

MicroArray Explorer – An Expandable Open-Source Java-based Data Mining Tool

Peter F. Lemkin, Ph.D.⁽¹⁾, Gregory Thornwall⁽²⁾,
Eric Shen⁽³⁾

¹LECB, NCI/FCRDC, Frederick, MD 21702,
lemkin@ncifcrf.gov; ²SAIC, NCI/FCRDC;

³Univ. Arizona

CHI-HealthTech Microarray Data Analysis, Sept 10-11, 2002, Washington, DC
Revised: 9-11-2002

Abstract

The Microarray Explorer (MAExplorer) is a versatile open-source Java-based data mining bioinformatic tool for analyzing quantitative DNA expression profiles across multiple microarray platforms and DNA labeling systems. Open-source means both the executables and source code may be freely downloaded, modified and redistributed. It is run as a stand-alone application on your computer and may be extended with plug-ins. With this program it is possible to: 1) analyze the expression of individual genes; 2) analyze the expression of gene families and clusters; 3) compare expression patterns; 4) directly access other genomic databases for gene subsets of interest. Spot data may be converted from GenePix, Scanalyze type spot quantified data using our Cvt2Mae data conversion “wizard”. Alternatively, you may download ready-to-run data sets from the NCI/CIT mAdb microarray Web server.

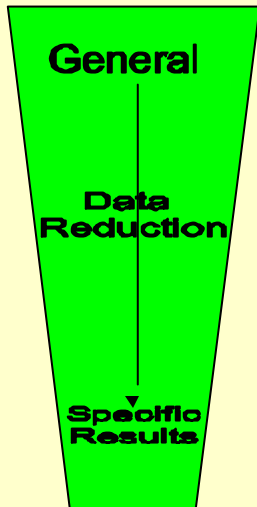
Analyses are performed in real-time, may be viewed and directly manipulated in pseudo-array images, scatter plots, reports, histograms, expression profile plots, and cluster analyses plots. The raw spot data may be normalized under a variety of methods. A key feature is the gene data filter for constraining a working set of genes to those passing the intersection of a variety of user-specified logical and statistical tests.

Resulting sets of genes may be saved as named sets and subsequent set operations may be performed resulting in derived gene sets. These in turn may be used in redefining the data filter. Condition sets of similar hybridized samples may be saved as named sets and used for reconfiguring experiment subsets. Reports may be generated with hypertext Web access to UniGene, GenBank, LocusLink, and other Internet databases for sets of genes found to be of interest. Users may save their exploration states on the local computer, and later recall or share them with other scientists. The emphasis on direct manipulation of genes and sets of genes in graphics and tables provides a high level of interaction with the data, making it easier for investigators to test ideas when looking for patterns.

Investigators may add their own types of analysis functionality by writing Java Plugins using the Open Java API which is fully documented on the Web site.

MAExplorer may be accessed and freely downloaded (runnable or source code) at <http://maexplorer.sourceforge.net/> with a mirror site at <http://www.lecb.ncifcrf.gov/MAExplorer>. This poster describes the MAExplorer plugin architecture

Array Data Mining - Finding Patterns of Genes



Quantified array spot data for multiple samples and replicates

Organize by: sample, gene expression, gene sets

Change views: normalization, data filters

Visualize and query: plots, cluster, reports

Explore: external genomic databases

Subset of genes for further analysis

What is MicroArray Explorer?

- MAExplorer is a Java microarray data-mining tool
- Developed initially for Mammary Genome Anatomy Program (MGAP) - Hennighausen (NIDDK)
- Handles multiple cDNA or oligo arrays with replicate spots
- Handles intensity or ratio (Cy3/Cy5) quantified array data
- Analyzes data for 2-conditions and expression profiles
- Data-filters gene sets by statistics, clustering, gene set
- Allows direct manipulation of data in graphics, spreadsheets
- Accesses genomic Web servers from plots and reports
- Oriented toward mRNA data, could extend to protein arrays
- Lemkin et al. *Nucleic Acids Res.* (2000) **28**:4452

Open Source at <http://maexplorer.sourceforge.net/>



The screenshot shows a web browser displaying the MicroArray Explorer website. At the top, there is a navigation menu with links for Home, Project Overview, Introduction, Tutorial, FAQs, Downloads, and Contact Us. Below the navigation menu, there is a section titled "Welcome to MAExplorer" followed by a paragraph describing the tool as a Java-based data-mining tool for cDNA or oligonucleotide microarray data. A list of features is provided, including handling multiple cDNA or oligo arrays, analyzing data for 2-conditions, and providing a graphical user interface. The website also includes a "SourceForge.net" logo and a "Download" button.

MAExplorer Download and Documentation

The screenshot shows the MAExplorer website interface. On the left is a navigation menu with links for Home, About, Tutorials, Examples, Download MAExplorer, Data conversion wizard, and Plugins. The main content area features the MAExplorer logo and the title "MicroArray Explorer for Data Mining Gene Expression Patterns". Below the title is a descriptive paragraph about the software's capabilities. Several yellow callout boxes with arrows point to specific elements: "Reference Manual" points to the "Reference Manual" link in the navigation menu; "Tutorials" points to the "Tutorials" link; "Examples" points to the "Examples" link; "Download MAExplorer" points to the "Download MAExplorer" link; "Data conversion wizard" points to the "Data conversion wizard" link; and "Plugins" points to the "Plugins" link.

MAExplorer Reference Manual

The screenshot shows the MAExplorer Reference Manual page. The page title is "Reference Manual - MAExplorer Microarray Exploratory Data Analysis". The main content area features a large heatmap visualization with a red arrow pointing to a specific cluster of genes. Below the heatmap is a table of contents for the manual, listing chapters from 1 to 17. At the bottom of the page, there is contact information for the author, Tomer E. Leshem, including his affiliation with the Laboratory of Experimental and Computational Biology at the Center for Cancer Research, Harvard Cancer Institute, and the Department of Pathology at Harvard Medical School. The page also includes a download link for the manual and a link to the MAExplorer website.

MAExplorer Downloads Available

Downloads for The MicroArray Explorer Project

You may download [program installers](#) for your particular computer for both MAExplorer and Cyt2Mar as well as the MAExplorer. If you are interested in the source code, that is also available. In desktop files it is available in the CVS directories listed in the table below (see [instructions on using CVS](#) to access these files directly with CVS). [Clicked for archives & packages](#) of the source code are also available.

If you want to upgrade your installation to the latest JAR files, simply download the JAR files and save them wherever you have installed the program replacing the previous jar files. A [Plugins.txt](#) file is available with all of the MAExplorer jar files. Simply replace the directory using Unzip or a Windows unzip program into a directory you can access when running MAExplorer. To let MAExplorer go directly to these files when you do a [Plugins] Load plugins menu command, copy the jar files into the Plugins directory where you previously installed MAExplorer.

You may also download the Mandatory Genetic Ancestry Program ([MGAP](#)) [public data set](#) that can be used with MAExplorer. There is a [list of PDF documents](#) describing MAExplorer and the Cyt2Mar data conversion wizard that may be downloaded.

See the [Release notes](#) for more information on what changes have been made and what new features are available.

Click on the entries to download the files.

1. Access of MAExplorer, MAExplorer.jar, Cyt2Mar from either Web server:

Program	Version/Program installers	Information on installing	Source	Jar file(s)
MAExplorer	0.96.12	MAExplorer	install MAExplorer	source code MAExplorer.jar
MAExplorer.jar	(not required)	Using MAExplorer	source code List of MAExplorer.jar	
Cyt2Mar	0.60	Cyt2Mar	install Cyt2Mar	source code Cyt2Mar.jar

Installing MAExplorer on User's Computer

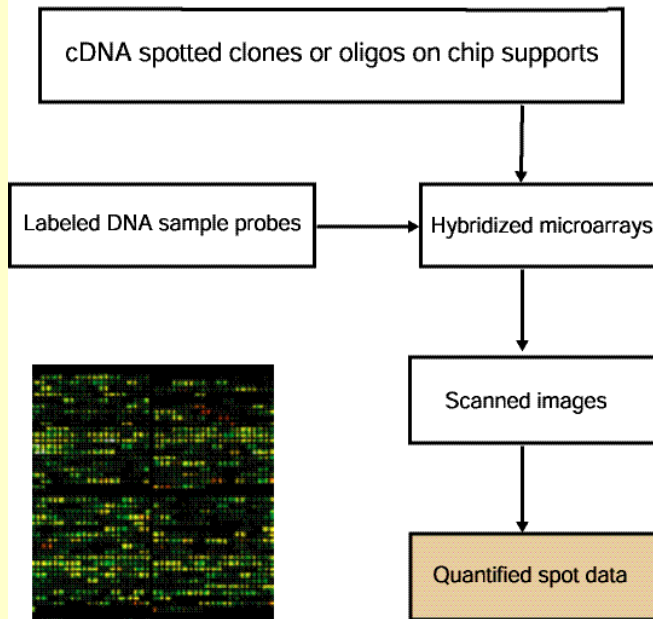
Version 0.96.12 - MAExplorer stand-alone application for data mining microarray data with MGAP data

Recommended installation for Your Platform:
 Windows (32-bit) Mac OS X (32-bit) Linux (32-bit) Solaris (32-bit) HP-UX Other Java-enabled Platforms

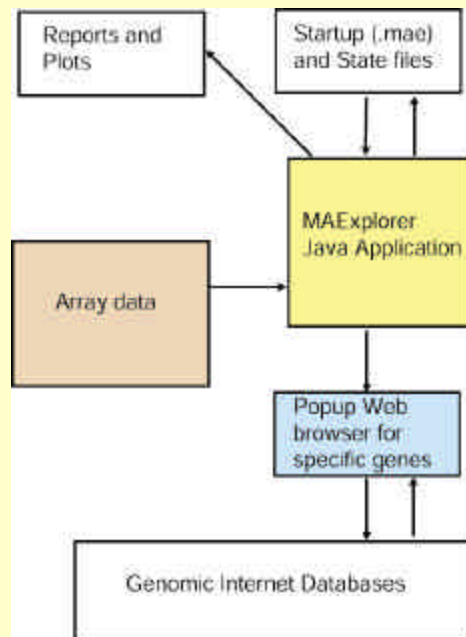
Available installers

Platform	Includes Java VM	Without Java VM	Instructions
Windows	Download (13.0M)	Download (13.0M)	View
Mac OS	Download (13.0M)	Download (13.0M)	View
Mac OS X	Download (13.0M)	Download (13.0M)	View
Linux	Download (13.0M)	Download (13.0M)	View
Solaris	Download (13.0M)	Download (13.0M)	View
HP-UX	Download (13.0M)	Download (13.0M)	View
Other Java-enabled Platforms	Download (13.0M)	Download (13.0M)	View

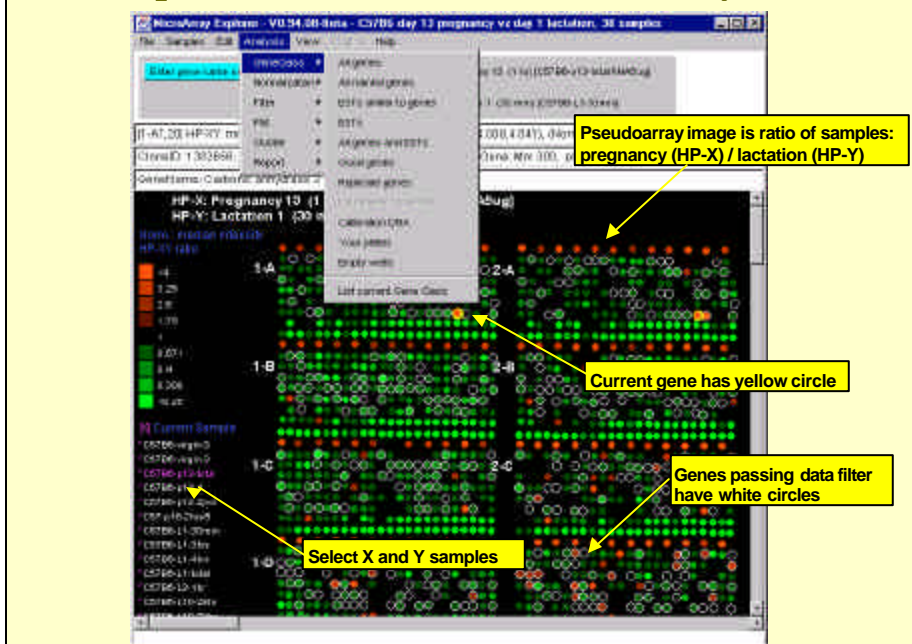
MAExplorer - Array Data Preparation



Paradigm: Local & Genomic-Web Databases



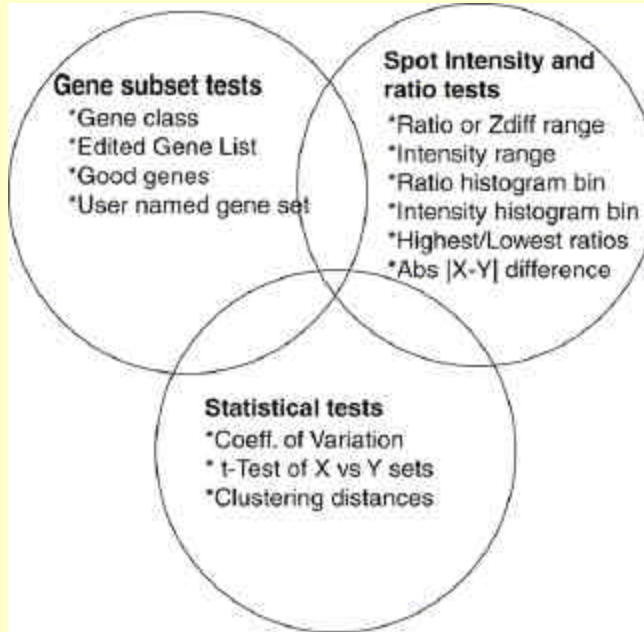
MAExplorer User Interface - Main Analysis Menu



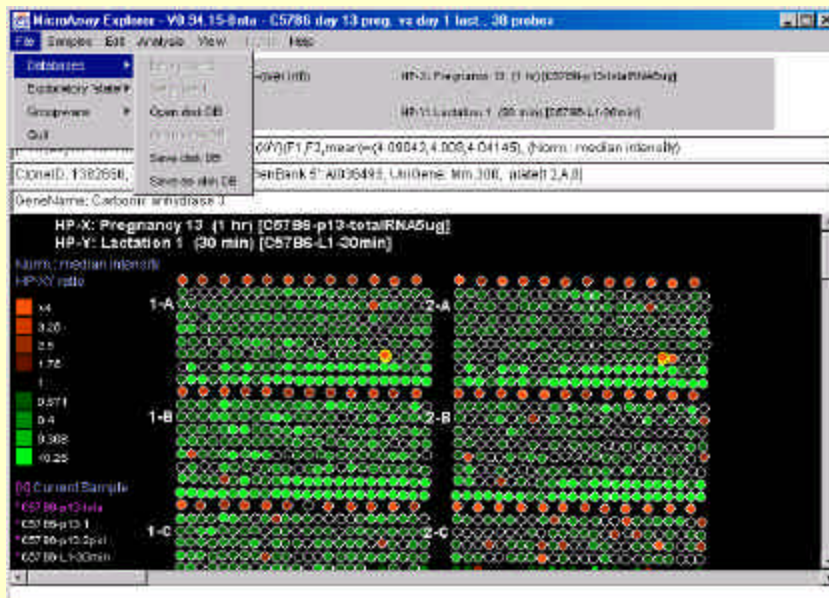
Multiple Data Filters - Finding a Gene Subset

- A data-filter is applied to all genes using selected filter tests
- It creates a working subset of genes used for subsequent clustering, plots, and reports
- Data filters compute intersection (AND) of sets from tests:
 1. Gene subsets defined in previous operations
 2. Spot intensity and ratio ranges
 3. Statistics: CV, t-Test
 4. Clustering: similar expr. profiles, K-means, hierarchical, dendrograms, clustergrams
- Changing filter parameters first recomputes the data filter and then recomputes all active plots or clustering methods

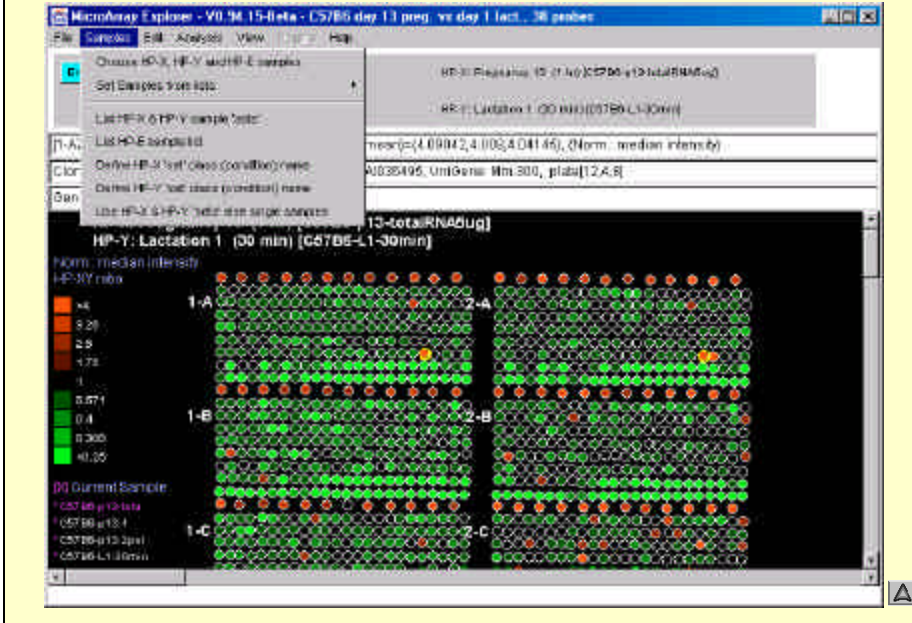
Gene Set Data Filter is Intersection of Tests



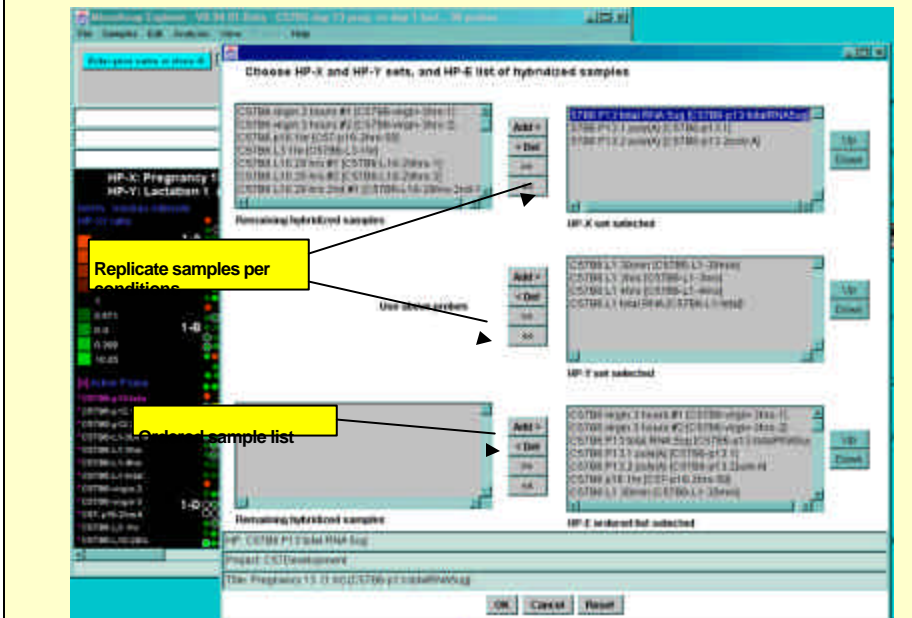
Menus: Files - Open and Save Exploration State



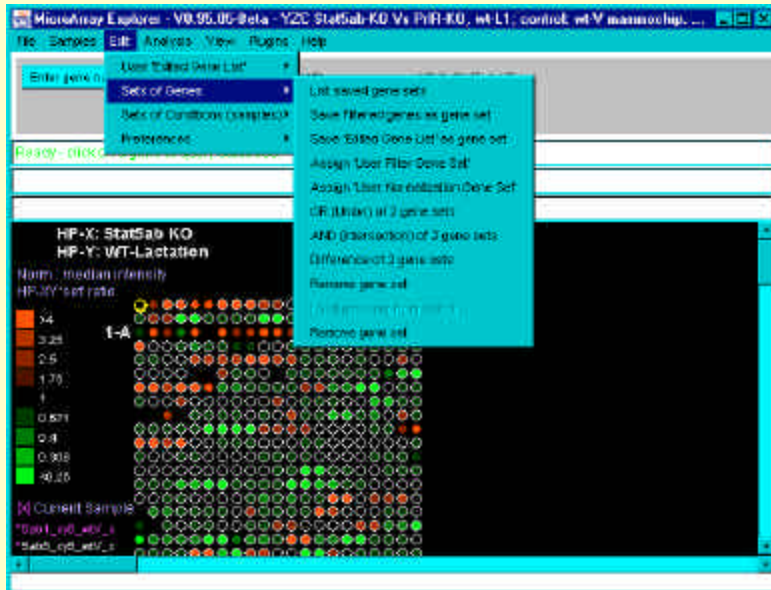
Menus: Samples - Select Condition Lists



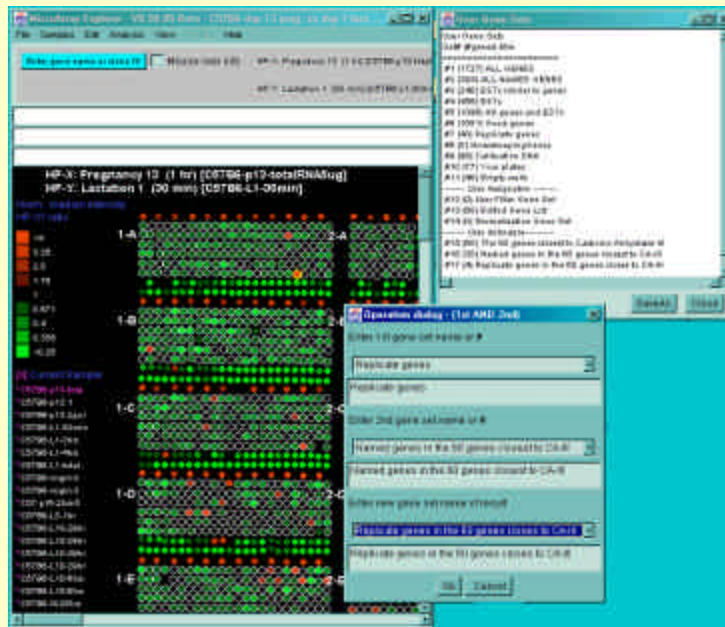
Sample Condition HP-X, HP-Y Sets & HP-E Lists



Menus: Edit - Gene Sets & Condition Lists



Gene Set Operations - e.g. 'AND' of Two Sets



Menus: Gene Class

The screenshot shows the MicroArray Explorer interface with the 'GeneClass' menu open. The menu options are:

- GeneClass
 - All genes
 - Normalization
 - All named genes
 - Filter
 - ESTs similar to genes
 - ESTs
 - Plot
 - All genes and ESTs
 - Cluster
 - Good genes
 - Replicate genes
 - Housekeeping genes
 - Empty wells
 - Set Gene Class subset
 - Report
 - List current Gene Class

The background shows a heatmap titled 'HP-X: Stat5ab KO' and 'HP-Y: WT-Lactation' with a color scale for 'Norm.: median intensity' ranging from <math>< -0.25</math> (red) to >4 (blue). The current sample is '*5ab1_oy5_wtV_c'.

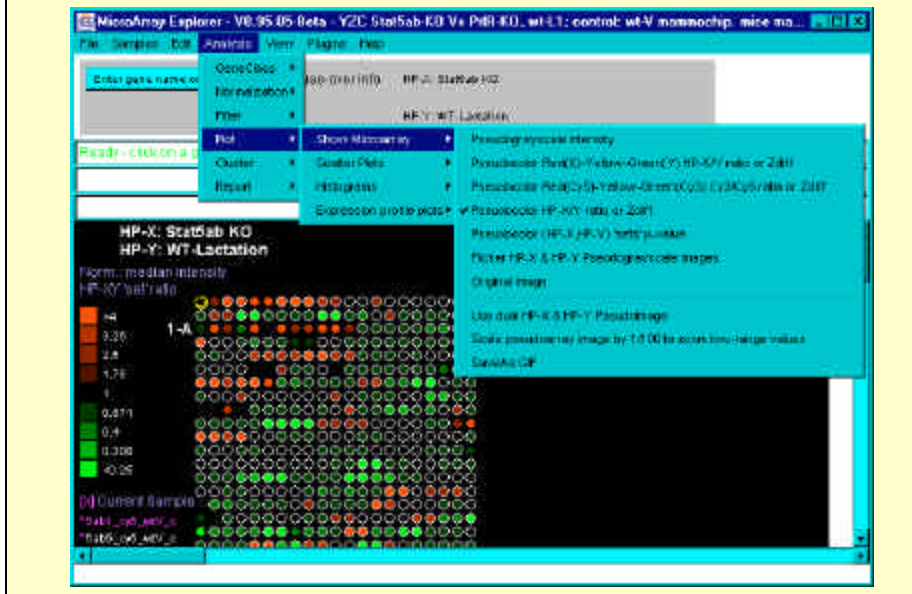
Menus: Normalization

The screenshot shows the MicroArray Explorer interface with the 'Normalization' menu open. The menu options are:

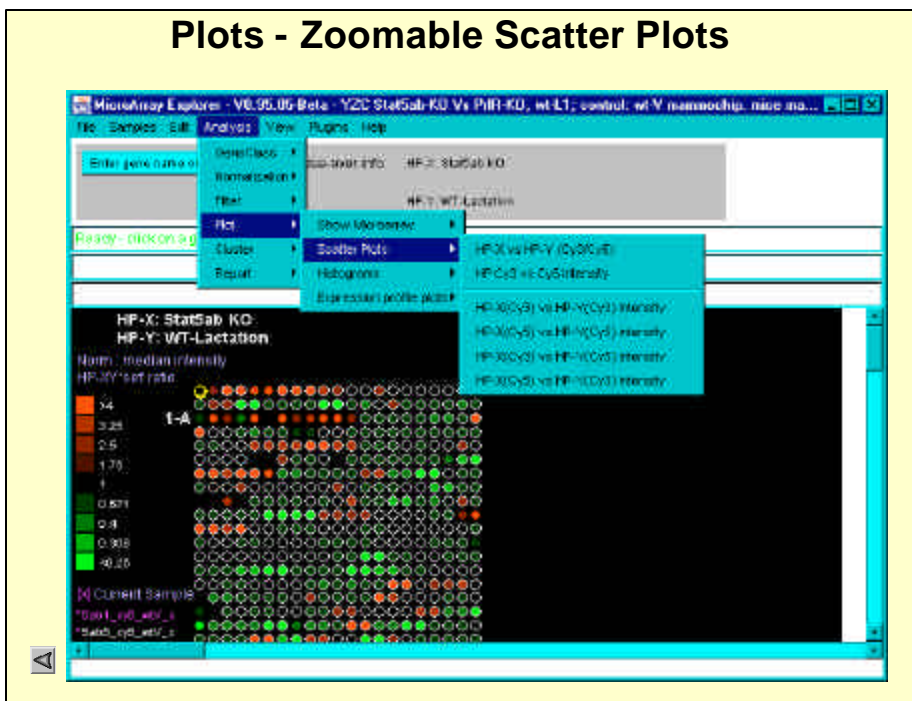
- Normalization
 - Zscore of intensity
 - Median intensity
 - Log median intensity
 - Zscore of log intensity, stdDev
 - Zscore of log intensity, mnAbsDev
 - By 'User Normalization Gene Set'
 - By housekeeping gene set
 - Scale intensity data to 65K
 - Unnormalized
 - Use background intensity correction
 - Use ratio median correction
 - Use per-sample Good Spots data for global array statistics

The background shows the same heatmap as the previous image, with the current sample '*5ab1_oy5_wtV_c' selected.

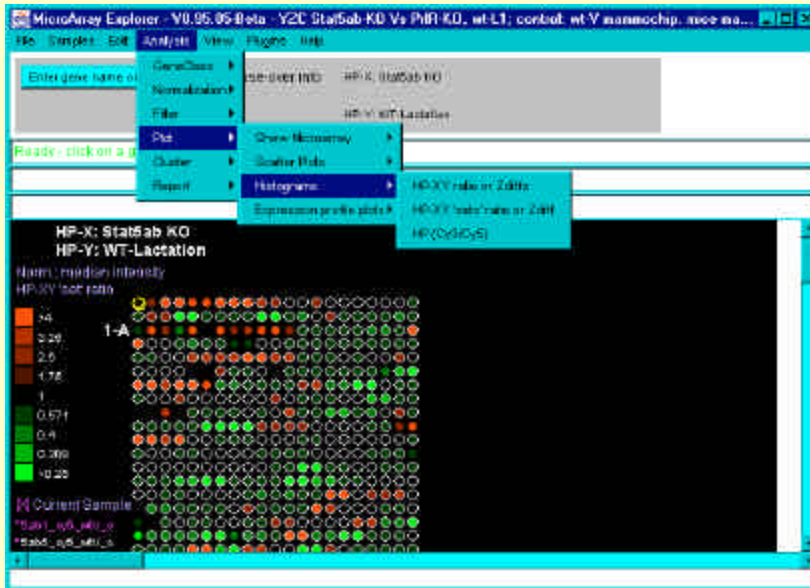
Menu: Plots - Pseudoarray Images of Functions of Samples



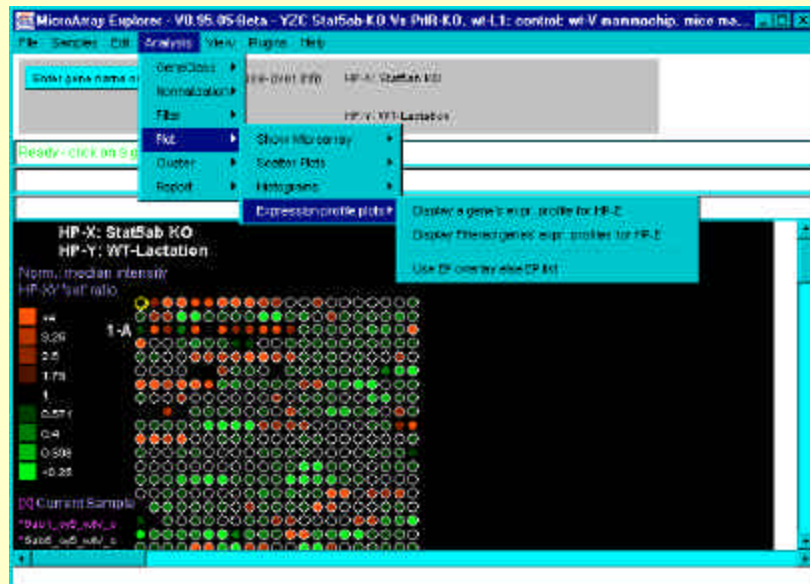
Plots - Zoomable Scatter Plots



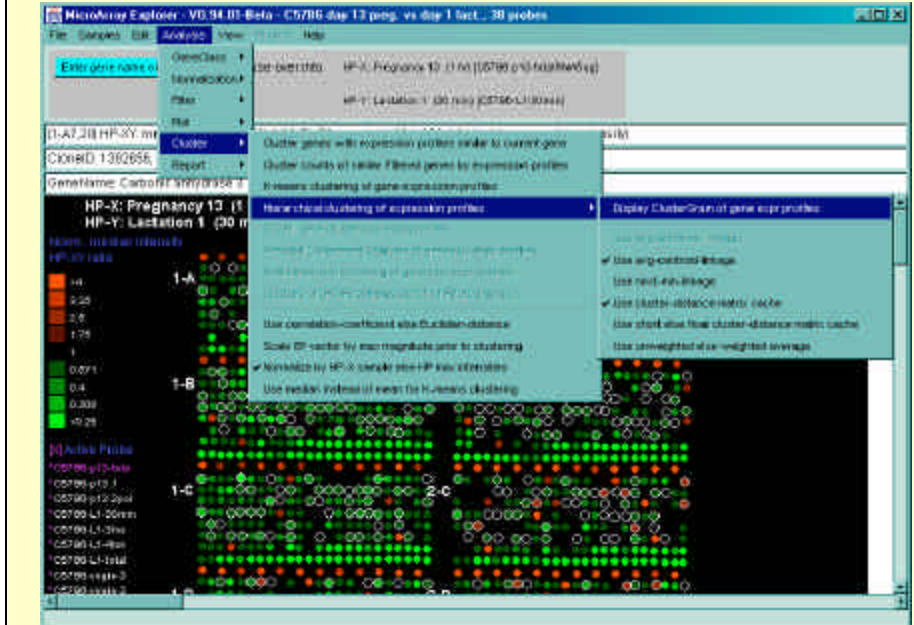
Menu: Plots - Histograms



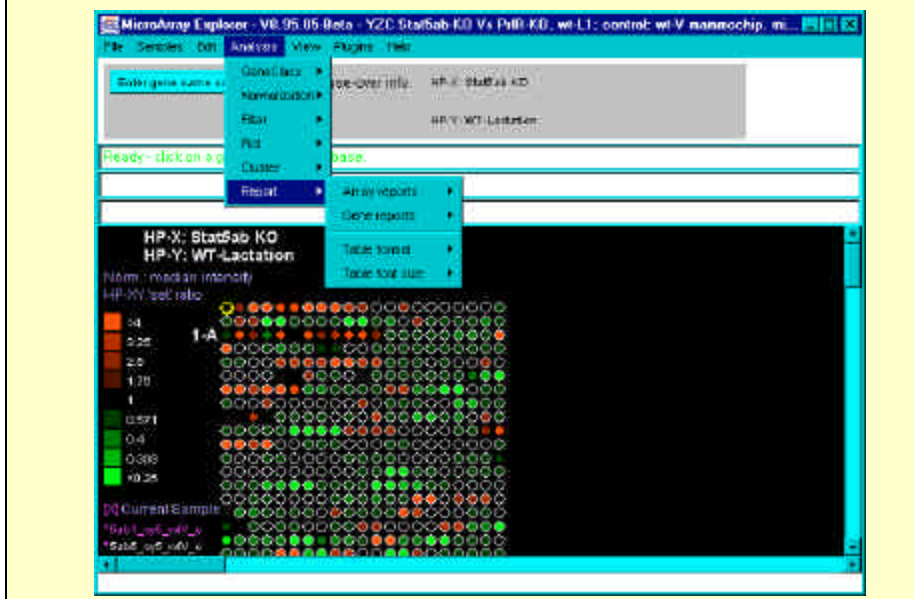
Menu: Plots - Expression Profile Plots



Clustering Similar Genes, K-means, Hierarchical



Report Operations - Samples and Gene Subsets



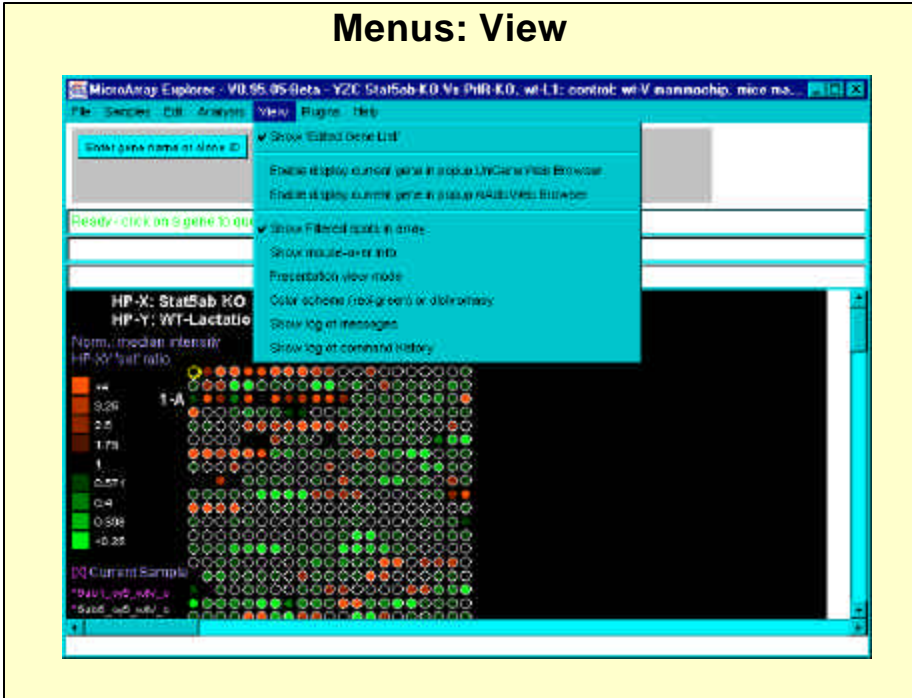
Report Menu - Samples Reports

The screenshot shows the MicroArray Explorer interface. The title bar reads "MicroArray Explorer - V0.95.05-Beta - YZC Stat5ab-KO Vs PriR-KO, wt-L1: control: wt-V mammochip. mi...". The menu bar includes File, Samples, Edit, Analysis, View, Plugins, and Help. The "Report" menu is open, showing options: Array reports, Gene reports, Table format, and Table font size. The "Gene reports" sub-menu is further expanded to show: Hybridized samples info, Hybridized samples Web links, Extra Info on Samples, MAE Projects DB, Samples vs Samples correlation coefficients, 'Calibration DNA' summary, and Samples mean & variance summary. The background displays a heatmap for "HP-X: Stat5ab KO" and "HP-Y: WT-Lactation" with a color scale from <math>< 0.25</math> to >4.

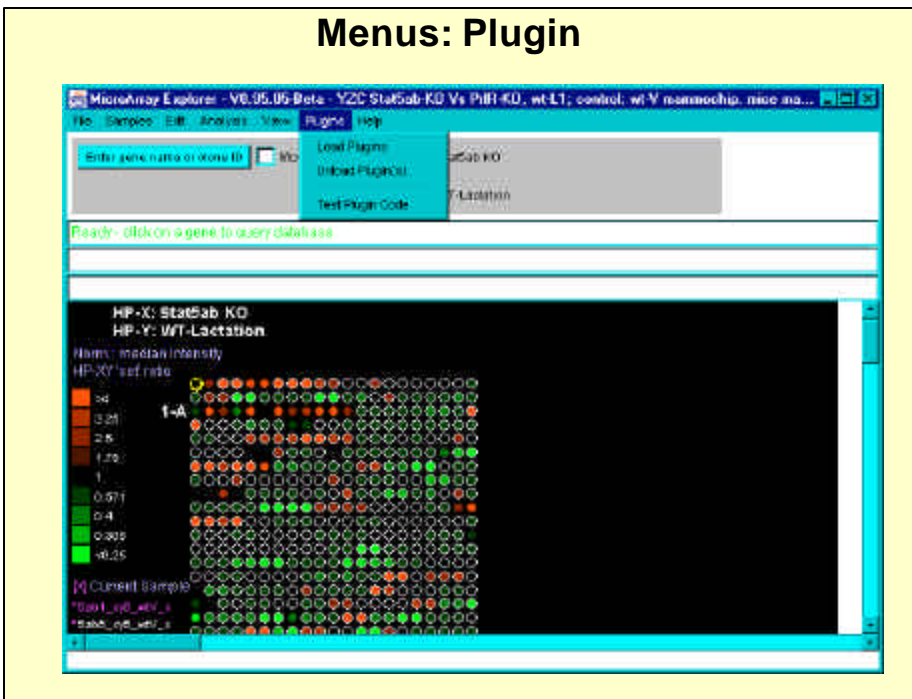
Report Menu - Gene Subsets Reports

The screenshot shows the MicroArray Explorer interface. The title bar reads "MicroArray Explorer - V0.95.05-Beta - YZC Stat5ab-KO Vs PriR-KO, wt-L1: control: wt-V mammochip. mice mammo...". The menu bar includes File, Samples, Edit, Analysis, View, Plugins, and Help. The "Report" menu is open, showing options: Array reports, Gene reports, Table format, and Table font size. The "Gene reports" sub-menu is further expanded to show: All named genes, Genes in Filtered Gene List, Genes in Normalization Data List, Genes in GeneClass, and Filtered gene reports. The "Filtered gene reports" sub-menu is further expanded to show: Done working the Filter, Highest HP:XY ratios in 20fts, Lowest HP:XY ratios in 20fts, Highest Cy3:Cy5 ratios in 20fts, Lowest Cy3:Cy5 ratios in 20fts, Expression profiles of Filtered genes, and HP:XY 'set' statistics of Filtered genes. The background displays a heatmap for "HP-X: Stat5ab KO" and "HP-Y: WT-Lactation" with a color scale from <math>< 0.25</math> to >4.

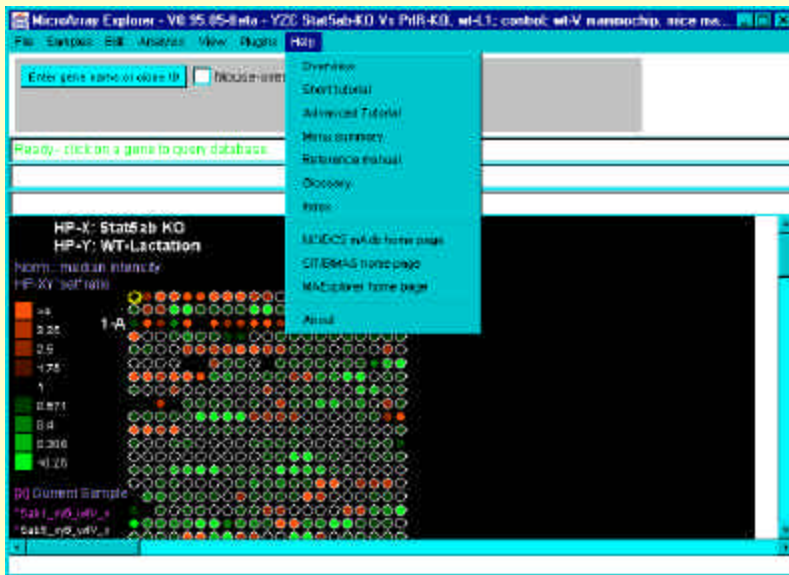
Menus: View



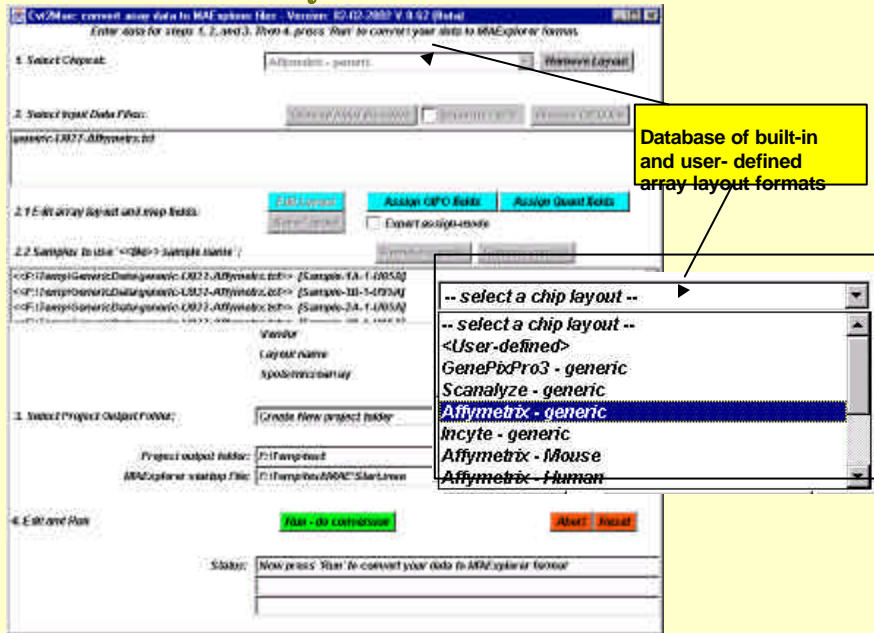
Menus: Plugin



Menus: Help



Cvt2Mae Array Data Conversion "Wizard"



Access Array Data from NCI/CIT mAdb Server <http://nciarray.nci.nih.gov/>

Procedure

1. Install MAExplorer
2. Select arrays to analyze
3. Download and unpack .zip file of this data to local disk
4. Run MAExplorer on this data

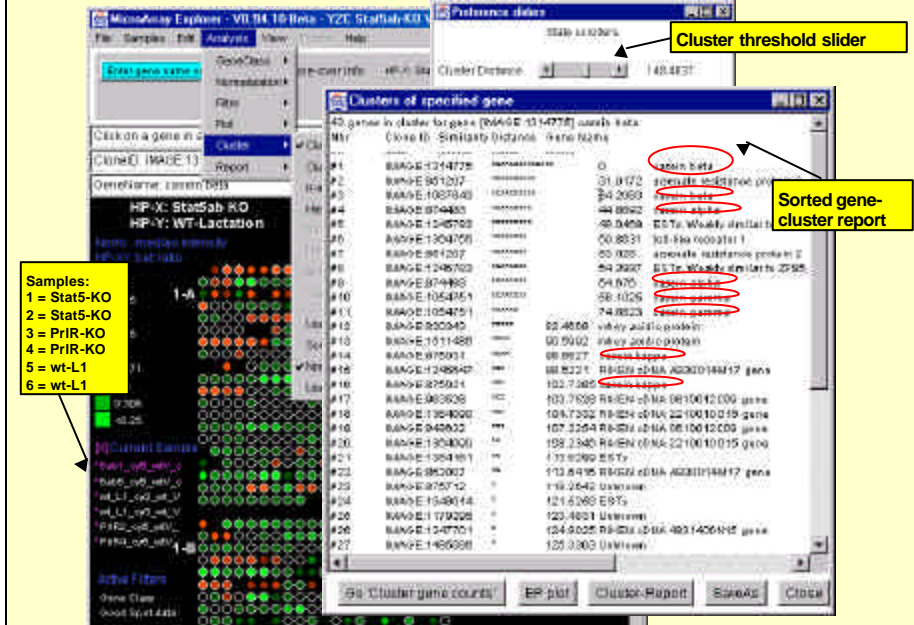
E.g. Similarity Clustering for Stat5ab-KO, PrIR-KO, wt-Lactation (wt-Virgin Ctrl) - Mammochip, Hennighausen et al.

Cluster command issued from menu

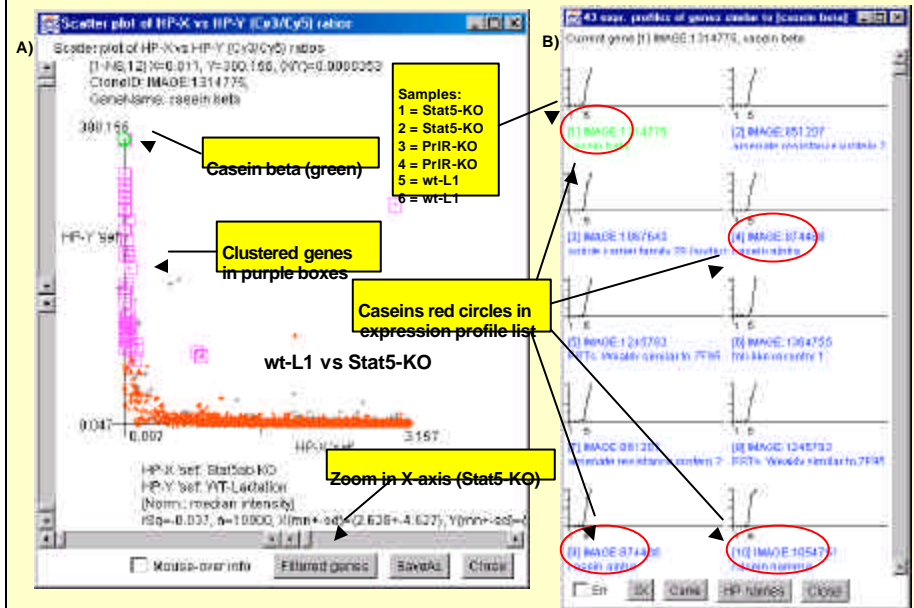
Samples:

- 1 = Stat5-KO
- 2 = Stat5-KO
- 3 = PrIR-KO
- 4 = PrIR-KO
- 5 = wt-L1
- 6 = wt-L1

Similar-Expression Cluster Report of Casein Beta



A: Scatter Plot of Genes Similar to Casein Beta (violet boxes) B: List of Expression Profiles of Most Similar Genes



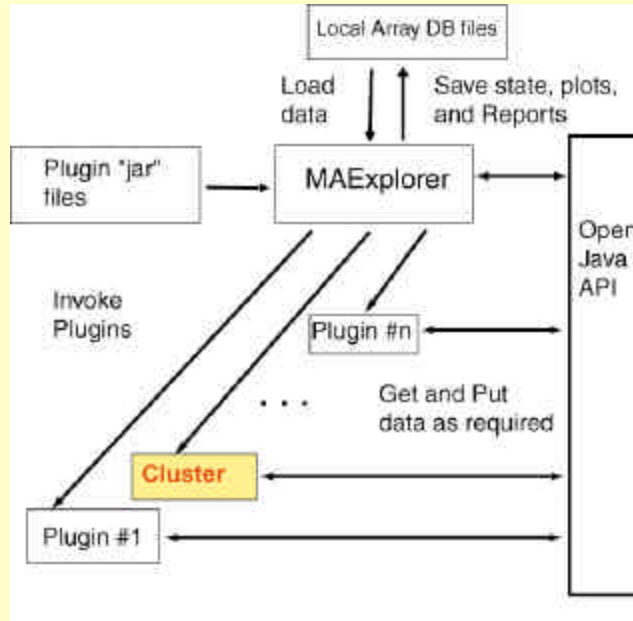
Plug-in Extensions to MAExplorer

- Java plugins allow investigators to extend capabilities of the core MAExplorer program to new analysis methods
- Web site contents: Open Java API, Java open-source examples, donated plugins & links are published and freely available. Sample plugins can serve as models for new ones
- MAEPlugin types: normalization, metrics, Filters, PCA, clustering, client-server, functional genomic analysis of cluster results, etc.
- MAEPlugins will have three types of implementations:
 1. Using 100% Java code
 2. Access local programs written in any language (e.g. 'R')
 3. Web-CGI or client-server to specialized genomic DBs

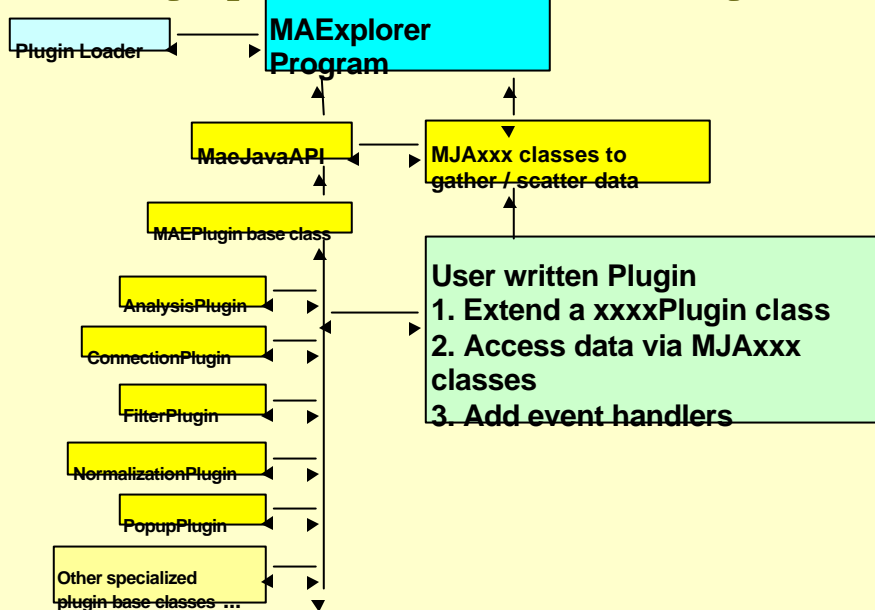
MAEPlugin schema

- Plugins are installed either when MAExplorer starts or later when it is running
- Plugins are usually invoked from the pull-down menus
- Client-server plugins may control MAExplorer by other programs
- Plugins can access any MAExplorer data structures through the MJA (MAExplorer Java Application Programming Interface or API)
- Plugins can provide their own GUI interfaces or may save data back into MAExplorer, and use its plot and report capabilities

MAEPlugin Design



Using Open Java API to Write a Plugin



MAExplorer Java API Classes Available for User MAEPlugins

MJxxxx Class	Objects and method access
MJAbase	base class and constants used by other MJA classes
MJAcluster	cluster data structures and methods
MJAcondition	condition lists of samples and ordered lists of condition lists
MJAeval	command interpreter to invoke MAExplorer commands
MJAexprProfile	expression profiles data
MJAfilter	gene data filters
MJAgene	access single gene genomic and normalized quantified data
MJAgeneList	lists of genes and get sets
MJAgenomicDB	genomic databases on the Internet
MJAgeometry	array geometry, spot to gene maps, etc.
MJAhelp	popup browser help methods
MJAhistogram	histogram plots
MJAmath	built-in math functions
MJAnormalization	normalization data and methods
MJAproperty	get and put individual properties
MJApropList	get lists of properties
MJAsample	get and put single sample low-level data
MJAsampleList	get lists of samples top-level data
MJAsort	built-in sort methods
MJAsstatistics	built-in statistics methods
MJAsstate	get and save state, get additional state info
MJAutil	built-in utility methods

E.g. Montage MAEPlugin of Current Gene for Visual Verification (Lemkin & Evans)

1. Load Montage MAEPlugin

2. Show Montage on current gene

3. Current gene

4. Montage MAEPlugin showing original image regions of current gene

E.g. ScatterPlot MAEPlugin of All Samples (Shen & Lemkin)

1. Load ScatterPlotAllPlugin

2. Show ScatterPlotAllSamples

3. ScatterPlotAllSamples MAEPlugin

E.g. ListFilter MAEPlugin

1. Load ListFilterPlugin

2. Show ListFilterPlugin

3. ListFilterPlugin MAEPlugin

Example: “ListFilter” MAEPlugin Java Code⁺⁺

- The `ListFilterPlugin.java` class is the one specified to the MAExplorer Java Plugin loader. It
 - installs the menu entry name in MAExplorer
 - it invokes a new instance of `ListFilter` when selected from the menu
- The `ListFilter.java` class is called by `ListFilterPlugin.java` when invoked from the menu. It
 - creates a popup GUI extending *Frame*
 - using the MJA classes, it gathers state information on the current MAExplorer data active data filters
 - draws this information in a *TextArea* in the *Frame*

⁺⁺ Source code is available on <http://maexplorer.sourceforge.net/>

Summary

- MAExplorer is a flexible fully Open-Source microarray data-mining tool
- Identified genes preferentially expressed during lactation, Lemkin et al. *Nucleic Acids Res.* (2000) **28**:4452
- Uses direct-manipulation, data filtering, built-in graphics, statistics, clustering, gene and sample set operations, reports
- Manages multiple samples, replicates, gene sets, expression profile lists - exploration state may be saved and restored
- Accesses genomic Web databases for further analysis
- Convert user data with Cvt2Mae “wizard” tool for commercial & academic chips
- Users may add new analytic methods using MAEPlugins extensions