

# There are some questions that can't be answered by Google

Jeremy Barker<sup>1</sup> and Dominique Gorse<sup>2</sup>

<sup>1</sup>Queensland Facility for Advanced Bioinformatics, Brisbane, Australia, j.barker@qfab.org

<sup>2</sup>Queensland Facility for Advanced Bioinformatics, Brisbane, Australia, d.gorse@qfab.org

Today's research questions require an integrative and collaborative approach to analyse the research data generated by different experimental platforms within the wealth of publicly available resources. In that context, the Queensland Facility for Advanced Bioinformatics (QFAB) helps life science researchers to unlock the full value of their research data by providing access to up-to-date integrated public datasets and by developing customised collaborative bioinformatics solutions. An eResearch platform has been developed to collect, share, integrate, analyse, store and retrieve life science research data. It is based around 4 main components:

- 1) Data integration technology, such as the Sequence Retrieval System [1], that provides a flexible mechanism for running complex queries across disparate private and public data sources,
- 2) Advanced data mirroring that enables the researchers to analyse their own data against public databanks while preserving confidentiality,
- 3) Workflow engine that allows the rapid elaboration of complex pipelines that combine local and remote data and tool web services.
- 4) Collaborative environment, such as Plone, that facilitates the sharing and collective publishing of project information and data [2].

In this presentation, we describe a range of research projects that have benefited from this platform and demonstrate the value of eResearch to the researchers in driving their research further. These examples will include:

## THE ARACHNOSERVER

Professor Glenn King from the Institute for Molecular Bioscience was limited in his research on spider toxins because of the lack of common nomenclatures and the absence of comprehensive curated resources. The ArachnoServer was developed to enable authorised researchers to collectively manage, curate and publish the toxin records [3]. A flexible database schema was designed to capture a wide variety of toxin related data such as the genomic and proteomic sequences, post-translational modifications, targets and activities. Furthermore, a nomenclature for spider species maintained by the Queensland Museum was integrated while a new ontology for molecular target was proposed. The Arachnoserver enables a wide range of biological scientists from all around the world, including neuroscientists, pharmacologists, and toxinologists, to readily access key data relevant to their discipline and to bring together their expertise, providing what is now the largest curated database on spider toxins. Data mining of this information resource will contribute to discoveries in pharmaceutical and environmental research.

## THE UCSC GENOME BROWSER MIRROR

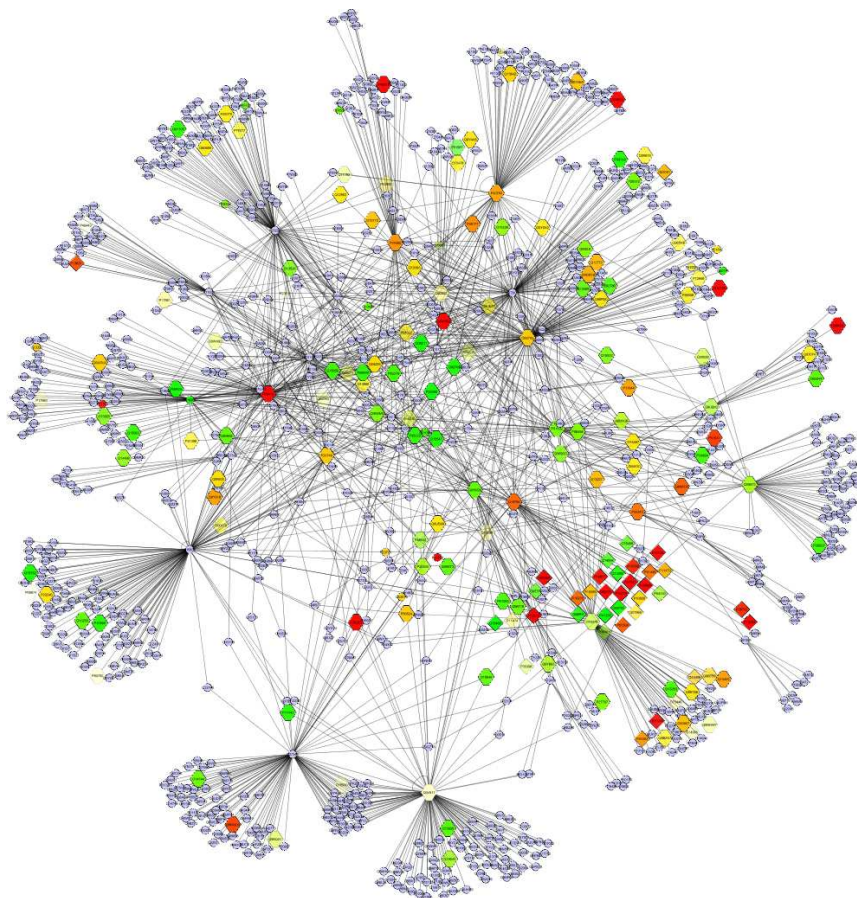
The UCSC genome browser is a world-leading application for visualizing and browsing annotated animal genomes. The database allows researchers to browse, query and visualize the genome assemblies and annotations of over 45 species. Researchers from the University of Queensland, Griffith University, Queensland University of Technology and CSIRO needed to have local access to such data and tools to support their leading-edge research. A full local mirror is now available with fortnightly updates [4]. The main value for the researchers is that each of the contributing partner organizations has access to private and secure hosted genome browser installations. These private mirrors have been optimized to allow the easy uploading of private annotations datasets and can be visualized as permanent tracks in the browser. In addition to the benefits gained from using the web-accessible UCSC tools, partner organizations also have access to a dedicated computing cluster that is integrated with the annotation database allowing advanced programmatic querying and computational analysis.

## THE KOALA CHILDHOOD OBESITY PROGRAM

The KOALA Childhood Obesity Program is led by Dr Gary Leong, a paediatric endocrinologist and researcher who divides his time between the Institute for Molecular Bioscience (IMB) and the Mater Children's Hospital. The Program is designed to improve the understanding of the basis of childhood obesity and its metabolic complications in order to develop effective interventions. To achieve such objectives, the researchers need the integration of studies in metabolism, genetics, environment (family and lifestyles), community influences and behaviour management. The KOALA program brings together researchers from the University of Queensland and a multidisciplinary clinic at the Mater Children's Hospital that includes medical, nursing, psychological and allied health professional team members. QFAB has developed a web portal [5] to capture data from participating families, health professionals and researchers, and continues to work with the project toward its goal of identifying biomarkers for the future onset of obesity-related illnesses.

## NUCLEAR RECEPTORS IN BREAST CANCER

The objectives of this discovery research program are to identify the nuclear receptor networks active in breast cancers and from these to select receptors that may be targeted for treatment in patients whose breast cancer does not express the estrogen receptor and for whom current treatment options are severely limited. The Chief Investigators of the NRBC project are drawn from virtually every State in Australia and contribute complementary expertise in hormone research, bioinformatics, high throughput analyses, human tissue research, animal models and clinical endocrinology. A collaborative environment based on Plone and SRB was developed to facilitate the exchange of idea and the sharing of data [6]. Furthermore, with Professor Mark Ragan, who leads the ARC Centre of Excellence in Bioinformatics and the Division of Genomics and Computational Biology of the Institute for Molecular Bioscience at the University of Queensland, QFAB is currently building one of the most comprehensive collections of protein interaction networks around Nuclear Receptors and their co-regulators to visualise gene expression data and facilitate decision support (Figure 1).



**Figure 1: An example of a protein-protein interaction network built around the pregnane X receptor (PXR). Visualisation of the network is performed using the Cytoscape software. The shapes represent different classes of proteins, the colour relates to the differential gene expression between two clinical cohorts.**

## REFERENCES

1. Sequence Retrieval System (SRS), <http://www.biowisdom.com>
2. Plone, <http://plone.org>
3. ArachnoServer, <http://www.arachnoserver.org>
4. UCSC Genome Browser mirror, <http://genome.qfab.org>
5. Koala Childhood Obesity Program, <http://koala.imb.uq.edu.au>
6. D.Gorse, M. Marshall1, D. Wood, N. Sim, A. Alexander, F. Eilert and J. Barker, *Collaboration Interface for Advanced Bioinformatics*, eResearch 2008, Melbourne, Australia