



Bio-grids and Applications

Jeremy Barker

QUEENSLAND FACILITY FOR ADVANCED BIOINFORMATICS
www.qfab.org.au

June 2007



Outline

- * QFAB overview
- * QFAB supplying data and tools for grids
- * Case study

June 2007



QFAB Overview

- ✳ Queensland Smart State grant of \$1.9M under the National and International Research Alliances Program
- ✳ Matched by partners – total \$3.8M
- ✳ Agreement to December 2009
- ✳ Collaboration of its partners, hosted at the University of Queensland

June 2007



QFAB Vision

- ✳ To build a bioinformatics facility to deliver advanced solutions to support the biotechnology, ICT, research biology and clinical communities.

June 2007



Partners

✧ Current

- ARC Centre in Bioinformatics
- e-Health Research Centre
- SRC for Functional and Applied Genomics
- Queensland University of Technology
- Griffith University
- Department of Primary Industries & Fisheries
- Australian Partnership for Advanced Computing
- Queensland Cyber Infrastructure Facility

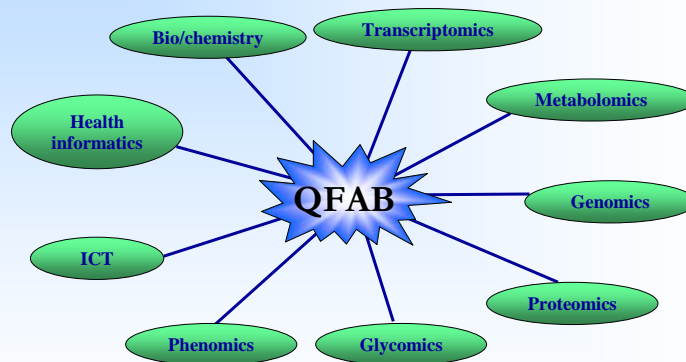
✧ Future



June 2007



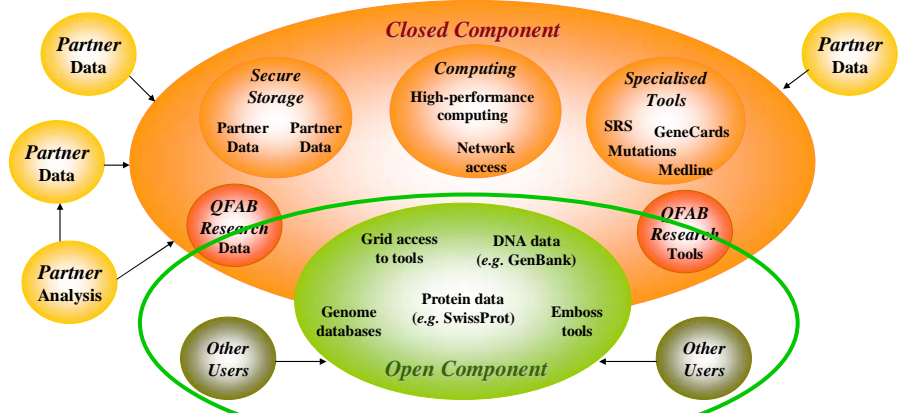
Catalyst for multidisciplinary research



June 2007



QFAB Operational Structure



QFAB provision of resources for partners and collaborative research



Grid Technology – our view

- * Provides access to distributed data, information and services
 - large data collections in collaborative environments
 - compute intensive services
- * Transparency of physical location and/or provider of resources
 - remote file operations where the user is unaware of where physical file is stored
 - access to one of many services offered by multiple suppliers



QFAB and Partners: enabling Bio -Grids

9

- * Maintaining public domain data sets and tools
 - Automated updating of public datasets
- * Making data available
 - Data put onto a grid with meta-data tags
- * Integrating private and public data
 - Linking proprietary and public data without compromising IP ownership
- * Providing web service interfaces to data and tools
 - Services for grids to access data (GENBANK, genome data) or to run compute intensive tools (BLAST, CLUSTAL)
- * Workflow engines for using services

June 2007



Computing and Storage

10

- * In-house
 - Secure cluster available for partner projects
- * QCIF and APAC
 - Use of QCIF and APAC high performance computing for running analyses and indexing data
- * APAC Grid
 - QFAB will contribute data and tools to the APAC grid program

June 2007



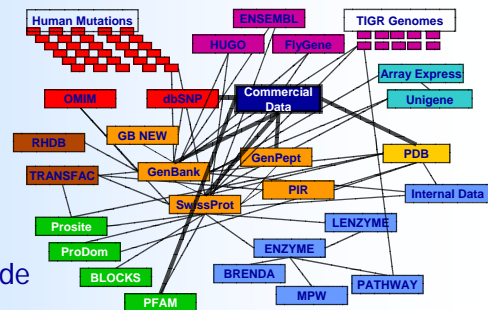
What is QFAB making available to the grid?

June 2007



SRS

- * Commercial tool used extensively in the pharma and biotech industries worldwide
- * Integrates partner data with public data
- * Over 1000 different data sets available for inclusion
- * Feature packed web browser for quickly searching and viewing data
- * Analysis tools available, with links to public data
- * Automated updating of public data
- * Web services interfaces for building custom applications



12

June 2007



Reactome for Visible Cell™

The screenshot shows a web browser window displaying the Reactome database interface. The main content area is titled "Entry view - REACTOME:19358-4" and features a list of "Highest Order Pathways in Reactome". The pathways listed include Apoptosis, Cell Cycle, Mitotic, DNA Replication, Epidermal Growth Factor Receptor (EGFR) signaling, Gap junction trafficking and regulation, HIV Infection, Immune System signaling, Insulin receptor mediated signaling, Maintenance of Telomeres, Metabolism of carbohydrates, Metabolism of nucleotides, Notch Signaling Pathway, Post-translational modification of proteins, Transcription, and mRNA Processing. To the right of the list is a 3D visualization of a cell model, rendered in various colors (red, purple, green, blue) to represent different cellular components and pathways. The QFAB logo is visible in the bottom right corner of the interface.

Showing highest order pathways

June 2007



Reactome for VisibleCell™

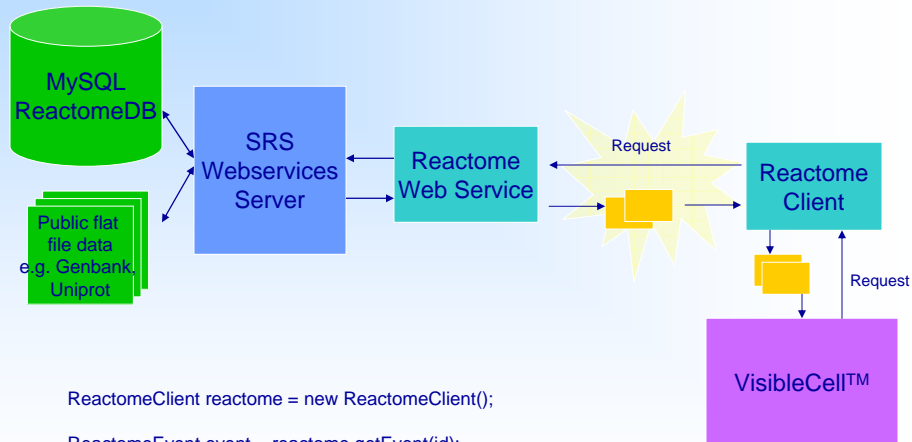
✱ Integrates

- Reactome pathway, reaction, gene ontology (GO) data
- Visualiser interface of the ACB VisibleCell™ project

June 2007



Reactome Web Service



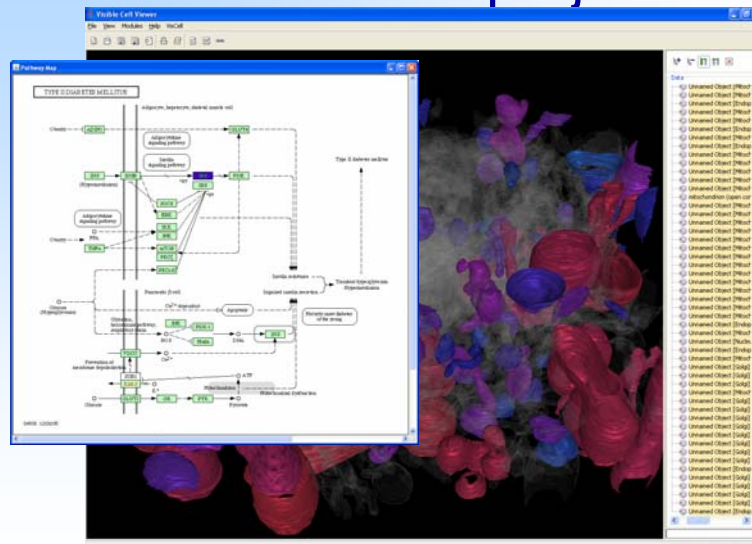
```
ReactomeClient reactome = new ReactomeClient();
```

```
ReactomeEvent event = reactome.getEvent(id);
ReactomeComponents[] components = reactome.getComponents(id);
```

June 2007



ACB VisibleCell™ project



June 2007



Storage Resource Broker (SRB) at QFAB

- ✱ Data grid for sharing data on an international scale
- ✱ Provides user with a single file hierarchy for data distributed across multiple storage systems
- ✱ At QFAB since March 2007
 - Currently at IMB and Griffith University
- ✱ Communication to more nodes planned

June 2007



Storage Resource Broker (SRB) at QFAB

- ✱ Maintain up-to-date copies of public data in SRB
 - Data sets: GENBANK, GO, PDB, UNIPROT etc
 - SRS updates data and automatically executes command to update SRB and make files available
 - Next step: Automatically add meta data to files
- ✱ UCSC and ENSEMBL genome datasets
- ✱ More data sets can be added as requested

June 2007



GPFlow: Bioinformatics workflows

- ✱ QUT Workflow engine technology
 - Builds and runs bioinformatics workflows for in-silico genomics experiments
 - Designed to meet the scientists at the bench without the need to write code
 - Interactive and immediate
 - Built on Microsoft business technology

June 2007



GPFlow – no code required

```
WsSessionImpl wsSession = WsSessionManager.GetWsSession(url)
```

```
SrsObject options = wsSession.GetToolOptions("Clustal", "");
```

```
String sequence = "CGATCTGATGCTGTAGCTGACTCGAT";
string jobId = wsSession.LaunchApplication(options, sequence, null, -1);
Console.WriteLine("Job Completed - ID: " + jobId);
```

```
SrsObject[] results = wsSession.GetQueryResults(jobId, "ClustalLoader");
```

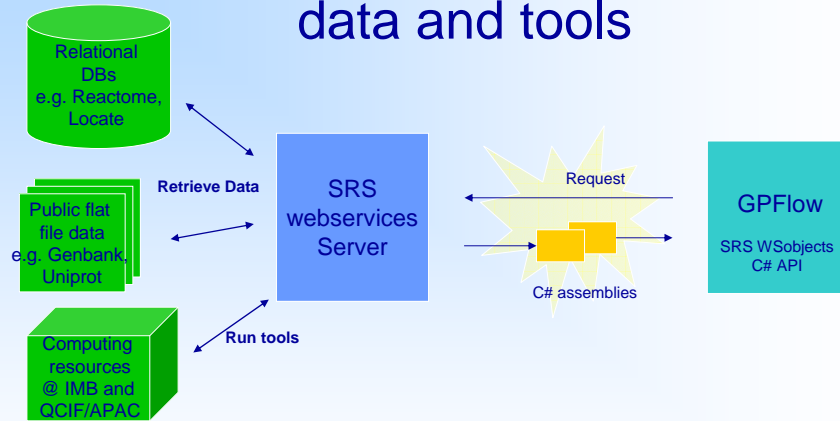
The screenshot shows the GPFlow web interface. On the left, a list of workflow steps is visible, including '1. Get homologs', '2. Get NCR', '3. Get promoters', '4. Run Meme', '5. Build tree', and '6. Run FootPrinter'. The main area displays the 'Result preview' for a motif found by Meme. The motif is 'CGATCTGATGCTGTAGCTGACTCGAT' (Motif 1 of 30, Parsimony = 0.0, Average offset = -71, std. dev. = 1). Below the motif, a table shows the alignment of this motif with surrounding DNA sequences from various genes (CAB665, CCA00695, TC0593, CT319, CP0698) and their locations. The 'Area surrounding motif' column shows the context of the motif within each sequence. At the bottom, there is an 'Output files' section.

name	Location	Area surrounding motif
CAB665	-73	aagaaatctGTATTcttattctctatctCGATCTGATGCTGTAGCTGACTCGATtagt
CCA00695	-71	aagaaatctGTATTcttattctctatctCGATCTGATGCTGTAGCTGACTCGATtagt
TC0593	-71	aagaaatctGTATTcttattctctatctCGATCTGATGCTGTAGCTGACTCGATtagt
CT319	-70	aagaaatctGTATTcttattctctatctCGATCTGATGCTGTAGCTGACTCGATtagt
CP0698	-71	gTTAAAtcagtgattcttattctctatctCGATCTGATGCTGTAGCTGACTCGATtagt

June 2007



GPFlow accessing QFAB data and tools



June 2007



Bringing it Together

June 2007



Case Study

- ✱ Using bioinformatics to underpin a research partnership to identify a regulatory pathway important in a specific cancer and hence identify targets
 - Researchers wish to identify all sources of information about a particular cancer
 - link these to clinical and experimental data to develop a validated model of action
 - ultimately identify key drivers which may lead to drug targets.

June 2007

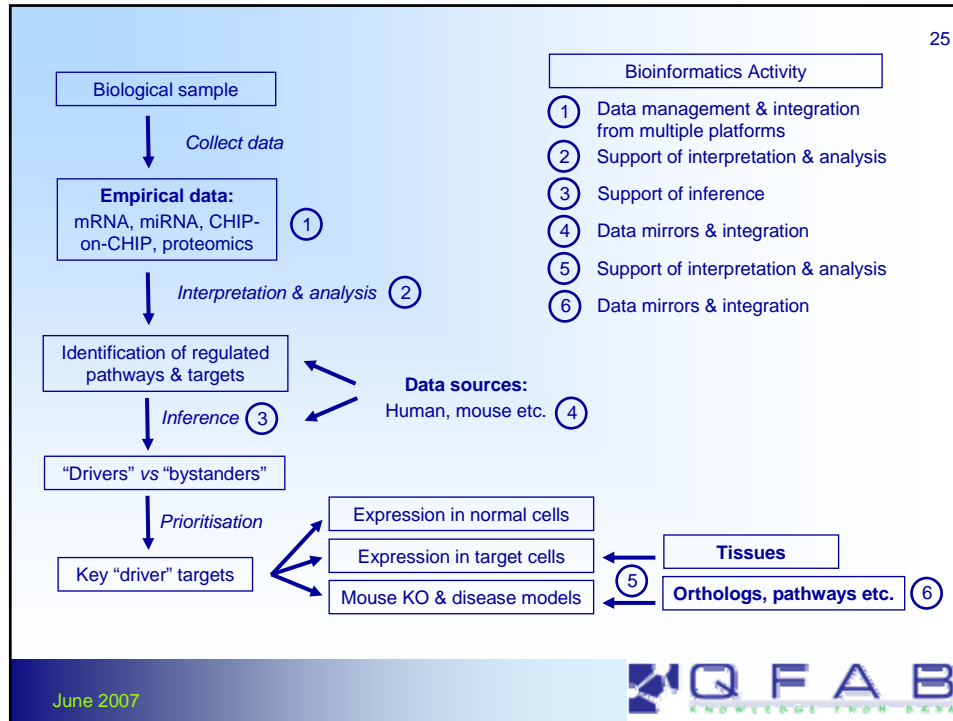


Key steps and challenges

- ✱ Identify sources of information about the cancer, including patient data
- ✱ Link to and integrate known data sources – genomic, proteomic, biomedical imaging, chemoinformatic, molecular modelling
- ✱ Develop model pathways and mode of action
- ✱ Test experimentally
- ✱ Validate model
- ✱ Identify key drivers and influences

June 2007





QFAB meeting these needs

- * Portal to provide researchers with access to data
 - Federated data from SRB
 - Tools to import data from dispersed technology platforms e.g. linking into the DART/ARCHER project at Monash
- * Secure data storage and access, authentication protocols
- * Develop ontology to enable reasoning across data sources
- * Data integration tools
 - SRS for public datasets
 - HDI for private clinical data integration
- * Providing or linking with data analysis tools for model development, testing and validating
- * APAC and QCIF Computing power

Bio-grid and QFAB applications

- ✱ QFAB is making data and tools available for grids to use:
 - Data already available via SRB grid applications
 - SOAP based services available for inclusion into Grids
- ✱ Underpins large collaborative research projects asking the big questions

June 2007



Contacts

www.qfab.org.au

Jeremy Barker
Chief Executive Officer
j.barker@imb.uq.edu.au
(07) 3346 2611

Dominique Gorse
Technical Manager
d.gorse@imb.uq.edu.au
(07) 3346 2024

June 2007

