

GebbaMA: archive and management of microarray data

Rossi E⁽¹⁾, Giuliani S⁽¹⁾, Gamberoni G⁽¹⁾, Volinia S⁽²⁾, Frangiamone G⁽³⁾, D'Ascia S⁽³⁾

⁽¹⁾High Performance Systems Division, CINECA, Bologna

⁽²⁾DAMA, University of Ferrara, Ferrara

⁽³⁾NSI - Nier Soluzioni Informatiche, Castel Maggiore - Bologna

Motivation

GebbaMA is a tool, accessible on the web, for archiving and organizing data produced by microarray experiments and high throughput genomics experiments in general. It is ideal for laboratories producing large amounts of data, allowing the information to be archived in a systematic way. GebbaMA is also useful for final users for retrieving data and reorganizing them for their research activities. By simply connecting to the web site, a microarray laboratory can upload the experimental data in a common repository, annotate them on the basis of information extracted automatically and organize the data as hybridizations or projects. The researcher can search the archive, upload data obtained from other sources or public repositories, design new projects and download the aggregated information for further processing. GebbaMA is a network of federated archives. Each local archive can be used off-line with its own data or online, in which case the network can be exploited to facilitate collaboration and data sharing. GebbaMA can be installed locally, therefore assuring security and privacy. Depending on the user preferences, it can also be hosted on a central site in order to make use of more powerful and robust institutional resources.

Methods

The efficient storage of microarray data is of considerable interest and there is much activity worldwide. In general most researchers adopt a "single workstation approach" for data management, but this method is rapidly becoming inconvenient for many reasons. We have decided to use the following criteria for the implementation of the GebbaMA microarray system. - Users must be given the opportunity to use a wide range of common and user-friendly tools for data entry and for the different platforms available, e.g. Affymetrix, Agilent, Illumina etc. Data should be distributed - Data must be recorded in a format which allows interoperability of all the data sources To satisfy these criteria GebbaMA has been designed to support a complex data management architecture and implemented with the most modern technologies. The system consists of a "central" node and many "satellite" nodes, each of which with its own data store. The system has been designed in a modular way and will work even if the central node is unavailable. In fact in our schema "central" merely indicates a central registry for distributed indexing and querying. Data is stored, analyzed and exchanged through a complex architecture built upon Alfresco, a content management system that provides a common interface and access to distributed data sources. Alfresco also includes user authentication and various levels of access privileges, thus allowing many degrees of data security and privacy. We have used the Alfresco structure to build software modules in order to manipulate MAGE-ML files, extract metadata from MAGE-MLs, and to store and index metadata in the repository for querying microarray data according to different search criteria.

Results

We have demonstrated a user-oriented, robust infrastructure for microarray data management realized within a project funded by the Emilia Romagna region. It allows the user to enter data which can then be distributed to the other sites. The first version of GebbaMA is already available for downloading and microarray researchers are invited to contact the authors if they wish to experiment with the system. GebbaMA is also accessible on the centralised node of Cineca at this web address:

<http://gebbama-prod.cineca.it/MicroarrayRepository/MicroarrayRepository.htm>.

You can get general information about requesting a test account here gebbama.cineca.it Future enhancements to GebbaMA will be added during the Biopharmanet project www.biopharmanet.eu and will include the possibility of performing sophisticated analyses and for managing gene expression values contained in the archived files.

Microarray Repository

Control Panel

- Upload Management
- Hybridization Management
- Project Management
- User Management

Videate Aperte

Project Management

User Management

Hybridization Management

Upload Management

Laboratory

Current logged user: cineca1 cineca1

Hybridization Management (DEMO)

+ Simple Search ?

Stored

Stored: Riga 1 di 4

<input checked="" type="checkbox"/>	Hybridization Name	Associated Files	Link	Check
>	01_8769	xml;chp;cel;ori;	0	
>	02_3368254	xml;cel;ori;	0	
>	20061130_B0-02_U133Plus	xml;cel;	1	
>	C29_PG37	xml;cel;ori;	1	
>				
>				

Contact : info.bio@cineca.it