate analyses and make tools more widely accessible. It aims to do so by establishing a common framework for the different model organism databases based around the InterMine system [REF], implementing shared standards, and setting up a series of interoperability links, in order to reduce the time and effort required for performing cross-organism analysis.

**Sequencing the genome of the dinoflagellate *Symbiodinium*, a symbiont of Red Sea corals**

Funding body: KAUST

Principal Investigators: G. Micklem (U. Cambridge) & C. Voolstra (KAUST)

Co-investigators: T. Ravasi (KAUST), V. Bajic (KAUST)

Dinoflagellates are ubiquitous marine and freshwater protists. As free-living photosynthetic plankton, they account for ~50% of the primary productivity of oceans and lakes. As photosynthetic symbionts, they provide essential nutrients to corals that are the architects of one of the most productive ecosystems: coral reefs. Dinoflagellates are adapted to a wide variety of environments as reflected by a tremendous diversity in form and nutrition. Additionally, they play important roles as parasites and predators, and form the evolutionary sister group to the apicomplexans that are best known for being human and animal pathogens (e.g. Plasmodium as the agent of malaria).

The sequencing of a dinoflagellate genome will not only inform us about the supposedly enormous gene repertoire of dinoflagellates, it will also help us understand the capacities, weaknesses, and evolution of parasitism and mutualism. Furthermore, it will aid in explaining some remarkable features of dinoflagellate biology such as their unique genome structure and gene regulation. Additionally, dinoflagellates are directly associated with coral bleaching, red tides, and paralytic shellfish poisoning. Together these data will help establish an unprecedented perspective on the evolutionary dynamics of this mutualistic relationship and the genes and pathways involved in inter-species communication.

**The metabolicMine project: integrated data and tools for the Common Metabolic Disease community**

Funding body: Wellcome Trust

Principal Investigator: G. Micklem (U. Cambridge)

642,389

metabolicMine ([www.metabolicmine.org](http://www.metabolicmine.org)) is a data warehouse containing a selection of data from rat, mouse and human relevant to research into common metabolic diseases such as diabetes. The datasets include GWAS, information on disease phenotypes and associations across species, as well as standard genomic data such as gene functions and interactions. metabolicMine also includes all standard InterMine functionality, such as report pages, list analysis and flexible querying, with the added special feature of a regions search tool, The project was created with close input from research collaborators, and is hoped that the integrated data combined with the analysis tools will be of utility to the common metabolic research community.

**modENCODE DCC: A Data Coordination Centre for the model**

**organism ENCODE project**

Funding body: NIH/NHGRI:

Principal Investigator: Lincoln Stein (OICR, Toronto)

Co-investigators: S. Lewis (LBNL, Berkeley), J. Kent (UCSC), G. Micklem (U. Cambridge)

The modENCODE project is a large-scale international initiative to characterise the functional genomic elements in the model organisms *Drosophila* and *C. elegans* through a series of coordinated high throughput experiments. The role of the Data Coordination Centre (DCC) is to provide data management support in terms of data storage, processing, validation, metadata collection and making the data from the project publicly accessible.

The Micklem group created modMine (link), an InterMine based data warehouse for accessing and analysing the modENCODE data. It includes standard InterMine features such as list analysis and flexible query capabilities, as well as modENCODE specific resources, such as a gene interaction browser and a region search tool.