

Using MicroArray Explorer with the NCI/CIT/ATC mAdb Web Server, Academic and Commercial Chips

P.F. Lemkin⁽¹⁾, G. Thornwall⁽²⁾, B. Stephens⁽²⁾, K. Walton⁽³⁾, Y. Cui⁽³⁾,
J-P. Renou⁽³⁾, M. Dimitrov⁽¹⁾, L. Anderson⁽⁴⁾, L. Hennighausen⁽³⁾

(1)LECB, NCI/FCRDC; (2)SAIC, NCI/FCRDC ; (3)LGP, NIDDK;
(4)LCC, NCI/FCRDC

MAExplorer: <http://www.lecb.ncifcrf.gov/MAExplorer>

Abstract

- The Microarray Explorer (**MAExplorer**) is a versatile Java-based data mining bioinformatic tool for analyzing quantitative cDNA expression profiles across multiple microarray platforms and other DNA labeling systems.
- These include common NIH chips such as NCI/ATC, Incyte and Affymetrix. Data is downloaded as required from the ATC server or other source and reside on the user's computer. Incyte and Affymetrix data is converted using the Cvt2Mae conversion program. With MAExplorer, 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 genes of interest.
- Analyses are performed in real-time and may be viewed and directly manipulated in array pseudo-images, scatter plots, reports, histograms, expression profile plots, and cluster analyses plots. Raw 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 or exported. Similar sets of 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, 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.
- MAExplorer, a sample set of 38 MGAP hybridizations, a reference manual, tutorials and documentation may be freely downloaded from **<http://www.lecb.ncifcrf.gov/MAExplorer>**.

MAExplorer is Compatible with mAdb Data

mAdb Downloads - Netscape

File Edit View Go Communicator Help






Back Forward Reload Home Search Netscape Print Security Shop Stop

Division of Clinical Sciences NCI

CIT Center for Information Technology

NCI/DCS μ Array Center mAdb Downloads

Page Updated: Thursday, 24-May-2001 14:01:17 EDT

		Program	Description	Author	Version	Updated	Download	Manual
<p>Axon Inc. Software</p> <p>This is commercial, licensed software and the GenePix application requires a "dongle" attached to the parallel port to run. The manual is accessible to all.</p> <p>Axon's Web Site</p>		GenePix Pro 3	Fully integrated acquisition and analysis software for the GenePix 4000A. Download to a folder of your choice and then run to extract the installation files. Then run the extracted file setup.exe and follow installation instructions		3.0.6.66 History	(Posted here 05/24/2001)	Download	GenePix3.0 Manual Axon Scanner Manual (PDFs)
<p>Stanford Genome Analysis Group Software</p> <p>It is available free of charge to academic and non-profit institutions.</p> <p>Stanford Download Site</p>		ScanAlyze	Image Analysis (extracts data from fluorescence images of arrays)	Michael Eisen	2.44	11/15/99	Download	Manual (PDF)
		Cluster	Perform Hierarchical Clustering, Self-organizing Maps, k-Means Clustering, and More	Michael Eisen	2.11.01	7/10/2000 (Posted here 10/26/2000)	Download	Manual (PDF)
		TreeView	Graphical Viewing and Browsing of Cluster Results	Michael Eisen	1.45	9/28/99	Download	
<p>MAExplorer</p> <p>Developed by and Available from LECB/FCRF/NCI.</p> <p>MAExplorer Web Site</p>		MAExplorer	A Java data mining application for gene expression data using a variety of statistical, clustering, direct-manipulation graphical, spreadsheet and Web access methods.	Peter Lemkin	Revision History	Current version	Link to Download	Manual (Online) Use with mAdb data (PDF)

Document: Done

MAExplorer Home Page

<http://www.lecb.ncifcrf.gov/MAExplorer>

The screenshot shows a Netscape browser window titled "MicroArray Explorer - MAExplorer - Netscape". The address bar shows the URL "http://www.lecb.ncifcrf.gov/MAExplorer/". The page content is as follows:

MAExplorer

Introduction
Overview
Demonstrations

Documentation

- [Manual \(on right\)](#)
- [Manual \(new window\)](#)
- [Manual \(7 Mb Zip\)](#)
- [Manual \(entire\)](#)
- [Newsletters](#)

• [Short Tutorial](#)

• [Advanced tutorial](#)


• [Menu summary](#)

• [Quick Start](#)

• [Glossary](#)


• [Index](#)

• [Help Desk](#)

•  [Documents \(PDF\)](#)


Data format conversion

- [User's array data](#)

•  [O2M Cvt2Mae converter](#)

Downloading MAExplorer

- [Stand-alone version](#)
- **NEW** [Revision notes](#)

•  [Install MAExplorer](#)

• [download](#)

[Disclaimer](#)

[MGAP-MAExplorer home](#)

Contact ([Leatkin](#) | leatkin@ncifcrf.gov | [LECB, NCI/FCRF](#))

Revised: \$Date: 2001/08/30 20:55:09 \$

MAE MAExplorer - MicroArray Explorer

(Stable release: 0.90.08, Beta release:)

MicroArray Explorer for Data Mining Gene Expression Patterns

The Microarray Explorer (MAExplorer) is a Java-based data-mining facility for cDNA microarray databases. It may be freely [downloaded](#) and run as a [stand-alone application](#) on your computer, or run as an applet in your Web browser. The exploratory data analysis environment provides tools for the [data-mining](#) of quantitative cDNA expression profiles across multiple microarrays.

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 and outliers; 4) directly access other genomic databases for clones of interest. In the applet version, data is downloaded as required from the server to the user's Web browser where real-time analyses are performed. The stand-alone version uses previously quantified array data copied to the local computer where it may save data from data mining sessions.

Microarray data may be viewed and directly manipulated in array pseudomages, scatter plots, histograms, expression profile plots, cluster analyses (similar clones, K-means, hierarchical clusters, etc.), and reports. A key feature is the clone data filters for constraining a working set of clones to those passing a variety of user-specified tests. Reports may be generated with hypertext Web access to genomic databases such as UniGene, GenBank, dbEST, I.M.A.G.E., NCI/CIT mAdb Clone DB and other Internet databases for sets of clones found to be of interest.

A major focus of this tool is interactive data mining with access to other supporting Web genomic databases. The emphasis on direct manipulation of clones and sets of clones 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.

It was developed by the NCI Laboratory of Experimental and Computational Biology ([LECB](#)) in collaboration with the NIDDK Laboratory of Genetics and Physiology ([LGP](#)). MAExplorer was created to help analyze microarray data for the LGP's Mammary Genome Anatomy Program (MGAP) designed to identify and understand genetic pathways operative during normal mammary gland development and tumorigenesis. Note that 38 hybridizations from the MGAP database are included as a demonstration database when you download the stand-alone version of MAExplorer.

MAExplorer may be used as an applet to access the Mammary Genome Anatomy Project (MGAP) microarray Web data through <http://mammary.nih.gov/mgap> or directly at <http://www.lecb.ncifcrf.gov/mae>.

The [Reference Manual](#) describes the details of MAExplorer as well as showing a number of screens illustrating various operations. Several [tutorials](#) are available as well 38 hybridizations from the MGAP database.

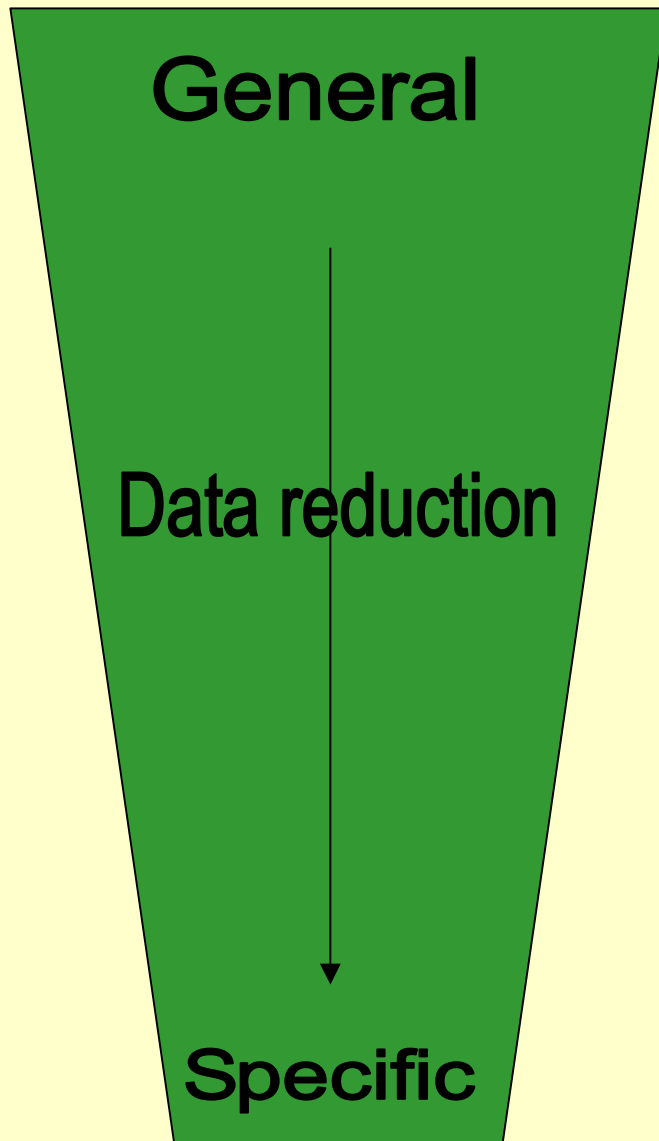
MAExplorer was Developed for the Ongoing Mammary Genome Anatomy Program (MGAP)

- MGAP models, histology, expression
<http://mammary.nih.gov/>
- Mouse models of
 1. normal mammary development C57B6
 2. Knock-outs: Stat5a (-,-) & (+,-), CEBP null, TGF α etc.
 3. Tumor models: WAP taq, p53 mutant, etc.
 4. Transgenics: WAP Int3, BRCA1, etc.
- Initially, MAExplorer was a Java applet for analyzing MGAP arrays <http://www.lecb.ncifcrf.gov/mae>
- MAExplorer identified genes in preferentially expressed during lactation, *Nucleic Acids Res.* (2000) **28**:4452

I. The Problem and Goals of Array Analysis

- Assume we have microarray data in a spreadsheet of quantified spots and the genes they represent - **What do we do with all those spots?**
- Could look for **patterns of changes** of experimental conditions with quantitative gene expression for well formed questions
- **Correlation of gene expression changes** with biological state implies a relationship but does not imply cause and effect
- Having found correlations, review findings with annotated bioinformatic data sources (PubMed) and validate with additional lab experiments
- Handles Cy3/Cy5 ratio (ATC, Incyte, etc), ^{33}P intensity (Affymetric, etc.), and academic chips using **Cvt2Mae** converter program

Data Mining - Finding Putative Relevant Patterns



Quantified array spot data for multiple samples

Organize by sample, gene expr, gene sets

Change views: normalization, data filters

Visualize and query: plots, cluster, reports

Explore external genomic databases

Results

I.1 What is the MicroArray Explorer?

- **MAExplorer is a flexible Java stand-alone (off-line) or applet (Web-based) microarray data-mining tool**
- Handles multiple arrays with replicates and expression profiles
- Handles intensity or ratio (Cy3/Cy5) quantified microarray data
- Data-filters gene sets by statistics, clustering, gene set operations
- Allows direct manipulation of data in graphics, spreadsheets
- Accesses genomic Web servers from plots and reports
- Freely downloadable from MAExplorer Web site
- Stand-alone installers are provided for Windows, MacOS, Solaris, Linux, Unix, etc.

I.2 Types of Experiments

- What **types of expression** could we analyze?
- Discover **patterns of differential expression** using visual, statistical, and clustering methods:
 - 1) of individual genes,
 - 2) of gene families and clusters of genes,
 - 3) as a function of conditions: development, time (eg. cell cycle), cell lines, disease progression, pathways models, etc.
- Finding genes with **similar gene expression** may help in understanding a gene's functional behavior or pathways
- These are **statistical entities**. The more data samples and replicates are available, the better these estimates will be

I.3 Operations on 2- & N-conditions of Samples⁽¹⁾

Set of HP-X replicate samples

Set of HP-Y replicate samples

Operations on sets of replicate samples- e.g. *t*-test, CV statistical tests

Order list of HP-E samples

Operations on order lists of samples: e.g. clustering, Expr. Profile plots

⁽¹⁾ “HP” is hybridization probe i.e. the sample hybridized to the chip

I.4a Choosing HP-X, HP-Y Sets and HP-E Lists

MGAP DB - MicroArray Explorer

File HybProbe Edit Analysis

Enter gene name or clone ID

HP-X: Pregnancy 13
HP-Y: Lactation 1 (3)

Norm.: median intensity
 HP-XY ratio

>4.0
 3.25
 2.5
 1.75
 1.0
 0.571
 0.4
 0.307
 <0.25

1-A
 1-B
 1-C
 1-D

[X] Active Probe
 * C57B6-virgin-3
 * C57B6-virgin-3
 * C57B6-p13-tota
 * C57B6-p13.1
 * C57B6-p13.2pol
 * C57-p16-2hrs-5
 * C57B6-L1-30min
 * C57B6-L1-3hrs
 * C57B6-L1-4hrs
 * C57B6-L1-total
 * C57B6-L3-1hr

Choose HP-X and HP-Y sets, and HP-E list of hybridization probes

Remaining hybridization sample probes

HP-X set selected

Use above probes

HP-Y set selected

Remaining hybridization sample probes

HP-E ordered list selected

HP: C57B6 virgin 3 hours #1

Project: C57Development

Title: Virgin 10 weeks old (3 hrs) [C57B6-virgin-3hrs-1]

OK Cancel Reset

I.4b Selectively Swapping Cy3,Cy5 for Particular Samples in Dye-Swap Experiments

The screenshot shows the MicroArray Explorer interface. At the top, the title bar reads "mAdb Hs-OC-2-23Cx24R DB - MicroArray Explorer - V0.89.35-Beta - Working title". The menu bar includes "File", "HybProbe", "Edit", "Analysis", "View", and "Help". Below the menu bar is a search bar with the text "Enter gene name or clone ID" and a checked "Mouse-over info" option. The search results show "HP-X: [MmOC8p6_46_16084] (Cy3/Cy5)" and "HP-Y: [MmOC8p6_52_16090] (Cy3/Cy5)".

A message below the search bar states: "Will swap (Cy3,Cy5) for [MmOC8p6_52_16090] to Cy5/Cy3".

The main area displays a heatmap with a legend on the left. The legend is titled "Norm.: median intensity" and "HP-XY ratio". It includes a color scale from red (>4.0) to green (<0.25). A "1-A" label is positioned above the heatmap. Below the heatmap, a list of probes is shown, with the first few highlighted in pink: "* MmOC8p6_46_160", "* MmOC8p6_47_160", "* MmOC8p6_48_160", "* MmOC8p6_49_160", "* MmOC8p6_49_160", "* MmOC8p6_50_160", "* MmOC8p6_51_160", "* MmOC8p6_52_160", "* MmOC8p6_53_160", "* MmOC8p6_54_160", "* MmOC8p6_55_160", and "* MmOC8p6_56_160".

A dialog box titled "Select Hybrid Probes to swap [Cy3/Cy5] to ..." is open in the foreground. It contains a list of probes with checkboxes: [MmOC8p6_46_16084], [MmOC8p6_47_16085], [MmOC8p6_48_16086], [MmOC8p6_49_16087] (checked), [MmOC8p6_50_16088], [MmOC8p6_51_16089], [MmOC8p6_52_16090] (checked), [MmOC8p6_53_16091], [MmOC8p6_54_16092], [MmOC8p6_55_16093], [MmOC8p6_56_16094], and [MmOC8p6_57_16095]. The dialog box has "Done" and "Cancel" buttons at the bottom.

I.5 Definition: Gene Expression Profile

- An **expression profile** e_j of an ordered list of N normalized spot values samples v_{jk} ($k=1$ to N) for a particular gene j
- The expression profile e_j for a particular gene j is:
$$e_j = (v_{j1}, v_{j2}, v_{j3}, \dots, v_{jN})$$
- A **difference** between two genes p and q may be estimated as a N -dimensional metric “**distance**” between e_p and e_q
- **Euclidean distance:** $d_{pq} = 1/N \left(\sum_{j=1:N} (v_{pj} - v_{qj})^2 \right)^{1/2}$
- Other distance measures: correlation coefficient, city-block, etc.
- If distance is scaled to $[0:1]$, then **Similarity** measure: $s_{pq} = 1 - d_{pq}$

I.6a Expression Profile Plots - Examples

MGAP DB - MicroArray Explorer - V0.89.12-Beta - Pregnancy day 13: C57BL/6 vs. stat5a (-.-) ...

File HybProbe Edit Analysis View Help

Enter gene name or clone ID MOUSE-OVER info

HP-X: Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug]

HP-Y: Pregnancy 13 (15 min) [Stat5a...p13-15min]

[1-A3,7] intensity[F1]=31.9504, intensity[F2]=26.4701, F1/F2=1.2070, (Norm.: median intensity)

CloneID: 1247513, dbEST3: 1696290, GeneBankAcc3: AA959964, UniGene: Mm.6800, plate[3,A,7]

GeneName: Keratin complex 2, basic, gene 8

Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug]

Norm.: median intensity
Intensity

>137.357 1-A 2-A

120.199

Gene Expression Profile #1

CloneID [1247513]
Keratin complex 2, basic, gene 8

Err 1X Line Show HPs SaveAs Close

Gene Expression Profile #2

#4 [C57B6-p13.1] 1:2.475 CV:0.049

CloneID [1247513]
Keratin complex 2, basic, gene 8

Err 1X Line Show HPs SaveAs Close

Gene Expression Profile #3

CloneID [1247513]
Keratin complex 2, basic, gene 8

Err 1X Circle Show HPs SaveAs Close

Gene Expression Profile #4

CloneID [1247513]
Keratin complex 2, basic, gene 8

Err 1X Curve Show HPs SaveAs Close

Current HP-E expression profile assi...

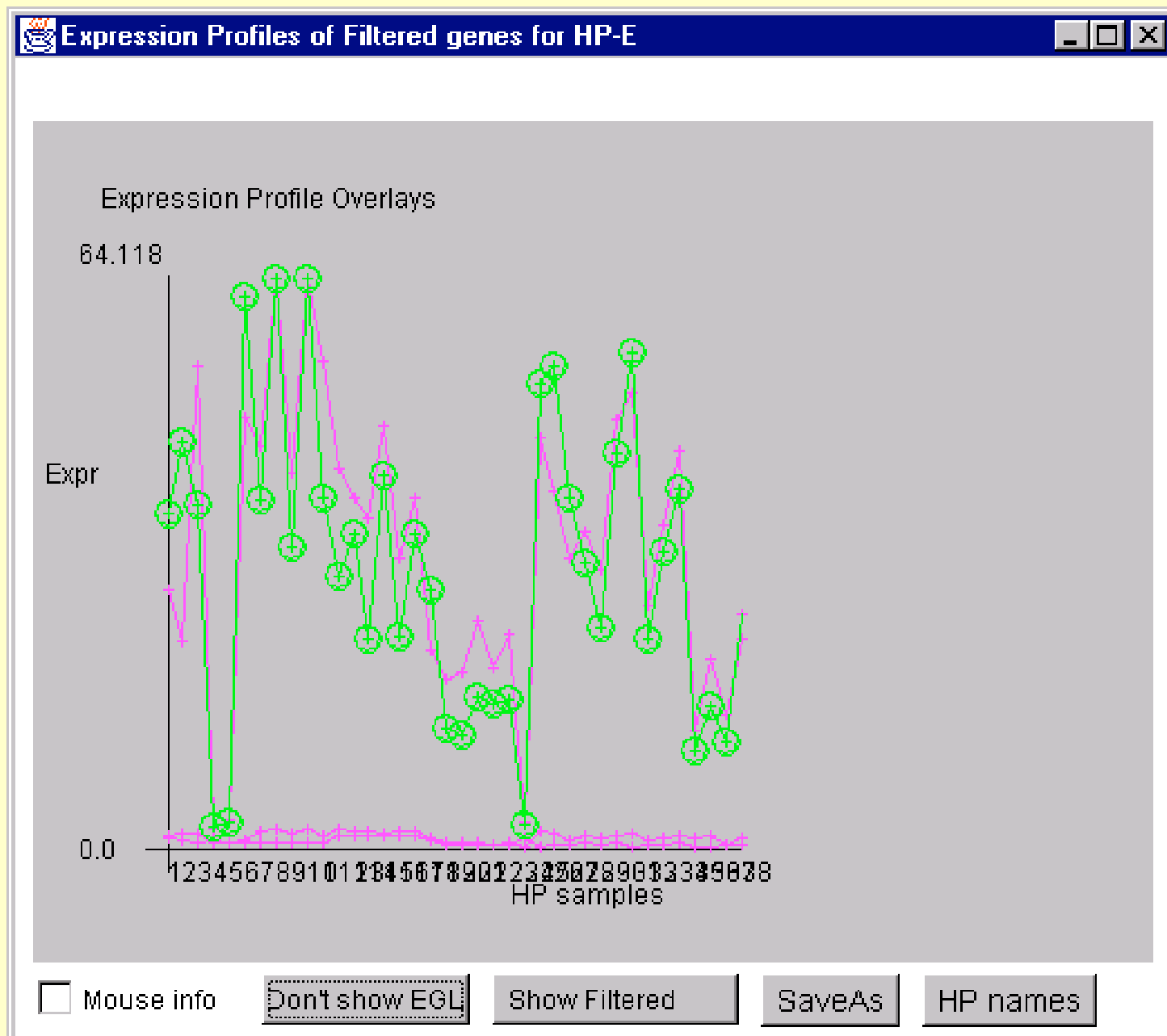
Current hybridization probe expression profile HP-E 'set a'

- #1 Virgin 10 weeks old (3 hrs) [C57B6-virgin-3hrs-1]
- #2 Virgin 10 weeks old (3 hrs) [C57B6-virgin-3hrs-2]
- #3 Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug]
- #4 Pregnancy 13 (1 hr) [C57B6-p13.1]
- #5 Pregnancy 13 (1 hr) [C57B6-p13.2poly-A]
- #6 Pregnancy 18 (1hr) [C57-p18-2hrs-50]
- #7 Lactation 1 (30 min) [C57B6-L1-30min]
- #8 Lactation 1 (3 hrs) [C57B6-L1-3hrs]
- #9 Lactation 1 (4 hrs) [C57B6-L1-4hrs]
- #10 Lactation 1 (30 min) [C57B6-L1-total]
- #11 Lactation 3 (1 hr) [C57B6-L3-1hr]
- #12 Lactation 10 (20 hrs) [C57B6-L10-20hrs-1]
- #13 Lactation 10 (20 hrs) [C57B6-L10-20hrs-2]
- #14 Lactation 10 (20 hrs) [C57B6-L10-20hrs-2nd-1]
- #15 Lactation 10 (20 hrs) [C57B6-L10-20hrs-2nd-2]
- #16 Lactation 10 (6 hrs) [C57B6-L10-6hrs-1]
- #17 Lactation 10 (hrs) [C57B6-L10-6hrs-2]
- #18 Involution-4 (25 hrs) [C57B6-I4-25hrs]
- #19 Pregnancy 13 (15 min) [Stat5a...p13-15min]
- #20 Pregnancy 13 (15 min) [Stat5a...p13-15min2]
- #21 Pregnancy 13 (1 hr) [Stat5a...p13-1hr2]
- #22 Pregnancy 13 (30 min) [Stat5a...p13-30min]
- #23 Pregnancy 13 (30 min) [Stat5a...p13-30min2]
- #24 Lactation 1 (30 min) [Stat5a...L1-30min]
- #25 Virgin (18 hrs) [CEBP-null-virgin-18hrs]
- #26 Pregnancy 13 (18 hrs) [CEBP-null-p13-18hrs]
- #27 Pregnancy 18 (18 hrs) [CEBP-null-p18-18hrs]
- #28 Lactation 1 (1 hr) [CEBP-null-1hr]
- #29 Pregnancy 18 (1 hr) [4A-1hr]
- #30 Pregnancy 18 (5 hrs) [4B-1hr]
- #31 Pregnancy 18 (1 hr) [4C-5hrs]
- #32 * (1 hr) [BA-1hr]
- #33 Virgin (1 hr) [BetaB-Hemo]
- #34 Virgin (1 hr) [BetaB-null-1hr]
- #35 * (1 hr) [HR1-1hr]
- #36 * (5 hrs) [HR2-5hrs]
- #37 Lactation 1 (1 hr) [Inh-31-null-1hr]
- #38 * (1 hr) [WAP-Tag-tumor-1hr]

#5 SaveAs Close

I.6b Expression Profile Overlay Plots

- Overlay EP plots for onco and proto-oncogenes in EGL for MGAP database



I.7 Why Do We Need to Cluster the Data?

- Clustering methods are one way to perform **gene discovery** by finding similar expression in a set of genes over a set of experiment samples
- A gene cluster is a set of genes that represent **similar gene expression** across that set of experiment samples
- Some clusters may contain genes be related by pathway or biological function
- **Many ways to cluster** the data (such as):
 - C.1 Find genes with similar expression to known gene
 - C.2 K-means clusters where the number of clusters K is fixed
 - C.3 Hierarchical clustering where a binary hierarchy is created

I.8 Find Similar Genes to Seed Gene

- Find a **sorted list** of all genes $\{g_j\}$ similar to gene g_s
- Determine if g_j is similar to seed gene g_s if distance $d_{js} < \text{threshold } T$

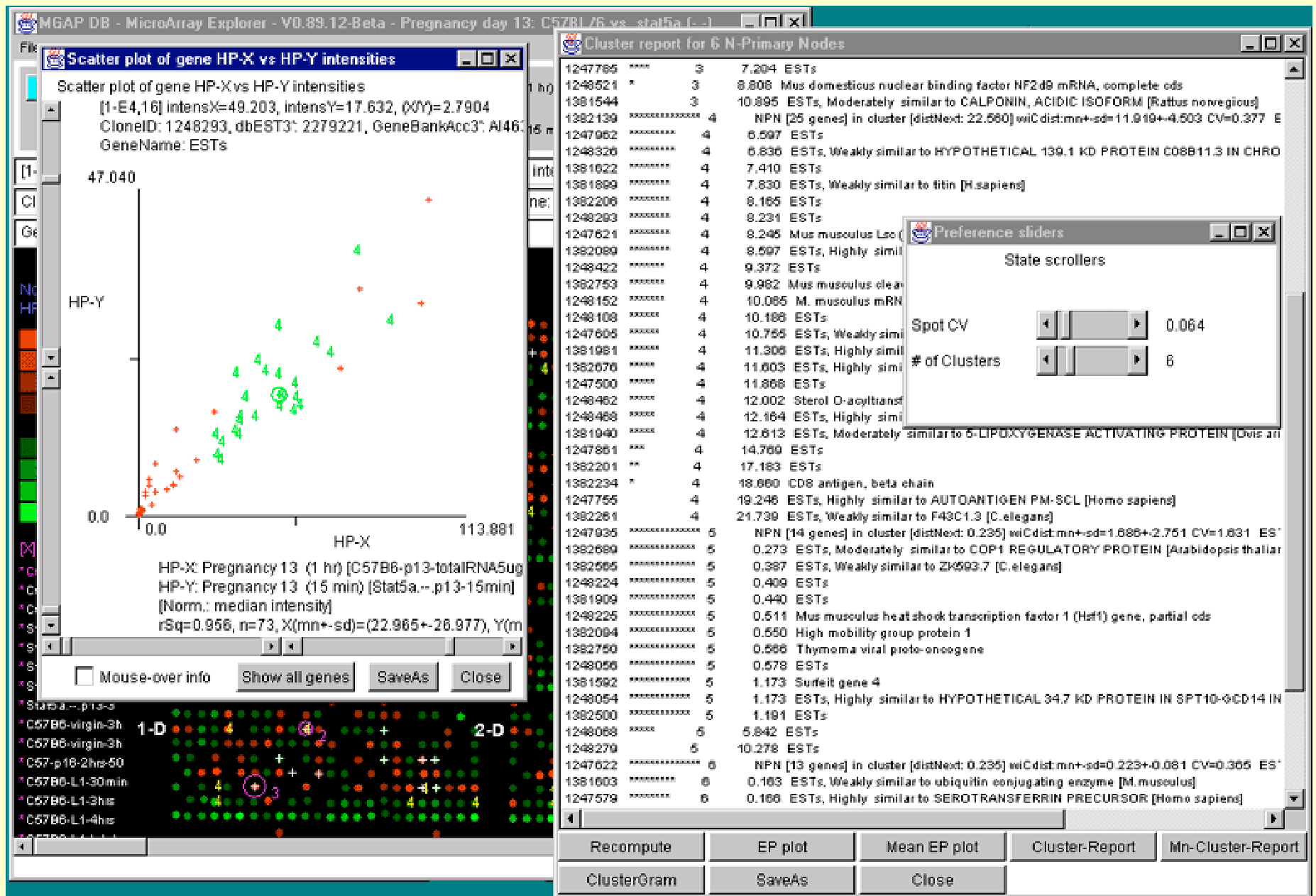
The screenshot displays the MGAP DB - MicroArray Explorer interface. The main window shows a heatmap of gene expression data for 'Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug]'. The heatmap is organized into three rows (1-A, 1-B, 1-C) and two columns (2-1, 2-2). A legend on the left indicates normalized median intensity levels, ranging from <math><0.092</math> to >137.357. Active probes are listed on the left, including C57B6-virgin-3h and C57B6-p13-total. A 'Preference sliders' window is open at the bottom, showing a 'Cluster Distance' slider set to 10.346.

A secondary window titled 'Clusters of specified gene' is open on the right, displaying a list of 43 genes in a cluster for the seed gene [1381538] Raf-related oncogene. The table below summarizes the data shown in this window:

Nbr	ClonID	Similarity	Distance	Gene_Name
#1	1381538	*****	0.0	Raf-related oncogene
#2	1248384	*****	3.3601	Mouse mRNA for 65-kDa macrophage cyt
#3	1248432	*****	3.0787	Prosaposin
#4	1382060	****	5.5852	ESTs
#5	1248050	****	5.6104	Mus musculus C57BL/6J ribosomal protein S28 mRt
#6	1248601	****	5.8478	EST
#7	1247522	****	5.9186	Mus musculus calpain small subunit mRNA, comple
#8	1248181	****	6.1459	ESTs, Weakly similar to endophilin II [M.musculus]
#9	1248071	****	6.2855	Histocompatibility 2, class II, locus Ma
#10	1248205	****	6.4081	Zinc finger protein 147
#11	1248612	****	6.4636	ESTs, Moderately similar to PIM-1 PROTO-ONCOG
#12	1248061	****	6.4987	Actin, alpha 1, skeletal muscle
#13	1382182	****	6.5649	ESTs
#14	1247698	****	6.5988	Adipocyte protein aP2
#15	1247603	****	7.1112	ESTs
#16	1247927	***	7.3611	ESTs, Highly similar to ATP SYNTHASE LIPID-BIN
#17	1247760	***	7.4661	Protein tyrosine phosphatase, non-receptor type sub
#18	1247612	***	7.5063	ESTs, Weakly similar to GLUTATHIONE S-TRANSFI
#19	1248312	**	7.6253	Protein kinase, mitogen activated kinase 3
#20	1382228	**	7.7015	ESTs, Moderately similar to 60S RIBOSOMAL PRC
#21	1382320	**	7.7464	ESTs, Weakly similar to unknown [S.cerevisiae]
#22	1247097	**	7.8393	ESTs
#23	1382089	**	7.9175	ESTs, Highly similar to EUKARYOTIC INITIATION I
#24	1247513	**	7.9329	Keratin complex 2, basic, gene 8
#25	1382133	**	8.1094	ESTs
#26	1248293	**	8.1281	ESTs
#27	1247621	**	8.1290	Mus musculus Lsc (Isc) oncogene mRNA, complete c
#28	1247553	**	8.2106	Mus musculus bodenin gene
#29	1248599	**	8.2694	ESTs
#30	1248418	*	8.3709	ESTs
#31	1381875	*	8.3764	ESTs
#32	1248239	*	9.1466	ESTs
#33	1381822	*	9.2046	ESTs
#34	1248469	*	9.2580	Properdin factor, complement
#35	1248025	*	9.3017	ESTs
#36	1248108	*	9.5685	ESTs
#37	1382597	*	9.7768	ESTs
#38	1247609	*	9.9234	ESTs
#39	1247778	*	9.9948	Membrane protein, palmitoylated (55 kDa)
#40	1382220	*	10.0418	ESTs
#41	1382676	*	10.1512	ESTs, Highly similar to MATRIN 3 (Rattus norvegicus)

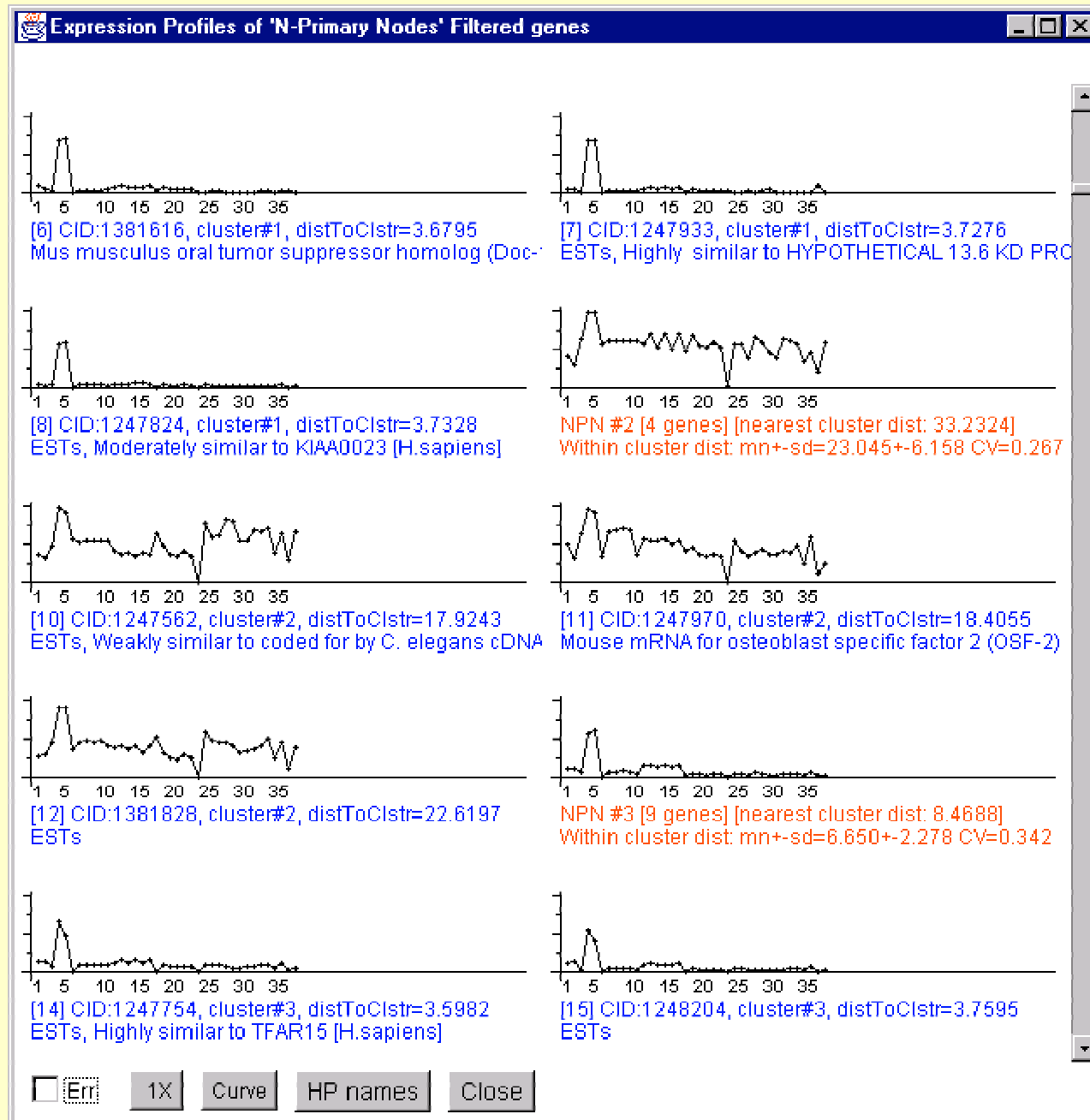
Buttons at the bottom of the cluster window include: Go 'Cluster gene counts', EP plot, Cluster-Report, SaveAs, and Close.

I.9 K-means Clusters K Sets of Similar Genes

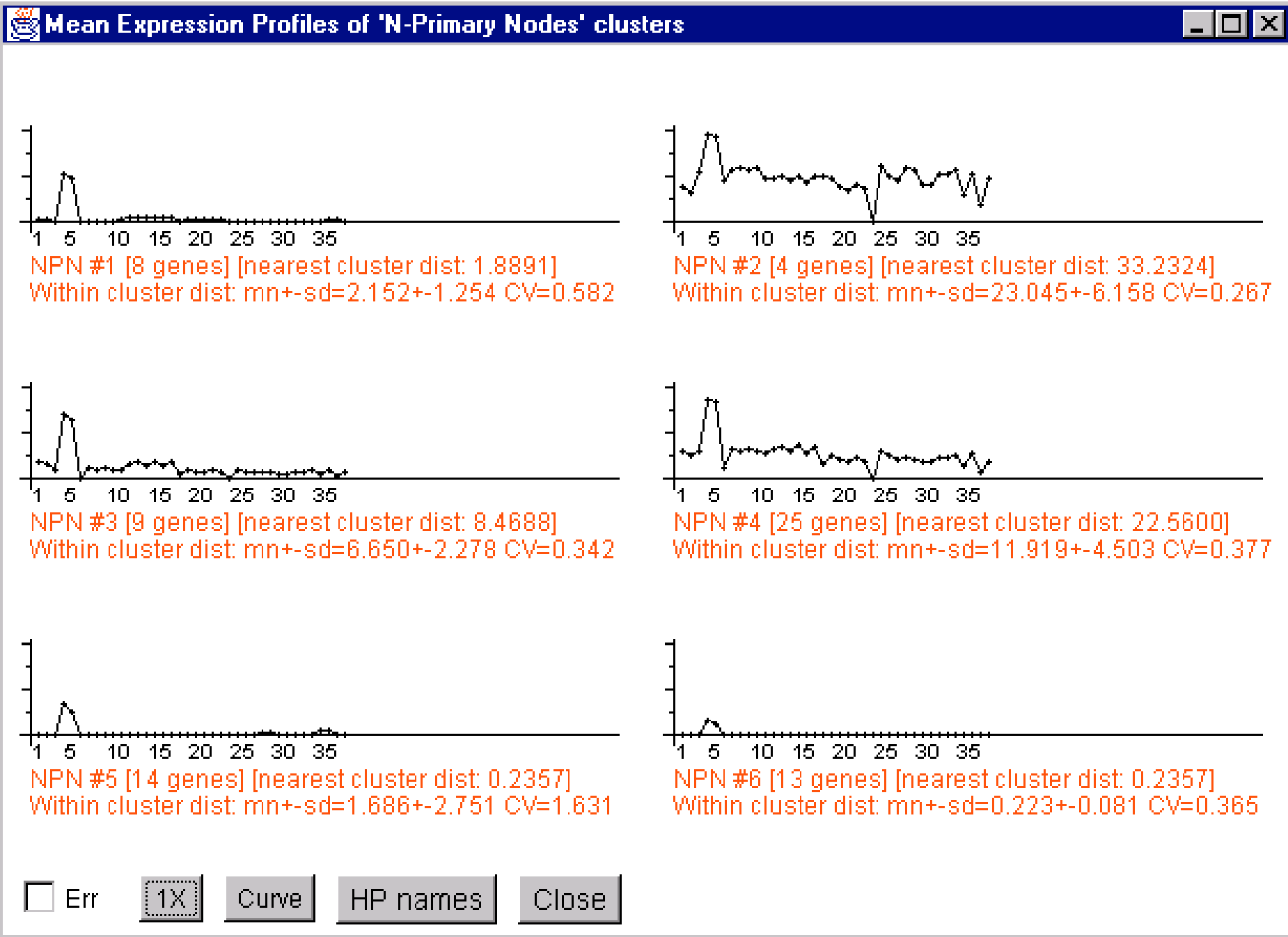


I.10a List of Expression Profiles Plots of Clusters

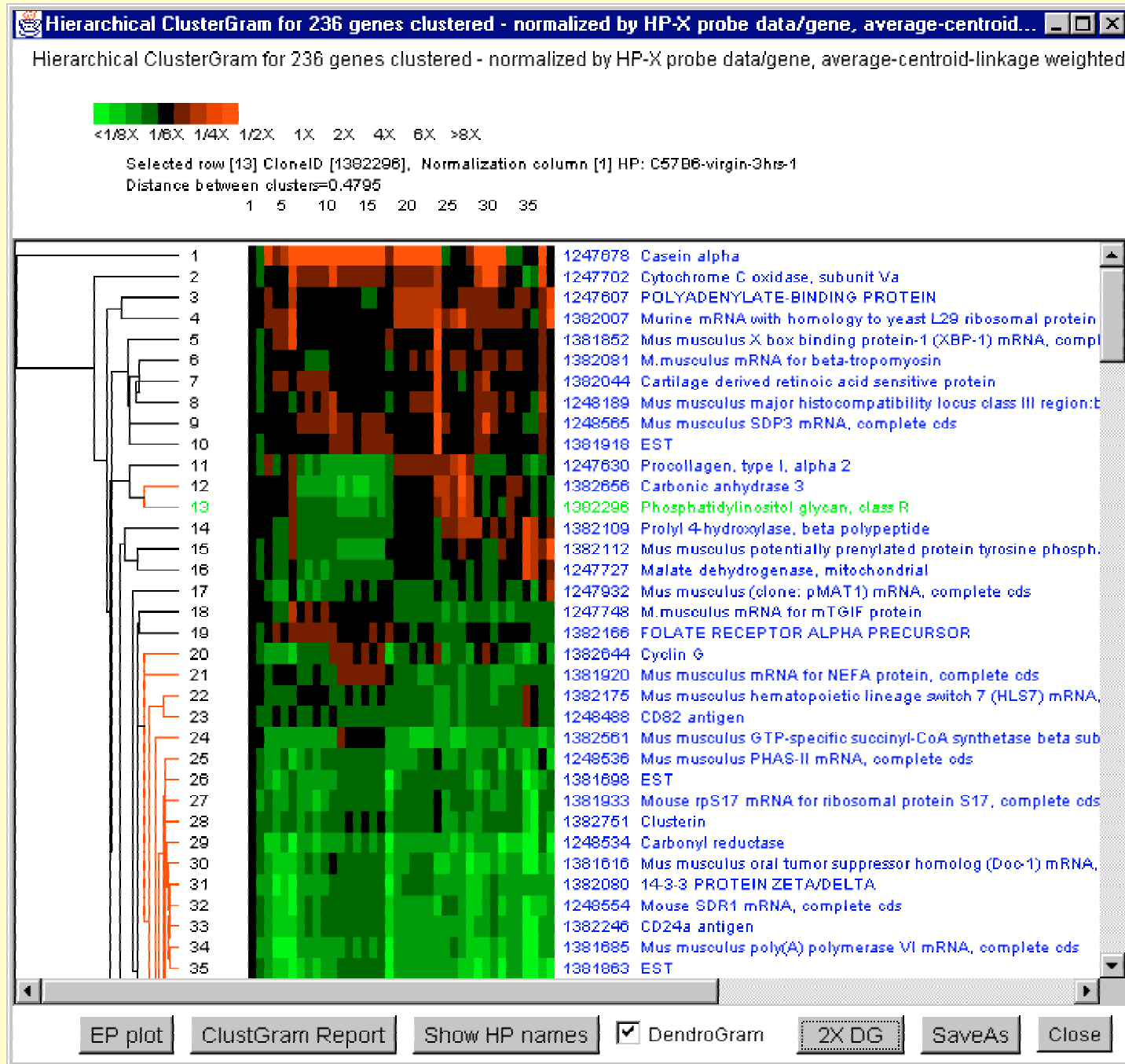
- Scrollable list of EP plots showing genes from clusters #1, #2, #3



I.10b Mean Expr. Profiles of K-means Clusters



I.11 Hierarchical Clusters of Similar Expression



I.12 The Data Mining Paradigm: “Drill Down” to Discover Patterns

Start

|

v

Have initial model of what may be related

|

v

+----->

Organize samples into sets of conditions

Set data pre-filters (normalization, stat. Filters, etc)

Examine Plots (scatter, expression, histograms, etc)

Cluster current gene subset and view cluster plots

|

|

|

Refine views

|

v

+<-----

Evaluate results for interesting data relationships

|

v

^

|

+<-----

Save interesting gene sets

|

Found interesting results, make reports, export results

v

Done

Summary

- MAExplorer is a flexible microarray data-mining tool running on the user's computer
- Uses direct-manipulation, data filtering, built-in graphics, statistics, clustering, gene and sample set operations
- Manages multiple samples, replicates, sets, expression profile lists where the state may be saved on the disk for later use
- Cvt2Mae data conversion tools allows use with common chips
- Plug-ins will allow extension with new analytic methods by users
- MAExplorer identified genes in preferentially expressed during lactation, *Nucleic Acids Res.* (2000) **28**:4452
- Freely available for download with documentation on Web site