

# The NCI/CIT MicroArray Database (mAdb) System –Bioinformatics for the Management and Analysis of Affymetrix and Spotted Gene Expression Microarrays

**J. M. Greene<sup>1,2</sup>, E. Asaki<sup>1,2</sup>, X. Bian<sup>2,3</sup>, C. Bock<sup>2</sup>, S. Castillo<sup>1</sup>, G. Chandramouli<sup>2,3</sup>, R. Martell<sup>1,2</sup>, K. Meyer<sup>1,2</sup>, T. Ruppert<sup>1,2</sup>, S. Sundaram<sup>1,2</sup>, J. Tomlin<sup>2</sup>, L. Yang<sup>2</sup>, and J. Powell<sup>2</sup>**

<sup>1</sup>SRA International, Inc., <sup>2</sup>Bioinformatics and Molecular Analysis Section, Center for Information Technology, NIH and <sup>3</sup>Medicine Branch, Center for Cancer Research, National Cancer Institute, NIH

A scalable, modular, enterprise-level system for both microarray databasing and analysis over the Internet has been developed over the past four years by the National Cancer Institute's Center for Cancer Research in collaboration with NIH's Center for Information Technology. This completely Web-based system, called mAdb (for microArray database), is currently supporting over 810 registered users and collaborators at NIH and contains over 22,000 microarray experiments, making it one of the largest collections of microarray data in existence. In addition, the mAdb system has been ported for the Netherlands Cancer Institute, the Genome Institute of Singapore, and the CDC. This system has been used for a wide variety of scientific experiments spanning the range from cancer to studies of early development, and for human, mouse, rat, yeast, and numerous microbial organisms.

The mAdb system uses both CGI and Java applet user interfaces to the data, and requires only a Web browser for the end user. The mAdb system currently accepts two-color spotted array data quantitated with Axon's GenePix<sup>TM</sup>, PerkinElmer's QuantArray<sup>TM</sup>, BioDiscovery's Imagen<sup>TM</sup>, and NHGRI's ArraySuite II. Composite images and quantitated data files are uploaded to the system over the Web. The composite image file is used to display individual array spot images upon request, and the quantitated data file is parsed into Sybase tables. Implementations have also been developed for single channel, radioactive filter data.

Recently, the ability to upload data from certain human and mouse Affymetrix chips was added to the mAdb system, with more Affymetrix layouts to be added soon. Data must be analyzed with Affymetrix's Microarray Analysis System (MAS) 5.0 software with certain parameters depending on whether one is analyzing a single array or comparing a group of arrays. The CEL image file can also be uploaded.

Users can filter the spots based on quality parameters to create reusable dataset files, from which subsets can be created by the application of additional filters and/or analysis tools. This approach was designed to reduce database input/output and increase analysis tool performance. Data is aligned by clone ID to compare gene expression across different array layouts.

Open source tools (the R statistical language, ImageMagick, the Apache Web server, Java, and Perl)

have been used wherever possible. Each feature on the array generates a summarized report, consolidating information from Unigene, GenBank, GeneCards, and other genomic and pathway databases.

Features and analysis tools currently in mAdb include:

- Agglomerative hierarchical, K-means, and self-organizing map (SOM) server-side clustering
- Principal Components Analysis (PCA) and Multidimensional Scaling (MDS)
- Interactive scatter plot
- Multiple array graphical viewer
- PAM (Prediction Analysis of Microarrays) classifier
- Boolean comparison of datasets
- Array group assignment and averaging
- T-tests, Wilcoxon Rank Sum, ANOVA, and Kruskal-Wallis statistical analyses
- Pathways Summary report (BioCarta, KEGG, GO)
- Configurable data display
- Ability to refresh gene information in stored datasets
- History of filtering of data subsets
- Keyword query of datasets
- User management of data

To support additional analysis needs, data can be exported to Excel, tab-delimited files, or other analysis tools, such as Silicon Genetics' GeneSpring<sup>TM</sup>. Basic and intermediate, hands-on training courses teach users how to use the system.

Data is secured and backed up on a regular basis, and investigators can authorize levels of access privileges to their projects, allowing data privacy while still enabling data sharing with collaborators.

Efforts are now underway towards bringing mAdb into compliance with the MIAME standard and adding XML export of the data in MAGE-ML format.

Our project oriented design approach allows supporting multiple, independent research projects through one platform, yet permits researchers the opportunity to access a shared tool collection that would otherwise be difficult for many individual labs to implement and manage. We will demonstrate the analysis and data management capabilities of the system, and discuss how mAdb has helped our researchers reach their scientific goals.