

The Genopolis Microarray Database

Splendiani A (1), Brandizi B (1), Even G (2), Ottavio B (2), Pavelka N (2), Pelizzola M (2),
Mayhaus M (2), Foti M (2), Mauri G (1), Ricciardi-Castagnoli P (2)

(1) Dept. of Informatics, Systemistics and Communication,
University of Milano-Bicocca.

(2) Genopolis Consortium, Dept. of Biotechnology and Bioscience,
University of Milano-Bicocca.

Motivation

Gene expression databases are key resources for microarray data management and analysis. Public repositories as well as microarray database systems that can be implemented by single laboratories exists. However, there is not yet a tool that can easily support a collaborative environment where different users with different rights of access to data can interact to define a common content. The scope of the Genopolis database is to provide a resource that allows different groups performing microarray experiments related to a common subject to create a common coherent knowledge base and to analyze it, while respecting confidentiality of information. The Genopolis database has been implemented as a dedicated system for the scientific community studying dendritic and macrophage cells functions and host-parasite interactions

Methods

All the data have been generated through the Affymetrix platform, and all experiments are annotated following MIAME recommendations. Experiment annotation is realized through a custom software. At the core of this software is set of objects that represent entities relevant to the experiment annotation, such as Experiment, Source, Stimulus, Sample, Hybridization, Measure. These objects are organized as a tree, and the system provides functions on this tree to navigate and check its components. Permission to edit and view these objects can be defined at the object level with a group/role system of authorization. Furthermore the system supports the creation of controlled vocabularies by the users. Integrity and consistency of data and annotation is enforced through checking procedures. These are both automatic, on file integrity and required fields, and humanly supervised, as for controlled vocabularies definition. The system is based on a web architecture and is implemented in PHP and Java. In its current version it is based on MySQL and is deployed on a Linux/Apache redundant server with high availability features. Several kind of data are managed by the system. Raw images and cell files are managed as files, and made available for download to authorized users, while expression values and experiment descriptions are managed by the SQL engine and used for basic data analysis and advanced visualization. The system also offers an automated export to ArrayExpress, parsing of Affymetrix MAGE-ML description files and an advanced interactive user interface. These interface allows users to visualize data matrices based on functional lists and sample characterization, and to navigate to other data matrices defined by similarity of expression values as well as functional characterizations of genes involved. A collaborative environment is also provided for the definition and sharing of annotation by users (eg.: functional annotations of genes).

Results

The Genopolis database system provides an advanced resource for a scientific community investigating a common topic through microarrays . Roles can be defined so that measurements and experiment descriptions can be provided by different users. Consistency of data and annotations is enhanced by the system and common controlled vocabularies are built as the result of the interaction of users. Data can be kept confidential among groups for a limited time, and can be later made public to other users and on the ArrayExpress public repository. The content of the knowledge base can be exported for analysis with external tools, or browsed with an interactive graphical system that intuitively allows users to browse related set of genes and experimental conditions.

Availability: <http://www.genopolis.it/>

Contact email: andrea.splendiani@unimib.it

Supplementary informations

Access to data is subordinated to proper agreement.

The screenshot displays the Genopolis web application interface. At the top, the browser window shows the URL <http://gc-lab32.btbs.unimib.it/genopolistest/html/search5/index.php>. The application header includes the Genopolis logo and search filters: array type (MG-U74Av), measure type (MAS5 scaled), and sort by (experiment-source-stimulus-time). A "DISCOVER!" button is visible.

The main content area is divided into several sections:

- Select your gene:** A search box with a "search" button and a "clear - all" link.
- Select Stimulus/i:** A dropdown menu with options: unstimulated, Schistosoma mansoni eggs, Schistosoma mansoni SLA, Zymosan, CMV, and Leishmania mexicana promastigote.
- Exp. Source Stimulus Time:** A table defining experimental conditions:

Exp.	D1+Leishmania mexicana	D1+Schistosoma mansoni
Source	promastigote	eggs
Stimulus	Leishmania mexicana	Schistosoma mansoni
Time	4-h, 8-h, 12-h, 24-h	4-h, 8-h, 12-h, 24-h
- HitMap:** A heatmap visualization with a grid of 50 rows and 64 columns. The heatmap shows varying intensities of red, indicating gene expression levels across different conditions.
- Gene Details:** A detailed view of the **Traf4** gene, including its description (Tnf receptor associated factor 4), signal (89.1), call P, and p-value (0.009985). It also lists the experiment (D1+Leishmania mexicana promastigote), source (musmusculus_C57BL/6), stimulus (Leishmania mexicana promastigote), sample (12-hours), and hybridization (A). The gene is noted as part of the NF-kB family, involved in antigen processing and presentation, and apoptosis.

On the left side, a list of genes is shown, including **Cd3g** (CD3 antigen, gamma polypeptide), **Itih3** (inter-alpha trypsin inhibitor, heavy chain 3), **Ryr1** (ryanodine receptor 1, skeletal muscle), **5930412E23Rik** (RIKEN cDNA 5930412E23 gene), **Traf4** (Tnf receptor associated factor 4), and **I** (interferon alpha family, gene 6).