Introduction to Data Mining of Microarrays using the MicroArray Explorer

Peter F. Lemkin

Lab. Experimental & Computational Biology, CCR, NCI
Frederick, MD 21702

MAExplorer: **http://www.lecb.ncifcrf.gov/MAExplorer**

Rev: 10-27-2001

Topics to be covered

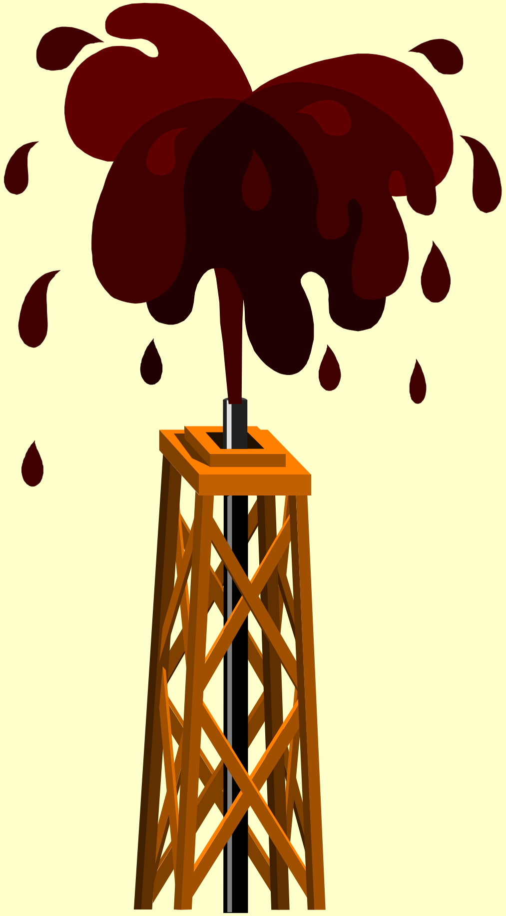
- **Need for data mining**
  1. What do you do with all that data?
  2. How do you manipulate it and find interesting correlations between particular genes and experimental conditions?

- **Capabilities of MAExplorer**
  1. Direct-manipulation data mining: graphics, statistics, clustering
  2. Freely available for download from Web to run on your computer
  3. Integrated with NCI/CIT mAdb server (nciarray.nci.nih.gov) to analyze your data on that server.

Outline

- I. Data Mining of microarray data
- II. MicroArray Explorer
- III. Installing MAExplorer on your computer
- IV. Using NCI/CIT mAdb data with MAExplorer

I. Data Mining of Microarrays



Outline

1. The problem

2. Types of experiments

3. Quantified data used

4. Normalization of data

5. Expression profiles

6. Clustering methods

7. Partition samples by 2 conditions or ordered list

8. Refine the search criteria

I. The Problem

- We assume we have a spreadsheet of quantified microarray 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.

- **Correlation of gene expression changes** with biological state implies a relationship but does not imply cause and effect

Types of Experiments

- What **types of expression** could we analyze?

- Look at expression patterns:
  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

Things To Consider in Data Mining:

- Initially, **don’t know what patterns to look for**
- Could **hypothesize experiments** where changes might be expected
- Then **look for the differences between patterns**
- How do these tools help **find patterns**?
- By visual, statistical and clustering methods

Example: the fold-change problem

- A measure of difference between 2 samples is “fold change”

  $f(x,y) = x/y$

- However f is sensitive to noise. If noise in all measurements is constant e, then $f_e(x,y,e)$ has a range of values

  $[\ (x-e)/(y+e)$ to $(x+e)/(y-e)\ ]$

- ***Example:*** for two points (x,y) = (6,3) & (600,300), and e = 0.5 then the range of fold change for these two points is

  $f(6,3)$ = 2.0

  $f_e(6,3,.5)$ = [5.5/3.5 to 6.5/2.5] = [**1.57 to 2.6**],

  and

  $f(600,300)$ = 2.0

  $f_e(600,300,.5)$ = [559.5/300.5 to 600.5/299.5] = [**1.995 to 2.005**].

[I. Kohane, Apr, 2001]

Quantified Data Used in Microarray Analysis

- 1) **Sets of samples** using either intensity (^{33}P radio-labeled) or ratio (Cy3/Cy5 fluorescent-labeled) DNA
- 2) Each **hybridized sample** contains thousands of spots correlated to spotted clones or oligonucleotides (denoted “genes” in MAExplorer)
- **If ^{33}P,** then normalize data **between** hybridized array samples by large numbers of common clones
- **If (Cy3, Cy5),** then use either Cy3 or Cy5 to normalized standard sample **within** an array sample

Dividing samples into 2-condition sets and ordered N-conditions sample lists

- The **2-class division allows using sets of replicates** for computing better gene expression estimates and allows using t-Tests etc. to determine statistical significance

- The **ordered N-list of samples** is used to represent an ordered time-series, development stages, drug-dose response, etc.

- [***In MAExplorer***]: 2-class data is represented by **HP-X and HP-Y** sets and an ordered list of N-samples data is represented by the **HP-E** expression profile list

Normalize intensity data (^{33}P) between samples

- Assuming linearity, for each array sample **j** get an **estimate T_j of total cDNA labeling** for a common subset of genes

- **Methods for estimating T_j** : mean, median, log median, Zscore, log Zscore, sum of calibration DNA, sum of gene set, etc.

- **Compute T_j over specific gene set**: calibration genes, all genes on the array, specific subset of genes

- Scale spot data **within** each sample (for samples 1 and 2, gene **k**):

$$s^*_{1,k} = s_{1,k} / T_1$$

and

$$s^*_{2,k} = s_{2,k} / T_2$$

- Then, we may **compare** normalized $s^*_{1,k}$ **and** $s^*_{2,k}$ values

Normalize ratio data (Cy3, Cy5) between samples

- Let Cy5-labeled spots be the **standard sample** hybridized to all arrays (could use Cy3 instead). Independent samples are labeled with Cy3

- Cy3 Data **within** each sample is scaled by corresponding Cy5 spot values (samples 1 and 2, and all genes k) to compute ratio values $\mathbf{s^r}$ where Cy5 labeled samples are common between samples 1 and 2:

  $$\mathbf{s^r_{1k} = s_{1k,cy3} / s_{1k,cy5}}$$

  and

  $$\mathbf{s^r_{2k} = s_{2k,cy3} / s_{2k,cy5}}$$

- Then scale ($\mathbf{s^*_{1k}, s^*_{2k}}$) from ($\mathbf{s^r_{1k}, s^r_{2k}}$) as for Intensity data.

- Then, we may **compare** the normalized $\mathbf{s^*_{1k}}$ and $\mathbf{s^*_{2k}}$ values

Definition: Gene Expression Profile

- An **expression profile** $\mathbf{e}_j$ of an ordered list of **N** normalized spot values samples $\mathbf{v}_{jk}$ (k=1 to **N**) for a particular gene **j**

- The expression profile for a particular gene **j** is:

  $$\mathbf{e}_j = (\mathbf{v}_{j1}, \mathbf{v}_{j2}, \mathbf{v}_{j3}, \ldots, \mathbf{v}_{jN})$$

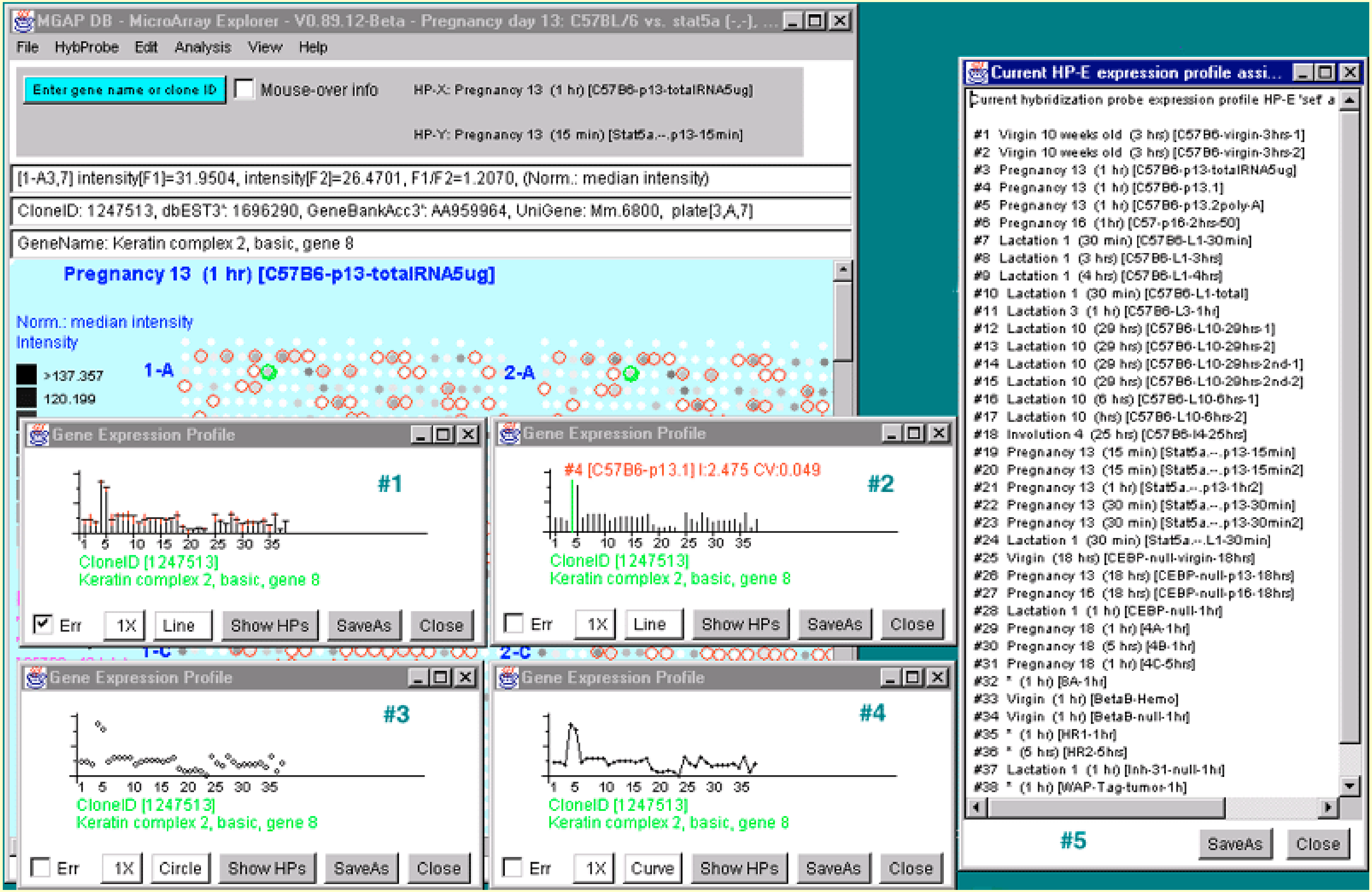
- A **difference** between two genes **p** and **q** may be estimated as a **N**-dimensional metric "**distance**" between $\mathbf{e}_p$ and $\mathbf{e}_q$

- **Euclidean distance:** $\mathbf{d}_{pq} = (1/N \sum_{j=1:N} (v_{jp} - v_{jq})^2)^{1/2}$

- Other distance measures: correlation coefficient, city-block, etc.

- If distance is scaled to [0:1], then **Similarity** measure: $\mathbf{s}_{pq} = 1 - \mathbf{d}_{pq}$

I.1 Expression profile plots - examples



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

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

Gene Expression Profile

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

Gene Expression Profile

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

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

Gene Expression Profile

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

Gene Expression Profile

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

Current HP-E expression profile assi...

Current hybridization probe expression profile HP-E 'sef 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 16 (1hr) [C57-p16-2hrs-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 (29 hrs) [C57B6-L10-29hrs-1]
#13 Lactation 10 (29 hrs) [C57B6-L10-29hrs-2]
#14 Lactation 10 (29 hrs) [C57B6-L10-29hrs-2nd-1]
#15 Lactation 10 (29 hrs) [C57B6-L10-29hrs-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 16 (18 hrs) [CEBP-null-p16-18hrs]
#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-1h]

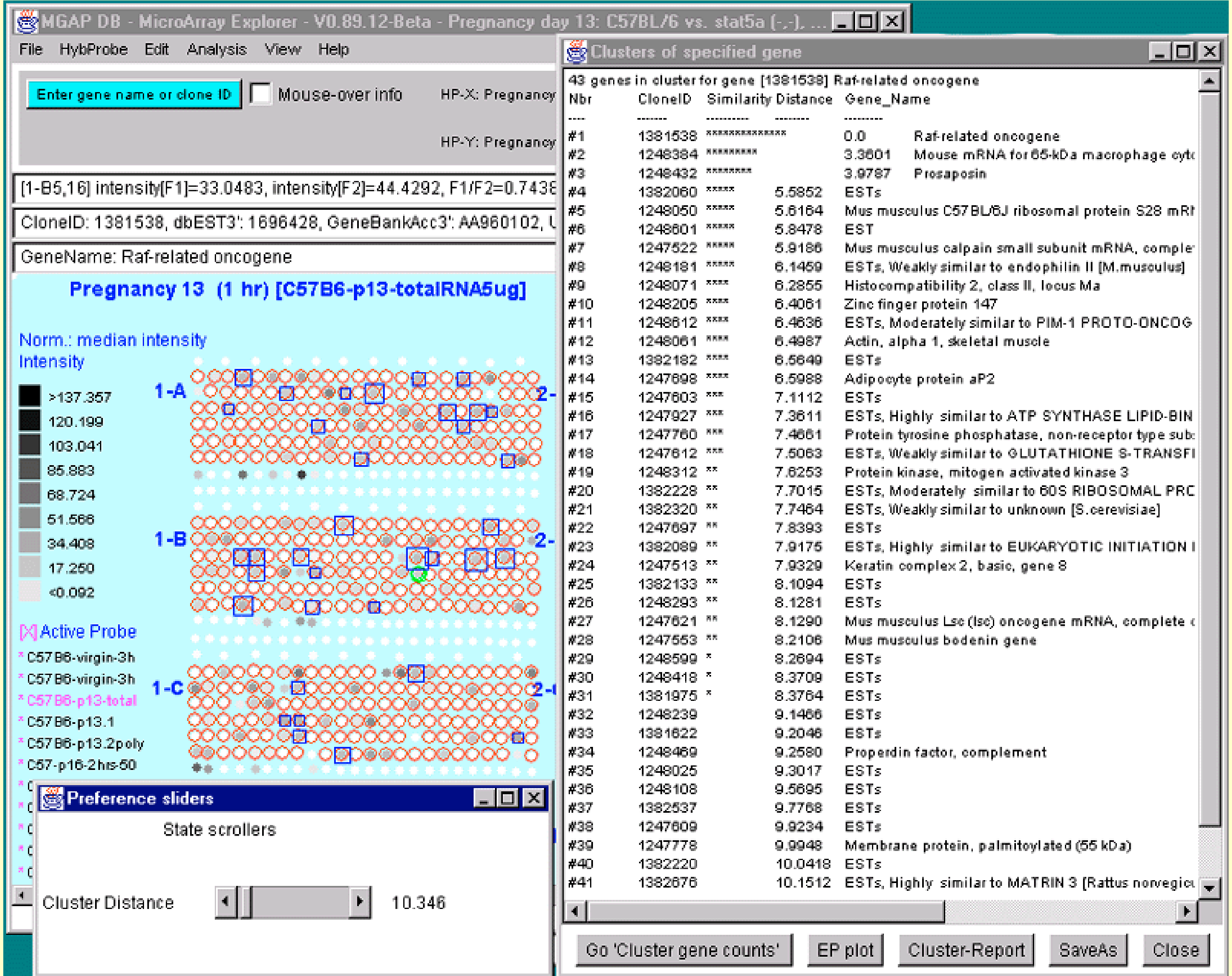
Why Do We Need to Cluster the Data?

- Clusters represent one way to **identify similar gene expression** across a set of experiment samples

- **Many ways to cluster** the data:
  - C.1 Find genes with similar expression
  - C.2 K-means clusters where the number of clusters K is fixed
  - C.3 Hierarchical clustering where a binary hierarchy is created
  - C.*n* Other methods: Self Organizing Memory (SOM), fuzzy clustering, Support Vector Machines (SVM), etc.

C.1 Finding similar genes

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



[1-B5,16] intensity[F1]=33.0483, intensity[F2]=44.4292, F1/F2=0.7438

CloneID: 1381538, dbEST3': 1696428, GeneBankAcc3': AA960102, U

GeneName: Raf-related oncogene

```
43 genes in cluster for gene [1381538] Raf-related oncogene
Nbr    CloneID  Similarity Distance  Gene_Name
----   -------  ---------- --------  ---------
#1     1381538  **************       0.0      Raf-related oncogene
#2     1248384  **********           3.3601   Mouse mRNA for 65-kDa macrophage cytc
#3     1248432  *********            3.9787   Prosaposin
#4     1382060  *****       5.5852   ESTs
#5     1248050  *****       5.6164   Mus musculus C57BL/6J ribosomal protein S28 mRI
#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  ****        6.4061   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    1247697  **          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 (lsc) oncogene mRNA, complete c
#28    1247553  **          8.2106   Mus musculus bodenin gene
#29    1248599  *           8.2694   ESTs
#30    1248418  *           8.3709   ESTs
#31    1381975  *           8.3764   ESTs
#32    1248239              9.1466   ESTs
#33    1381622              9.2046   ESTs
#34    1248469              9.2580   Properdin factor, complement
#35    1248025              9.3017   ESTs
#36    1248108              9.5695   ESTs
#37    1382537              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 norvegicu
```

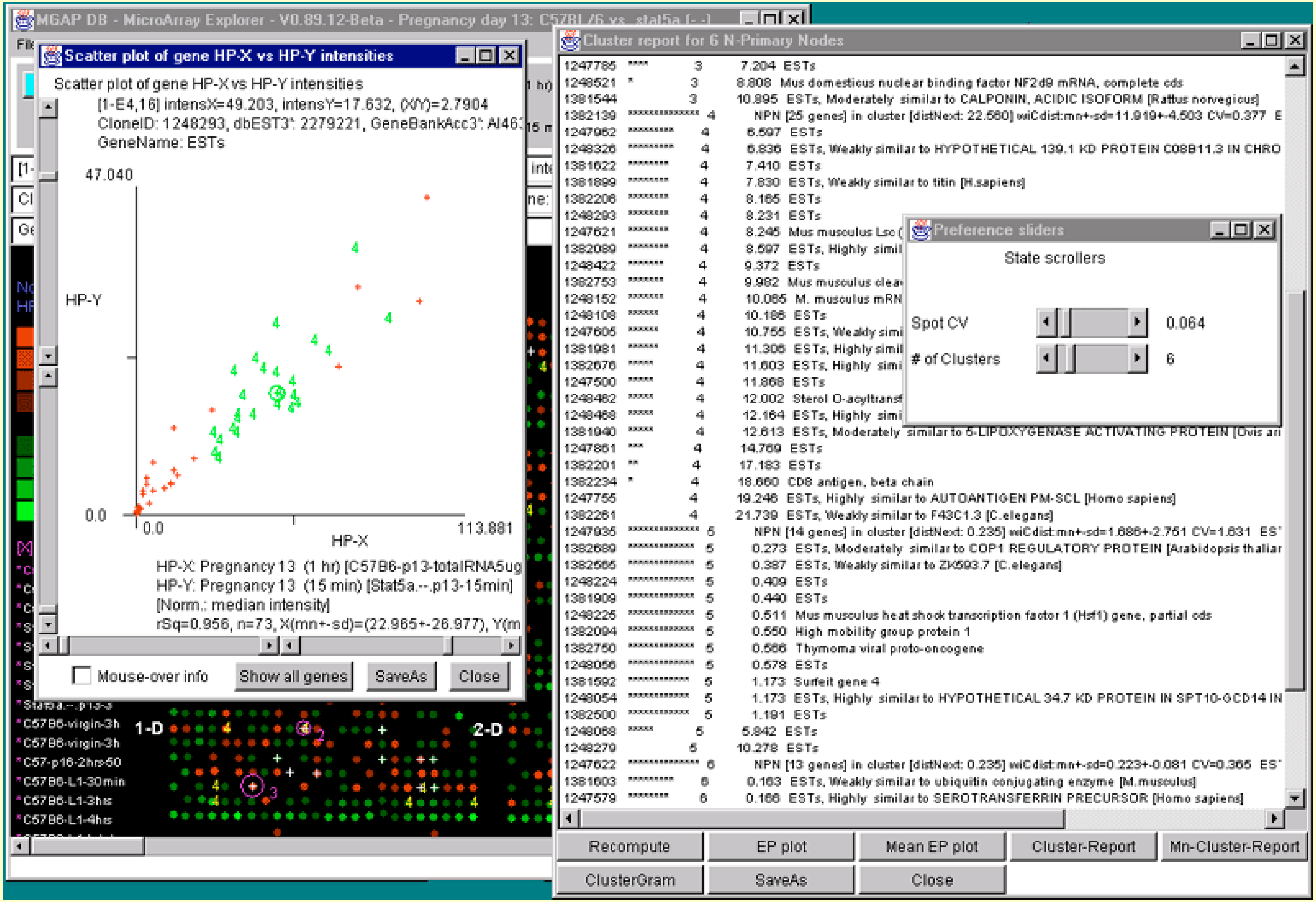
C.2 K-means Clustering

- **K-means clustering** finds **K** clusters of similar genes. Could use variance of clusters to determine if split into sub-clusters by increasing **K**

- Don't need distance matrix - faster clustering large numbers of **N** genes

- **Algorithm**:
  1. Pick seed gene **s** and put it into cluster 1 (let **k=1**)
  2. For all clusters **j=1 to k** , find gene **q** such that $\mathbf{d_{jq}}$ is a maximum
  3. Set **k=k+1**. Put gene **q** into new cluster **k**
  4. For **j= k to K**, repeat steps 2 and 3 until there are **K** clusters
  5. Then, assign **(N-K)** remaining genes **q** into one of the **K** clusters **j** with minimum $\mathbf{d_{jq}}$
  6. Compute new *virtual* genes as means $\{\mathbf{e_k}\}$ for each of **K** clusters
  7. Reassign all **N** genes **q** into **K** new clusters with minimum $\mathbf{d_{pq}}$ using virtual genes $\{\mathbf{e_p}\}$
  8. Variants: use multiple seed genes, range of K values, minimize COV

I.2 Example of K-means clustering



Scatter plot of gene HP-X vs HP-Y intensities

[1-E4,16] intensX=49.203, intensY=17.632, (X/Y)=2.7904
CloneID: 1248293, dbEST3': 2279221, GeneBankAcc3': AI46
GeneName: ESTs

HP-X: Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug
HP-Y: Pregnancy 13 (15 min) [Stat5a.--.p13-15min]
[Norm.: median intensity]
rSq=0.956, n=73, X(mn+-sd)=(22.965+-26.977), Y(m

```
1247785  ****             3      7.204  ESTs
1248521  *               3      8.808  Mus domesticus nuclear binding factor NF2d9 mRNA, complete cds
1381544                  3     10.895  ESTs, Moderately  similar to CALPONIN, ACIDIC ISOFORM [Rattus norvegicus]
1382139  ***************   4     NPN [25 genes] in cluster [distNext: 22.580] wiCdist:mn+-sd=11.919+-4.503 CV=0.377  E
1247962  **********       4      6.597  ESTs
1248326  **********       4      6.836  ESTs, Weakly similar to HYPOTHETICAL 139.1 KD PROTEIN C08B11.3 IN CHRO
1381622  *********        4      7.410  ESTs
1381899  *********        4      7.830  ESTs, Weakly similar to titin [H.sapiens]
1382206  *********        4      8.165  ESTs
1248293  *********        4      8.231  ESTs
1247621  *********        4      8.246  Mus musculus Lsc (
1382089  *********        4      8.597  ESTs, Highly  simil
1248422  ********         4      9.372  ESTs
1382753  ********         4      9.982  Mus musculus clea
1248152  ********         4     10.065  M. musculus mRN
1248108  *******          4     10.186  ESTs
1247605  *******          4     10.755  ESTs, Weakly simi
1381981  *******          4     11.306  ESTs, Highly simil
1382676  ******           4     11.603  ESTs, Highly  simi
1247500  ******           4     11.868  ESTs
1248462  ******           4     12.002  Sterol O-acyltransf
1248468  ******           4     12.164  ESTs, Highly  simi
1381940  ******           4     12.613  ESTs, Moderately  similar to 5-LIPOXYGENASE ACTIVATING PROTEIN [Ovis ari
1247861  ***              4     14.769  ESTs
1382201  **               4     17.183  ESTs
1382234  *                4     18.660  CD8 antigen, beta chain
1247755                   4     19.246  ESTs, Highly  similar to AUTOANTIGEN PM-SCL [Homo sapiens]
1382261                   4     21.739  ESTs, Weakly similar to F43C1.3 [C.elegans]
1247935  ***************   5     NPN [14 genes] in cluster [distNext: 0.235] wiCdist:mn+-sd=1.686+-2.751 CV=1.631  ES
1382689  ***************   5      0.273  ESTs, Moderately  similar to COP1 REGULATORY PROTEIN [Arabidopsis thaliar
1382565  ***************   5      0.387  ESTs, Weakly similar to ZK593.7 [C.elegans]
1248224  ***************   5      0.409  ESTs
1381909  ***************   5      0.440  ESTs
1248225  ***************   5      0.511  Mus musculus heat shock transcription factor 1 (Hsf1) gene, partial cds
1382094  ***************   5      0.550  High mobility group protein 1
1382750  ***************   5      0.566  Thymoma viral proto-oncogene
1248056  ***************   5      0.578  ESTs
1381592  **************    5      1.173  Surfeit gene 4
1248054  **************    5      1.173  ESTs, Highly  similar to HYPOTHETICAL 34.7 KD PROTEIN IN SPT10-GCD14 IN
1382500  **************    5      1.191  ESTs
1248068  ******           5      5.842  ESTs
1248279                   5     10.278  ESTs
1247622  ***************   6     NPN [13 genes] in cluster [distNext: 0.235] wiCdist:mn+-sd=0.223+-0.081 CV=0.365  ES
1381603  **********        6      0.163  ESTs, Weakly similar to ubiquitin conjugating enzyme [M.musculus]
1247579  *********         6      0.166  ESTs, Highly  similar to SEROTRANSFERRIN PRECURSOR [Homo sapiens]
```

State scrollers

Spot CV 0.064

# of Clusters 6

C.3 Hierarchical clustering

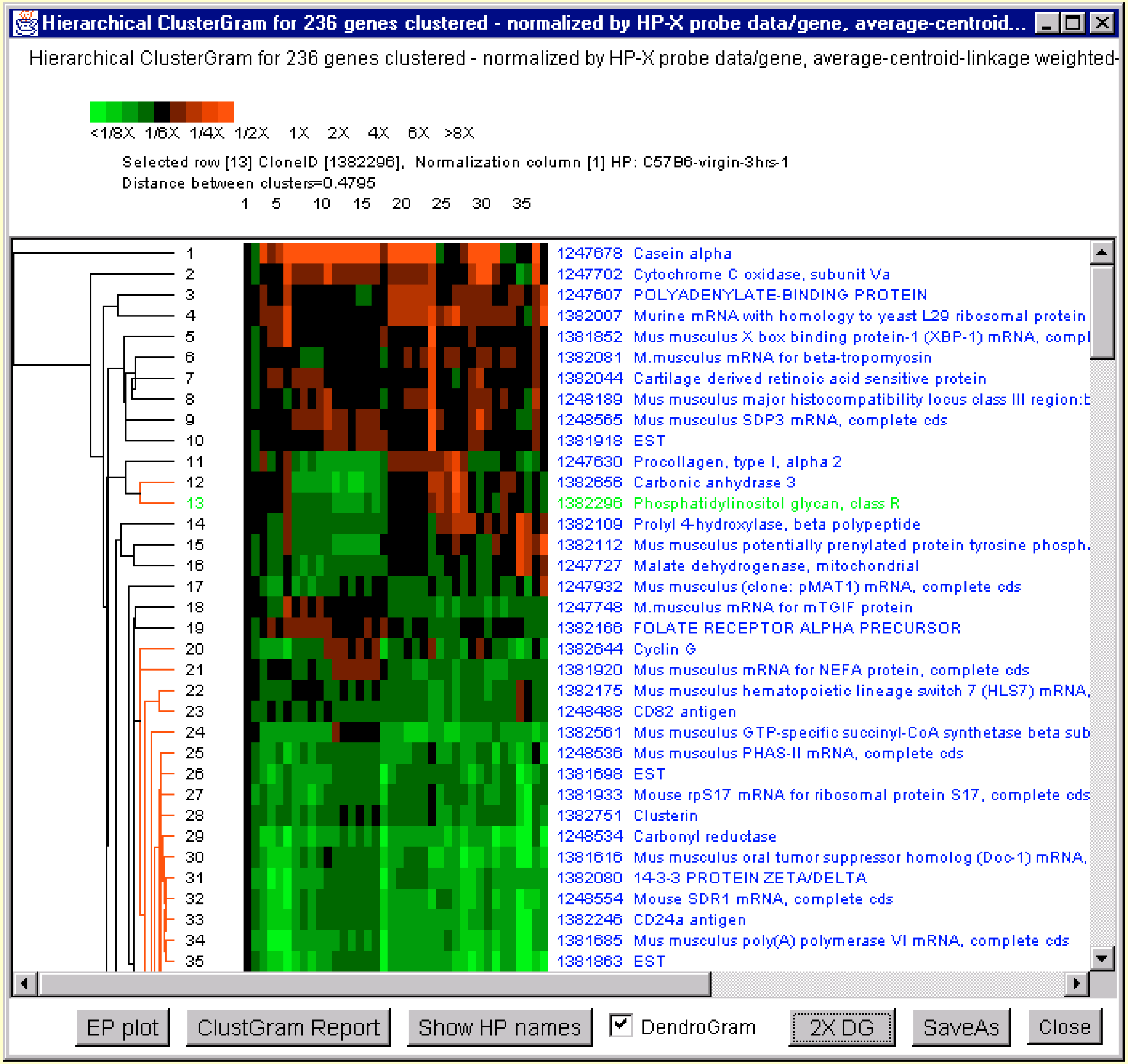
- **Hierarchical clustering** requires a distance matrix. For **N** genes (terminal gene clusters), it generates **2N-1** clusters.

- **Distance matrix** is upper diagonal matrix **D** of $\mathbf{d_{pq}}$ of size **N(N-1)/2**

- **D** can get quite large for clustering a large number of genes **N** [for **N**=5000, this is > 50 Mbytes!]

- **Algorithm**:
  1. Assign all **N** genes to clusters 1 to **N**, set **n** to **N**
  2. Find two clusters **p** and **q** such that $\mathbf{d_{pq}}$ is a minimum
     - 2.1 Compute a *virtual* cluster vector $\mathbf{e_{p,q}}$ = *average* $(\mathbf{e_p},\mathbf{e_q})$
     - 2.2 Set **n = n+1**
     - 2.3 Assign "virtual" cluster to new cluster **n** with estimated value $\mathbf{e_{p,q}}$
  3. Repeat step 2 until **n = 2N-1**.

I.3 Example of Hierarchical Clustering



Hierarchical ClusterGram for 236 genes clustered - normalized by HP-X probe data/gene, average-centroid...

Hierarchical ClusterGram for 236 genes clustered - normalized by HP-X probe data/gene, average-centroid-linkage weighted-

Selected row [13] CloneID [1382296], Normalization column [1] HP: C57B6-virgin-3hrs-1
Distance between clusters=0.4795

1	1247678	Casein alpha
2	1247702	Cytochrome C oxidase, subunit Va
3	1247607	POLYADENYLATE-BINDING PROTEIN
4	1382007	Murine mRNA with homology to yeast L29 ribosomal protein
5	1381852	Mus musculus X box binding protein-1 (XBP-1) mRNA, compl
6	1382081	M.musculus mRNA for beta-tropomyosin
7	1382044	Cartilage derived retinoic acid sensitive protein
8	1248189	Mus musculus major histocompatibility locus class III region:b
9	1248565	Mus musculus SDP3 mRNA, complete cds
10	1381918	EST
11	1247630	Procollagen, type I, alpha 2
12	1382656	Carbonic anhydrase 3
13	1382296	Phosphatidylinositol glycan, class R
14	1382109	Prolyl 4-hydroxylase, beta polypeptide
15	1382112	Mus musculus potentially prenylated protein tyrosine phosph.
16	1247727	Malate dehydrogenase, mitochondrial
17	1247932	Mus musculus (clone: pMAT1) mRNA, complete cds
18	1247748	M.musculus mRNA for mTGIF protein
19	1382166	FOLATE RECEPTOR ALPHA PRECURSOR
20	1382644	Cyclin G
21	1381920	Mus musculus mRNA for NEFA protein, complete cds
22	1382175	Mus musculus hematopoietic lineage switch 7 (HLS7) mRNA,
23	1248488	CD82 antigen
24	1382561	Mus musculus GTP-specific succinyl-CoA synthetase beta sub
25	1248536	Mus musculus PHAS-II mRNA, complete cds
26	1381698	EST
27	1381933	Mouse rpS17 mRNA for ribosomal protein S17, complete cds
28	1382751	Clusterin
29	1248534	Carbonyl reductase
30	1381616	Mus musculus oral tumor suppressor homolog (Doc-1) mRNA,
31	1382080	14-3-3 PROTEIN ZETA/DELTA
32	1248554	Mouse SDR1 mRNA, complete cds
33	1382246	CD24a antigen
34	1381685	Mus musculus poly(A) polymerase VI mRNA, complete cds
35	1381863	EST

Data mining:

- ***Data mining is a <u>pattern discovery activity</u> - use all the tools you have***.

- *It is **open-ended** because of the variety of ways data may be partitioned, normalized, pre-filtered, clustered, and viewed.*

- *When data mining microarray data, look at correlated genes from the point of view of **what relationships might be interesting from a biological view**. I.e. check out the results with PubMed, genomic databases, other lab experiments, etc.*

I.4 The Data Mining Paradigm: the Refinement Process

```
            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
```

A Possible Analysis Scenario

1. **Select set of samples** from database

2. **Organize samples** as 2-class (X vs Y) sets or ordered list of N samples

3. **Select normalization** method

4. **Preview the data** with scatter plots and histograms

5. **Restrict search** using data filter to pre-filter a robust set of genes

6. **Cluster genes** & visualize with EP plots, clustergram, dendrogram, etc

7. **Make report** and **access genomic Web databases** with resulting genes

8. **Save results** for later use or continued investigation

II. MicroArray Explorer (MAExplorer)

Outline

1. Description

2. Importing data

3. Examples of analysis capabilities

II. What is the MicroArray Explorer?

- **MAExplorer is a Java stand-alone (off-line) or applet (Web-based) microarray real-time data-mining tool**
- Install **stand-alone** from the Web site for MS Windows, MacOS, Solaris, Linux, Unix
- Helps makes sense of **large complex sample data sets** with replicates
- Data mining is accomplished using **data filtering** with **direct manipulation** of data in **graphics** and **spreadsheets**
- Data filtering includes set-operations, statistics and clustering
- MAExplorer **handles a variety of quantified microarray data**

MAExplorer Home Page

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

MAExplorer

Introduction
Demonstrations

Documentation
- Manual (on right)
- Manual (new window)
- Manual (7Mb Zip)
- Manual (entire)
- Newsletters

- Short Tutorial
- Advanced tutorial
- Menu summary
- Quick Start
- Glossary
- Index
- Help Desk

- Overview (PDF)
- Examples (PDF)
- Intro Data Mining (PDF)
- *N.A.R.* paper (PDF)
- Use with mAdb (PDF)

Downloading MAExplorer
- User's array data
- Cvt2Mae data converter
- Stand-alone version
- NEW Revision notes
- Installer information

- download

Disclaimer

MGAP-MAExplorer home

Contact [Lemkin | lemkin@ncifcrf.gov | (LECB, NCI/FCRDC)]

Revised: $Date: 2001/08/10 18:05:25 $

MAExplorer - MicroArray Explorer

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.

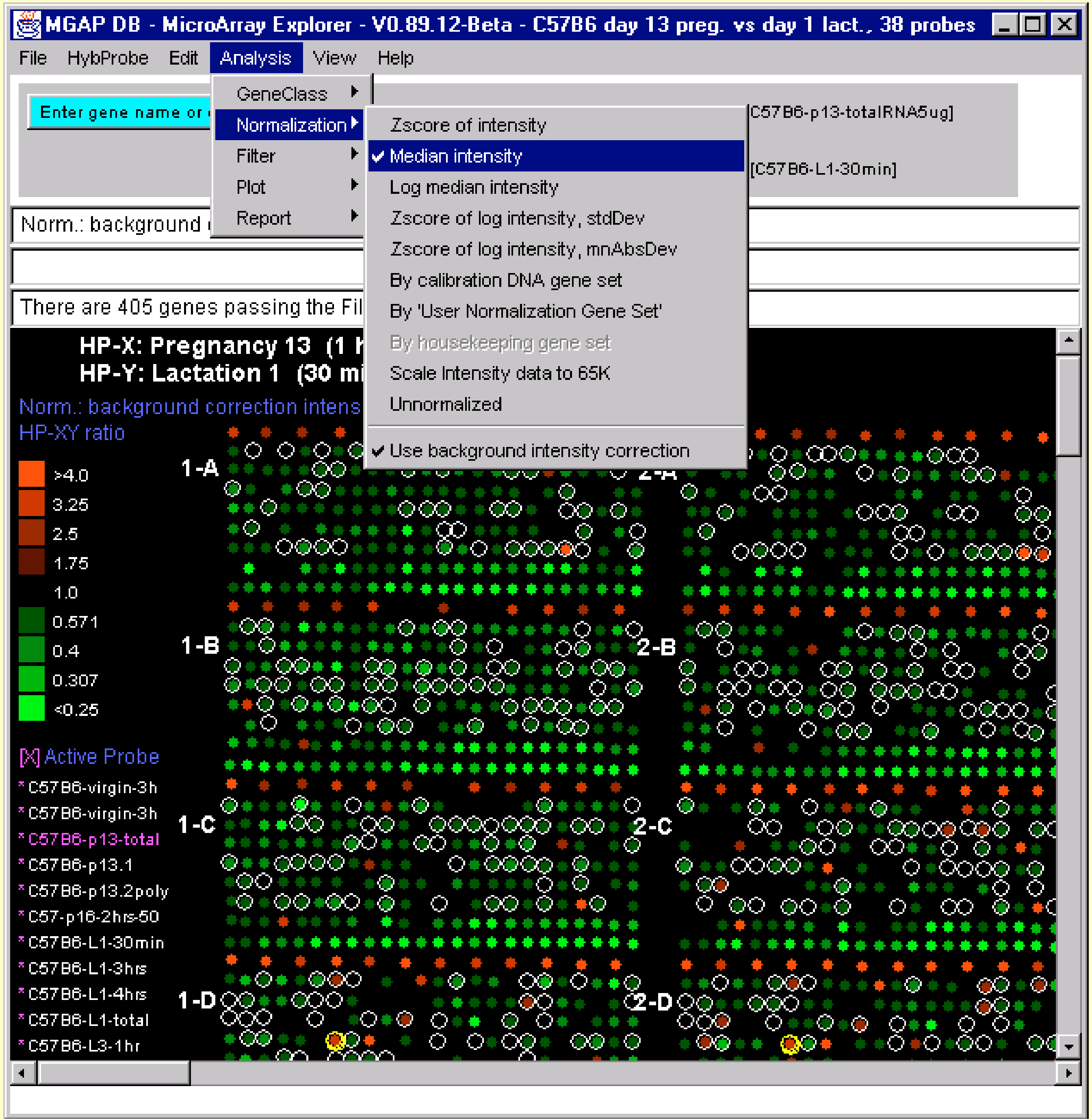
Microarray data may be viewed and directly manipulated in array pseudoimages, scatter plots, histograms, expression profile plots, cluster analyses (similar clones, K-means, hierarchical clusters, etc.), and reports. A key feature is the clone data filters for constraining a working set of clones to those passing a variety of user-specifed 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.

II.1 MAExplorer Menu Interface



MGAP DB - MicroArray Explorer - V0.89.12-Beta - C57B6 day 13 preg. vs day 1 lact., 38 probes

Enter gene name or

[C57B6-p13-totalRNA5ug]

[C57B6-L1-30min]

Norm.: background

There are 405 genes passing the Fil

HP-X: Pregnancy 13 (1 h

HP-Y: Lactation 1 (30 mi

Norm.: background correction intens

✓ Median intensity

✓ Use background intensity correction

What is the MicroArray Explorer? (continued)

- Developed for **Mammary Genome Anatomy Program**
  http://www.lecb.ncifcrf.gov/mae

- First use **statistical data filters to pre-filter** data (eg. sets of genes) so remaining data is robust

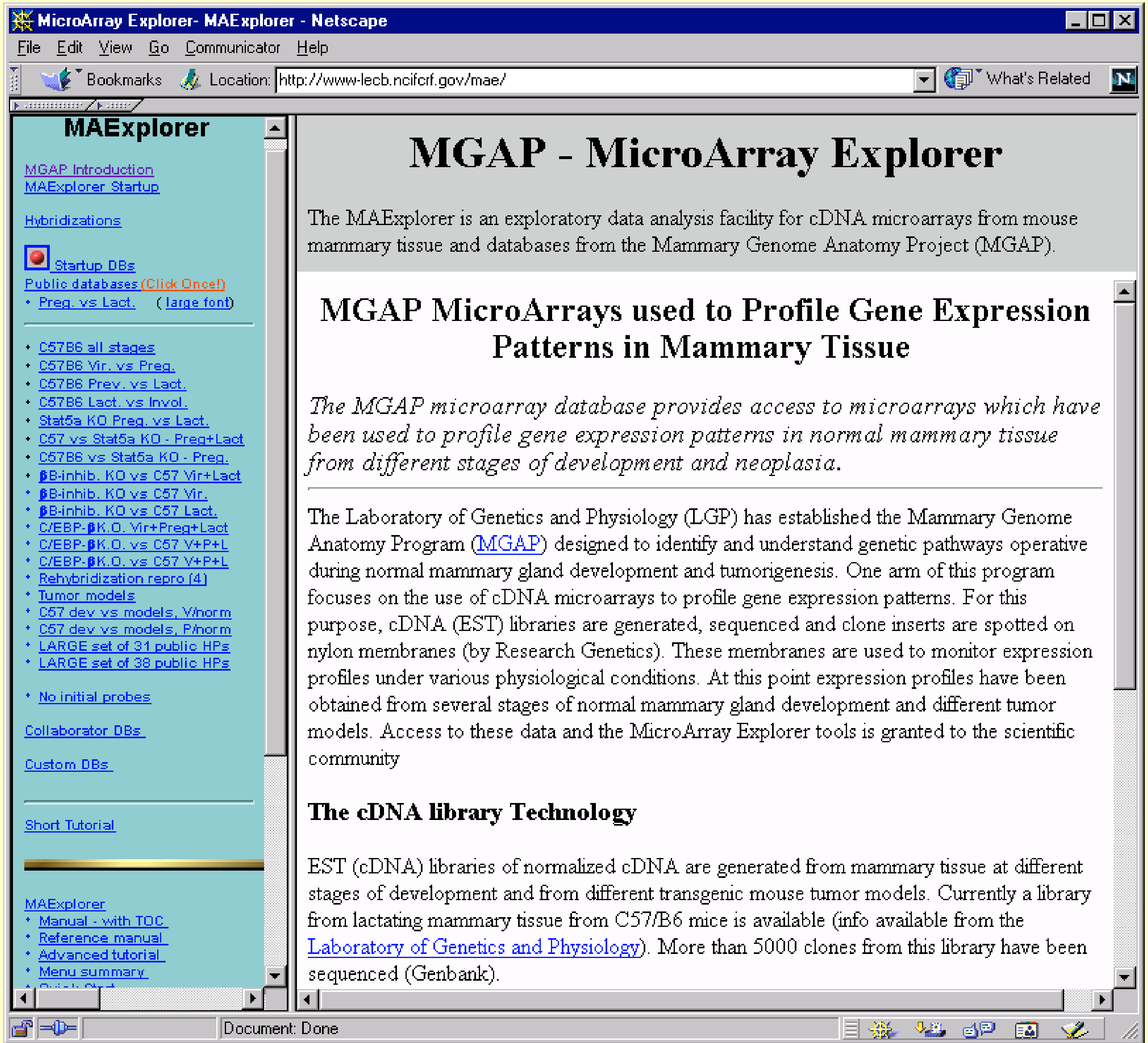
- Then use methods such as **cluster analysis to discover patterns** observed with direct-manipulation graphical plots and reports

- Save, restore, and compare results using **gene sets** and **condition lists.** Save current state of **data mining analyses** locally in files (i.e. “bookmark”)

- Access third-party **genomic data** such as UniGene using links to Web databases

- **Online documentation** (HTML manual, tutorials, examples, etc.) on Web site

II.2 Mammary Geneome Anatomy Program
MAExplorer http://www.lecb.ncifcrf.gov/mae



MAExplorer

[MGAP Introduction](#)
[MAExplorer Startup](#)

[Hybridizations](#)

[Startup DBs](#)
[Public databases](#) (Click Once!)
- [Preg. vs Lact.](#) ([large font](#))

- [C57B6 all stages](#)
- [C57B6 Vir. vs Preg.](#)
- [C57B6 Prev. vs Lact.](#)
- [C57B6 Lact. vs Invol.](#)
- [Stat5a KO Preg. vs Lact.](#)
- [C57 vs Stat5a KO - Preg+Lact](#)
- [C57B6 vs Stat5a KO - Preg.](#)
- [βB-inhib. KO vs C57 Vir+Lact](#)
- [βB-inhib. KO vs C57 Vir.](#)
- [βB-inhib. KO vs C57 Lact.](#)
- [C/EBP-βK.O. Vir+Preg+Lact](#)
- [C/EBP-βK.O. vs C57 V+P+L](#)
- [C/EBP-βK.O. vs C57 V+P+L](#)
- [Rehybridization repro (4)](#)
- [Tumor models](#)
- [C57 dev vs models, V/norm](#)
- [C57 dev vs models, P/norm](#)
- [LARGE set of 31 public HPs](#)
- [LARGE set of 38 public HPs](#)

- [No initial probes](#)

[Collaborator DBs](#)

[Custom DBs](#)

[Short Tutorial](#)

[MAExplorer](#)
- [Manual - with TOC](#)
- [Reference manual](#)
- [Advanced tutorial](#)
- [Menu summary](#)

MGAP - MicroArray Explorer

The MAExplorer is an exploratory data analysis facility for cDNA microarrays from mouse mammary tissue and databases from the Mammary Genome Anatomy Project (MGAP).

MGAP MicroArrays used to Profile Gene Expression Patterns in Mammary Tissue

The MGAP microarray database provides access to microarrays which have been used to profile gene expression patterns in normal mammary tissue from different stages of development and neoplasia.

The Laboratory of Genetics and Physiology (LGP) has established the Mammary Genome Anatomy Program ([MGAP](#)) designed to identify and understand genetic pathways operative during normal mammary gland development and tumorigenesis. One arm of this program focuses on the use of cDNA microarrays to profile gene expression patterns. For this purpose, cDNA (EST) libraries are generated, sequenced and clone inserts are spotted on nylon membranes (by Research Genetics). These membranes are used to monitor expression profiles under various physiological conditions. At this point expression profiles have been obtained from several stages of normal mammary gland development and different tumor models. Access to these data and the MicroArray Explorer tools is granted to the scientific community

The cDNA library Technology

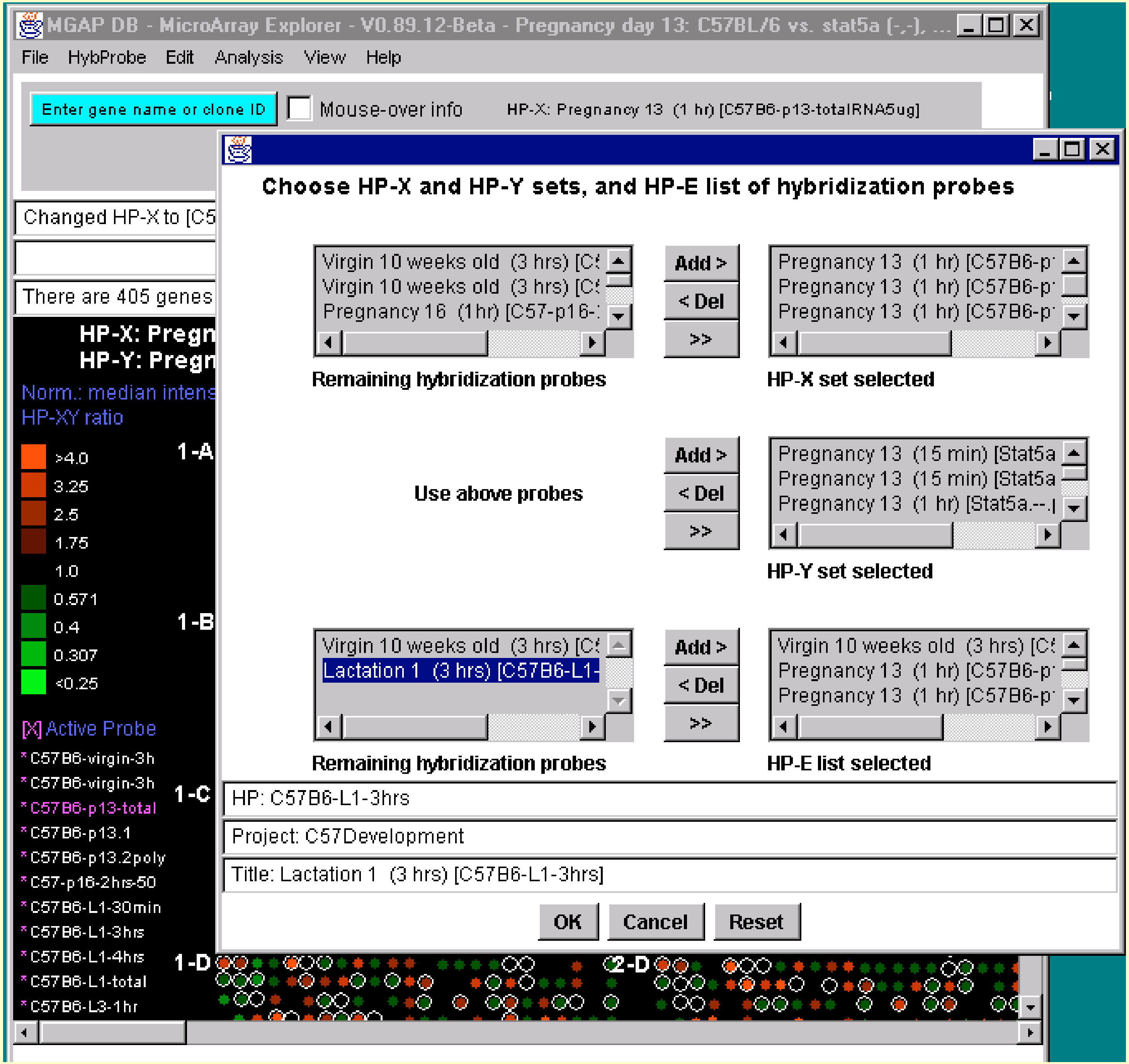
EST (cDNA) libraries of normalized cDNA are generated from mammary tissue at different stages of development and from different transgenic mouse tumor models. Currently a library from lactating mammary tissue from C57/B6 mice is available (info available from the [Laboratory of Genetics and Physiology](#)). More than 5000 clones from this library have been sequenced (Genbank).

Sample Organization

- **Samples are organization by**:
    1. X-Y paired samples
    2. sets of X-Y replicate samples (X and Y-sets)
    3. ordered expression profile lists of samples (E-list)

- **Dynamically choose** hybridized probe samples as **HP-X, HP-Y** and **HP-E**

II.3 Choosing HP-X, HP-Y sets and HP-E lists



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

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

Changed HP-X to [C5

There are 405 genes

HP-X: Pregn
HP-Y: Pregn

Norm.: median intens
HP-XY ratio

>4.0
3.25
2.5
1.75
1.0
0.571
0.4
0.307
<0.25

[X] Active Probe
* C57B6-virgin-3h
* C57B6-virgin-3h
* C57B6-p13-total
* C57B6-p13.1
* C57B6-p13.2poly
* C57-p16-2hrs-50
* 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 probes:
Virgin 10 weeks old (3 hrs) [C5
Virgin 10 weeks old (3 hrs) [C5
Pregnancy 16 (1hr) [C57-p16-

HP-X set selected:
Pregnancy 13 (1 hr) [C57B6-p
Pregnancy 13 (1 hr) [C57B6-p
Pregnancy 13 (1 hr) [C57B6-p

Use above probes

HP-Y set selected:
Pregnancy 13 (15 min) [Stat5a
Pregnancy 13 (15 min) [Stat5a
Pregnancy 13 (1 hr) [Stat5a.--.]

Remaining hybridization probes:
Virgin 10 weeks old (3 hrs) [C5
Lactation 1 (3 hrs) [C57B6-L1-

HP-E list selected:
Virgin 10 weeks old (3 hrs) [C5
Pregnancy 13 (1 hr) [C57B6-p
Pregnancy 13 (1 hr) [C57B6-p

HP: C57B6-L1-3hrs

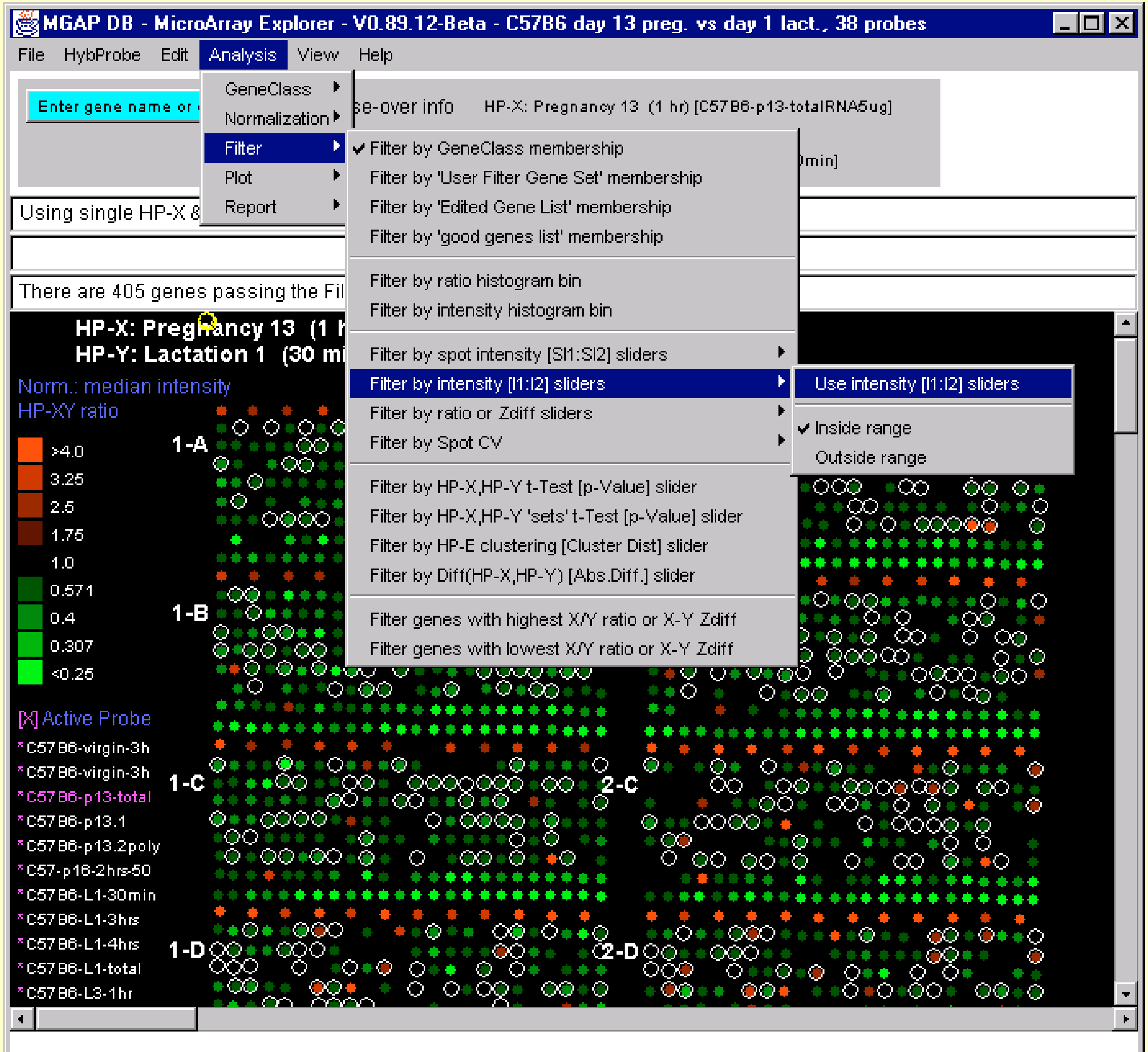
Project: C57Development

Title: Lactation 1 (3 hrs) [C57B6-L1-3hrs]

Data Filters

- **Data filters are used to help converge on genes of interest:**
  1. normalization methods
  2. gene sets
  3. spot intensity and ratio ranges
  4. statistics
  5. clustering (similar-genes, K-means, hierarchical clustering)

II.4 Select One or More Simultaneous Data Filters



MGAP DB - MicroArray Explorer - V0.89.12-Beta - C57B6 day 13 preg. vs day 1 lact., 38 probes

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

Using single HP-X &

There are 405 genes passing the Fil

HP-X: Pregnancy 13 (1 h

HP-Y: Lactation 1 (30 mi

Norm.: median intensity

HP-XY ratio

Data Views Using Pop-up Plots and Reports

- **Plots**: pseudo-array images, scatter-plots, histograms, expression profiles, clustergrams, dendrograms, silhouette-plots
- **Reports**: dynamic genomic Web-accessible spreadsheets, tab-delimited data for Excel
- **Report data**: gene reports, array information, correlation of samples, statistics on subsets of genes or samples
- **Direct manipulation**: select genes from plots and reports, select samples, choose HP-X, HP-Y and HP-E
- **Web linkage to genomic DB**: hyperlinked plots and reports

Sources of Quantified Microarray Data

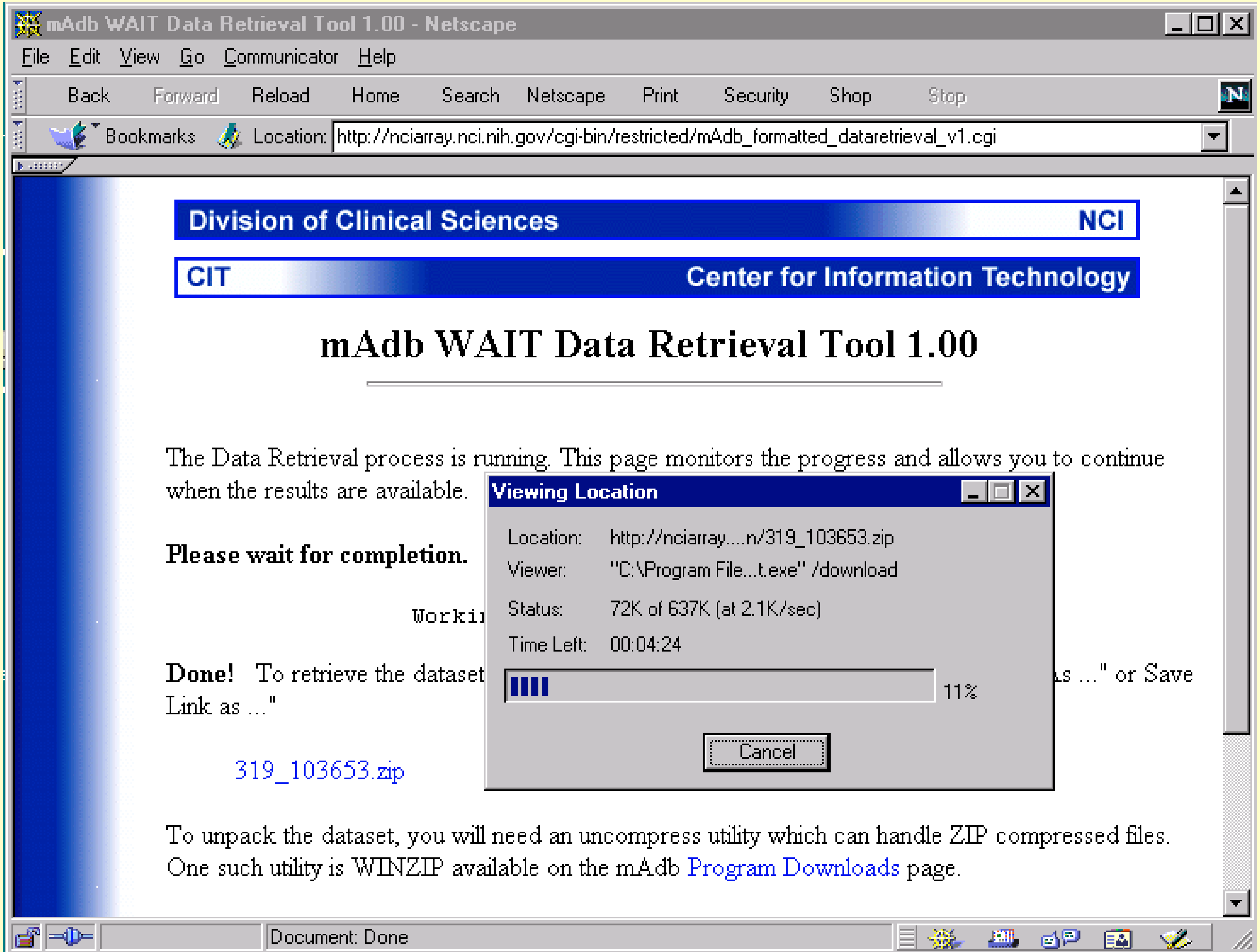
- MAExplorer **handles variety of quantified microarray data**

- Data is specified by array-specific tab-delimited files that include:
    1. **GIPO** file - Gene In Plate Order (i.e. Print) table listing spot grid coords, Clone Id, gene name, GenBank & UniGene Ids, etc.
    2. **Configuration** file describing array geometry, spot labeling, etc.
    3. **Quantification** files of hybridized sample quantified spot data
    4. **Samples DB** file listing the names of the hybridized samples

- Download quantified data from **NCI/CIT-ATC mAdb** database
  http://nciarray.nci.nih.gov/

- Developing Java tool **Cvt2Mae** to convert commercial & academic quantified array data (Incyte, Affymetrix, etc.) to MAExplorer format

II.2a Download NCI/CIT mAdb Data for MAExplorer



Location: http://nciarray.nci.nih.gov/cgi-bin/restricted/mAdb_formatted_dataretrieval_v1.cgi

mAdb WAIT Data Retrieval Tool 1.00

The Data Retrieval process is running. This page monitors the progress and allows you to continue when the results are available.

Please wait for completion.

Worki

Done! To retrieve the dataset …s ..." or Save Link as ..."

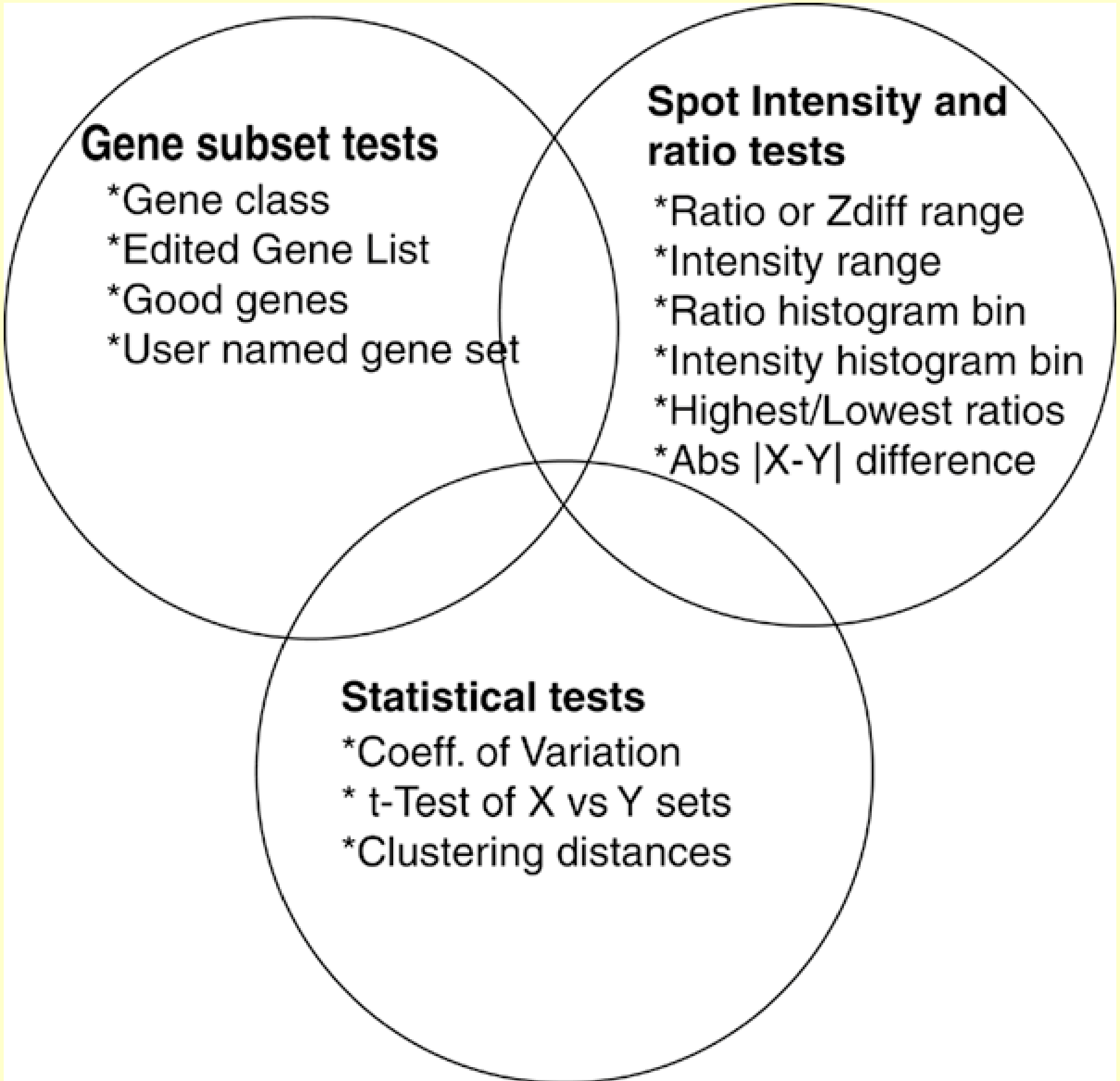
319_103653.zip

To unpack the dataset, you will need an uncompress utility which can handle ZIP compressed files. One such utility is WINZIP available on the mAdb Program Downloads page.

Location: http://nciarray....n/319_103653.zip
Viewer: "C:\Program File...t.exe" /download
Status: 72K of 637K (at 2.1K/sec)
Time Left: 00:04:24
11%

II.3 Gene Data Filter is Intersection of Tests

- Current set of genes is **intersection** of gene sets each passing selected filter tests
- Filtered gene subset is used as **pre-filter** for subsequent clustering, plots, and tables
- Changing any filter parameters causes the data filter to be re-computed



Gene subset tests

- *Gene class
- *Edited Gene List
- *Good genes
- *User named gene set

Spot Intensity and ratio tests

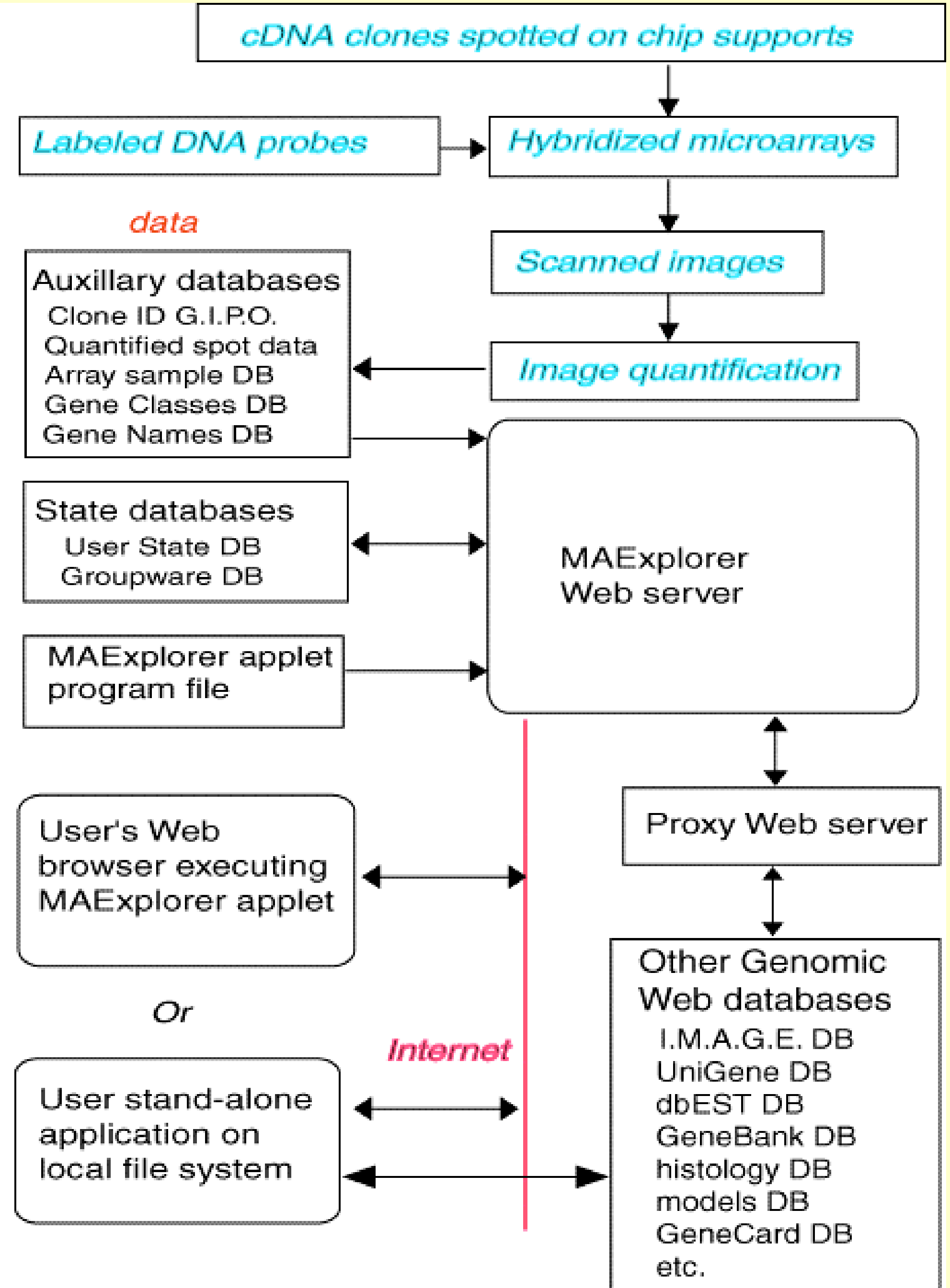
- *Ratio or Zdiff range
- *Intensity range
- *Ratio histogram bin
- *Intensity histogram bin
- *Highest/Lowest ratios
- *Abs |X-Y| difference

Statistical tests

- *Coeff. of Variation
- * t-Test of X vs Y sets
- *Clustering distances

II.4 Overview of MAExplorer Database System

(Steps in cyan are performed before MAExplorer analysis.)



cDNA clones spotted on chip supports

Labeled DNA probes

Hybridized microarrays

data

Scanned images

Auxillary databases
Clone ID G.I.P.O.
Quantified spot data
Array sample DB
Gene Classes DB
Gene Names DB

Image quantification

State databases
User State DB
Groupware DB

MAExplorer
Web server

MAExplorer applet
program file

Proxy Web server

User's Web
browser executing
MAExplorer applet

Or

Internet

Other Genomic
Web databases
I.M.A.G.E. DB
UniGene DB
dbEST DB
GeneBank DB
histology DB
models DB
GeneCard DB
etc.

User stand-alone
application on
local file system

Examples of MAExplorer

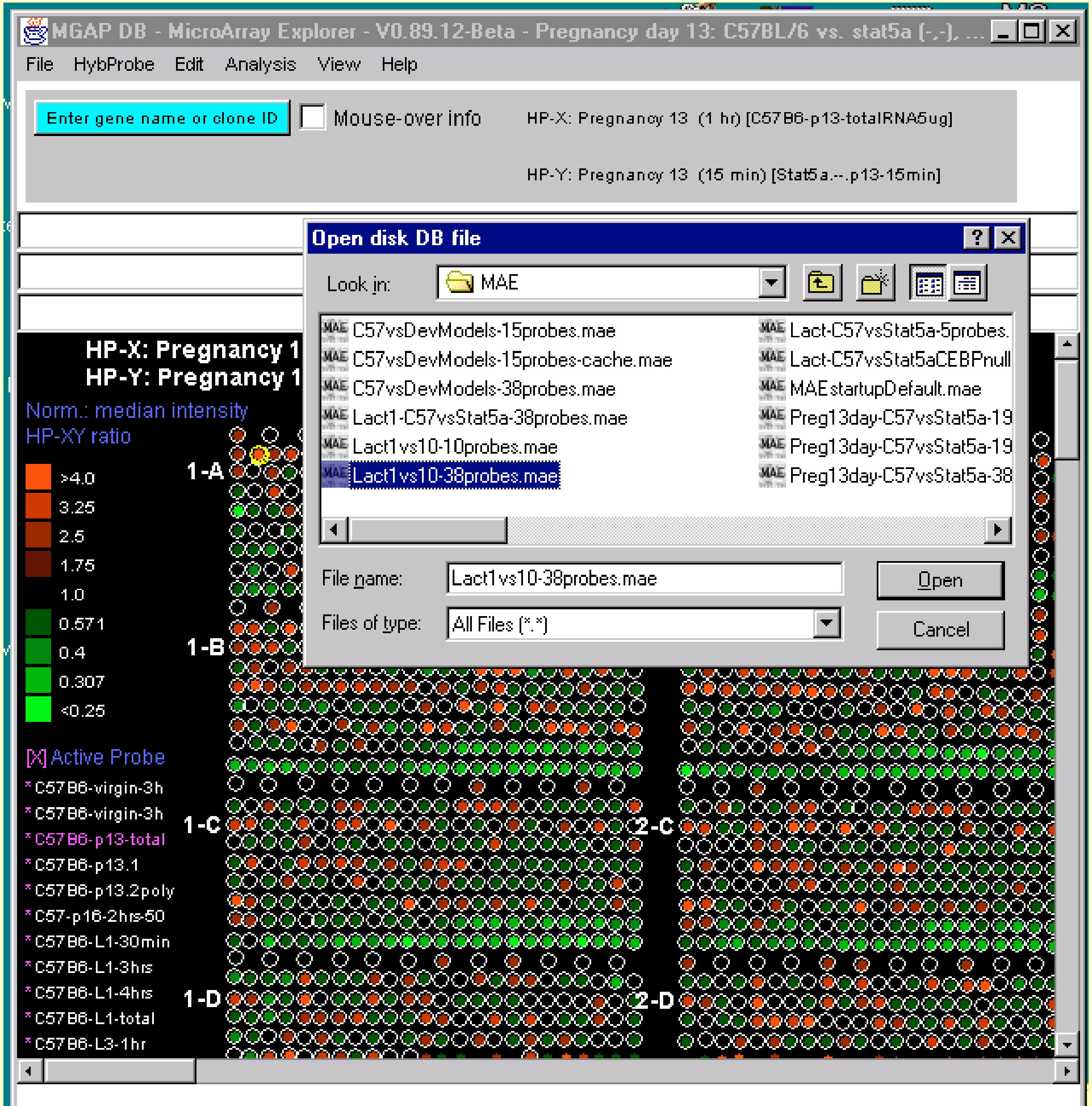
- The following examples demonstrate some of its capabilities

- Note: many more examples and discussion of the various analysis plots and reports may be found in the online reference manual at

  http://www.lecb.ncifcrf.gov/MAExplorer/hmaeHelp.html

II.5. Opening a database from local disk

- In stand-alone mode, you may browse a project database containing many startup databases.



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

☐ Mouse-over info

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

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

Open disk DB file

Look in: MAE

C57vsDevModels-15probes.mae
C57vsDevModels-15probes-cache.mae
C57vsDevModels-38probes.mae
Lact1-C57vsStat5a-38probes.mae
Lact1vs10-10probes.mae
Lact1vs10-38probes.mae
Lact-C57vsStat5a-5probes.
Lact-C57vsStat5aCEBPnull
MAEstartupDefault.mae
Preg13day-C57vsStat5a-19
Preg13day-C57vsStat5a-19
Preg13day-C57vsStat5a-38

File name: Lact1vs10-38probes.mae

Files of type: All Files (*.*)

HP-X: Pregnancy 1
HP-Y: Pregnancy 1

Norm.: median intensity
HP-XY ratio

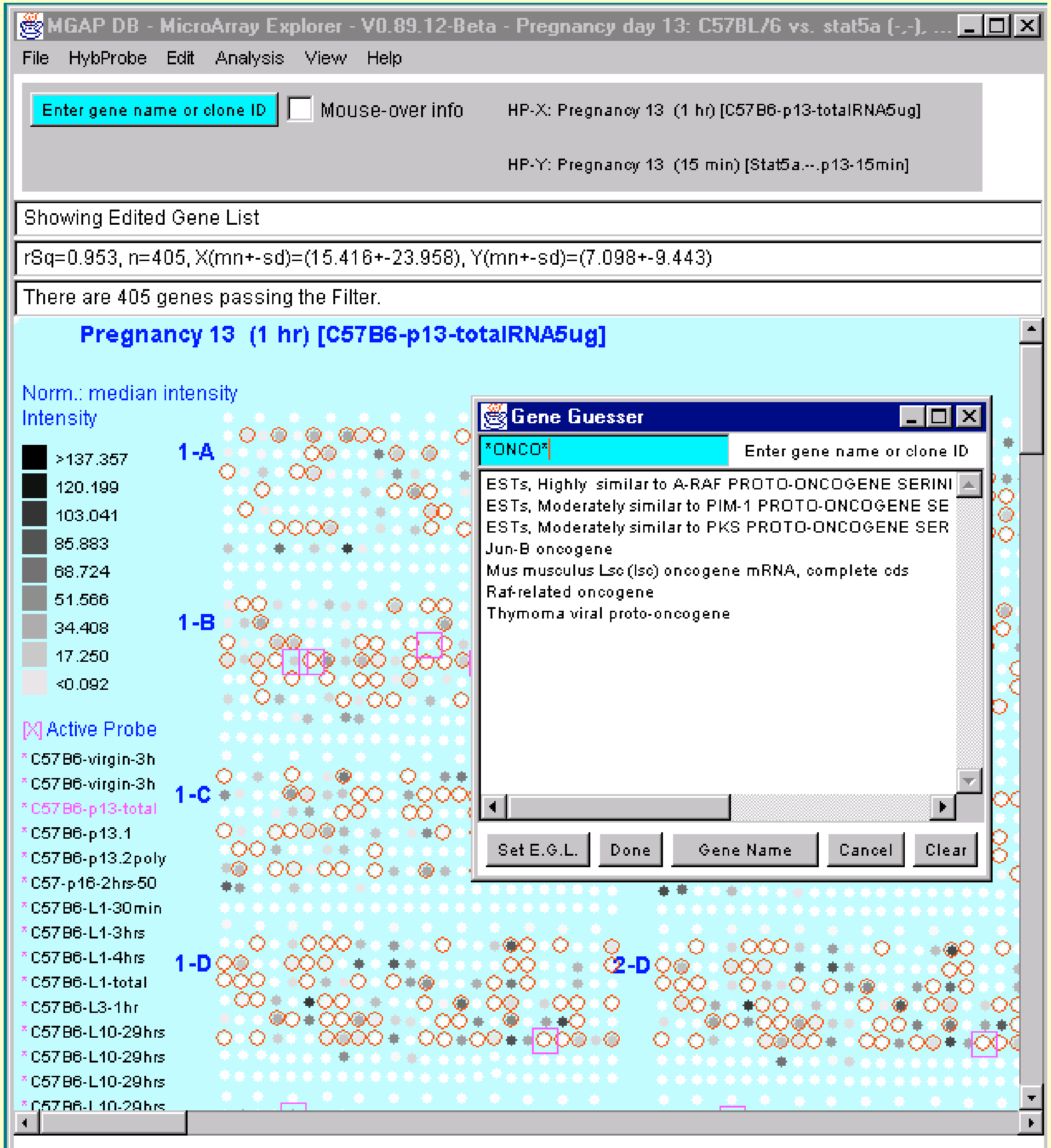
🟥 >4.0
🟥 3.25
🟥 2.5
🟥 1.75
1.0
🟩 0.571
🟩 0.4
🟩 0.307
🟩 <0.25

[X] Active Probe

*C57B6-virgin-3h
*C57B6-virgin-3h
*C57B6-p13-total
*C57B6-p13.1
*C57B6-p13.2poly
*C57-p16-2hrs-50
*C57B6-L1-30min
*C57B6-L1-3hrs
*C57B6-L1-4hrs
*C57B6-L1-total
*C57B6-L3-1hr

II.6 Specify Gene or Gene Subset by Name

- Specify gene or gene subset by gene name guesser using wildcard sub-strings eg. "*ONCO*" indicated by magenta boxes - saved in 'Edited Gene List'. [MGAP DB]



Showing Edited Gene List

rSq=0.953, n=405, X(mn+-sd)=(15.416+-23.958), Y(mn+-sd)=(7.098+-9.443)

There are 405 genes passing the Filter.

ONCO

ESTs, Highly similar to A-RAF PROTO-ONCOGENE SERINE
ESTs, Moderately similar to PIM-1 PROTO-ONCOGENE SE
ESTs, Moderately similar to PKS PROTO-ONCOGENE SER
Jun-B oncogene
Mus musculus Lsc (lsc) oncogene mRNA, complete cds
Raf-related oncogene
Thymoma viral proto-oncogene

MAExplorer User Interface

- The **MAExplorer menus** are similar to most Windows PC applications where pull-down menu selections are used to invoke operations.

- The current hybridization sample is displayed as a **pseudo image of spot intensity**.

- Names of the current HP-X and HP-Y **samples** are listed above the pseudo image.

- The "**Enter gene name or Clone ID**" button pops up a dialog box to assign the current gene (or set of genes) by name or wildcard.

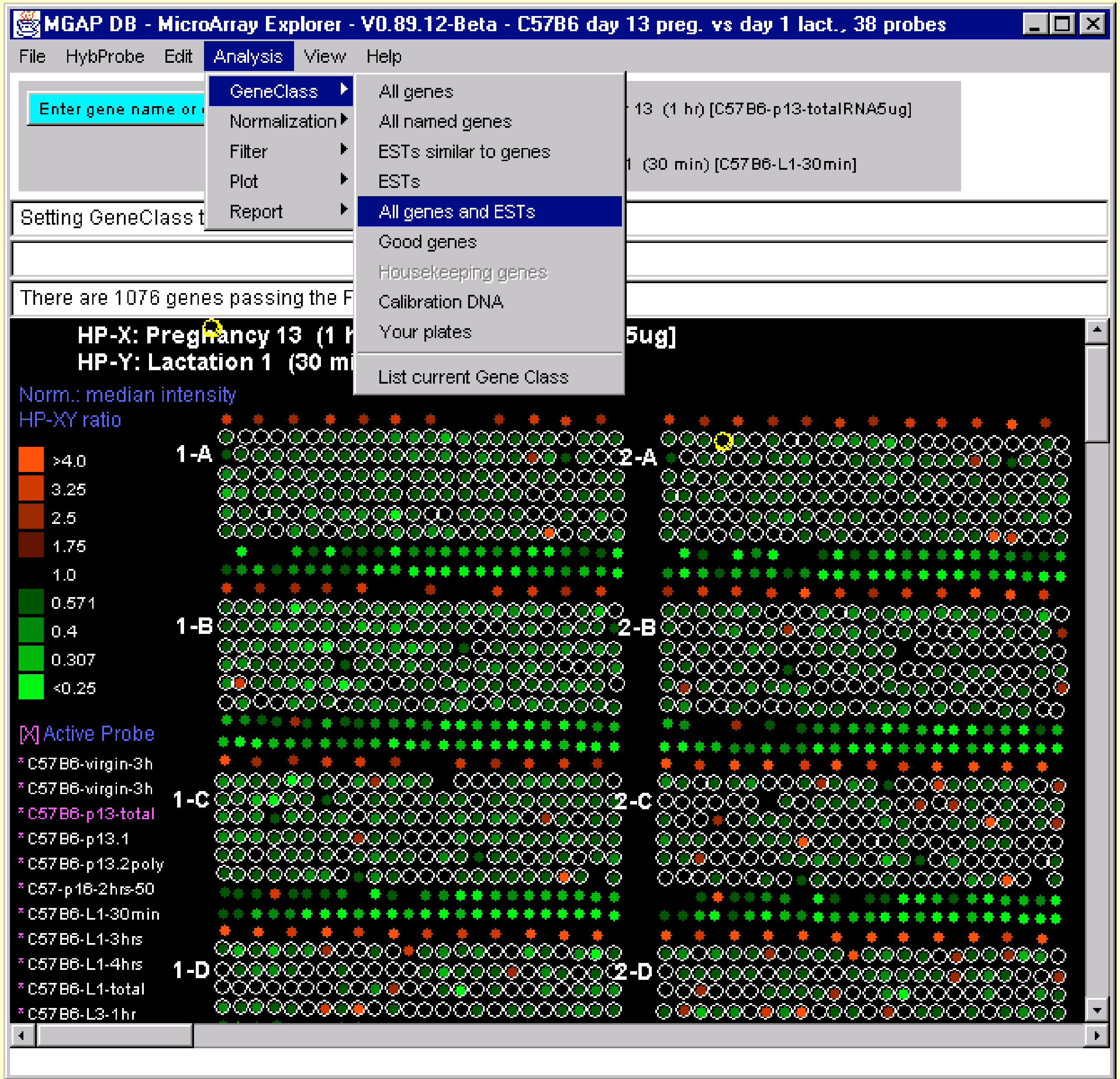
- **Clicking** on spots, points in plots or cells in spreadsheet reports **assigns the current gene**, displays information on it, and accesses Web genomic databases.

- The MGAP microarrays (shown here) contain 1,700 duplicated ^{33}P-labeled clones indicated as fields 1 and 2 in the array pseudo image.

- Duplicated grids of cDNA spots are labeled as 1-A, 2-A, 1-B, 2-B, etc.

II.7a Named Genes and ESTs

- Specify sets of genes for *all named genes* and *all ESTs* indicated in the microarray by white circles. [MGAP data]



MGAP DB - MicroArray Explorer - V0.89.12-Beta - C57B6 day 13 preg. vs day 1 lact., 38 probes

Enter gene name or

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

1 (30 min) [C57B6-L1-30min]

Setting GeneClass t

There are 1076 genes passing the F

HP-X: Pregnancy 13 (1 h 5ug]

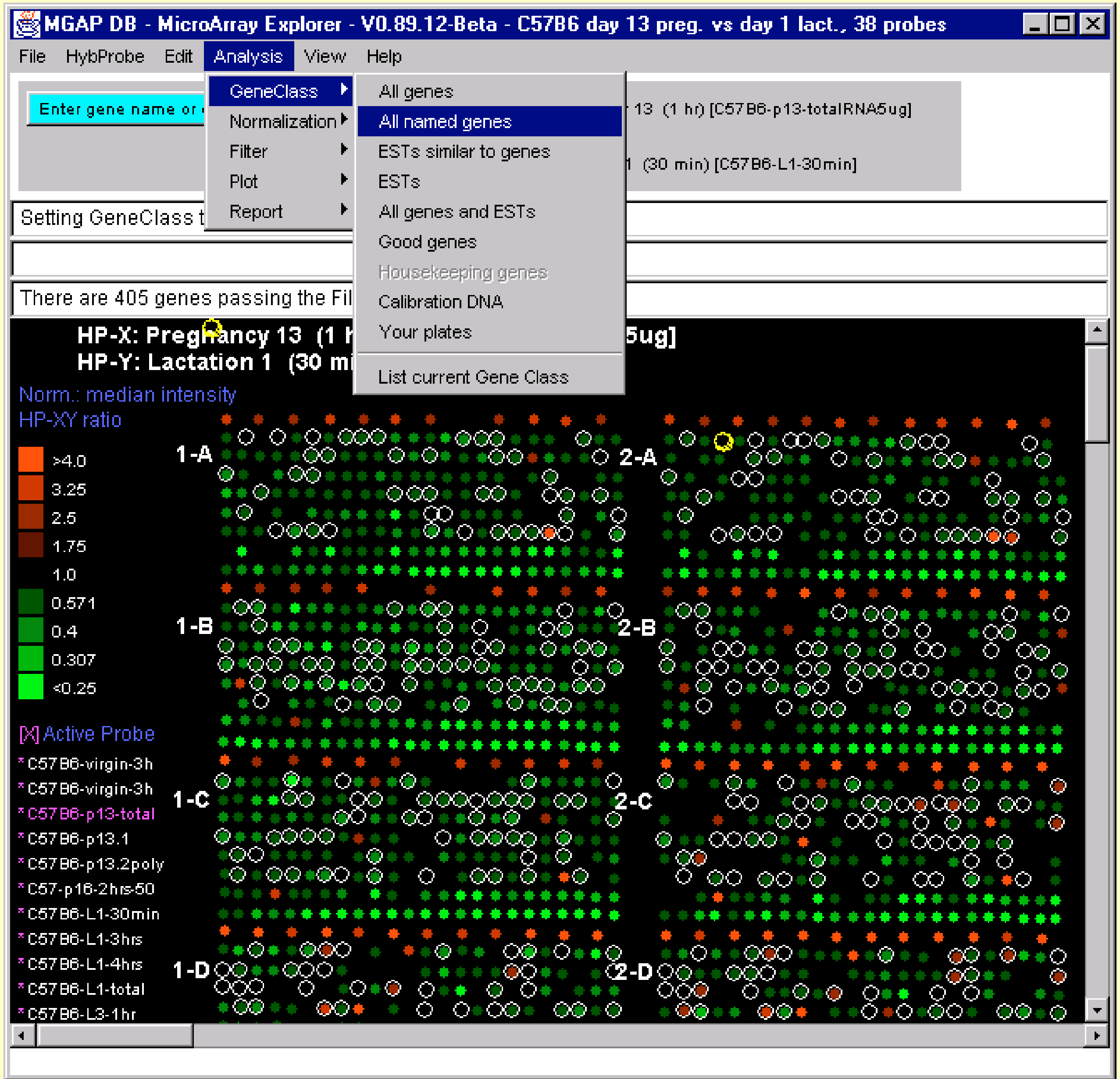
HP-Y: Lactation 1 (30 mi

Norm.: median intensity

HP-XY ratio

II.7b Named Genes

- Specify sets of genes for *all named genes* indicated in ratio X/Y array plot by white circles



MGAP DB - MicroArray Explorer - V0.89.12-Beta - C57B6 day 13 preg. vs day 1 lact., 38 probes

Setting GeneClass t

There are 405 genes passing the Fil

HP-X: Pregnancy 13 (1 h 5ug]

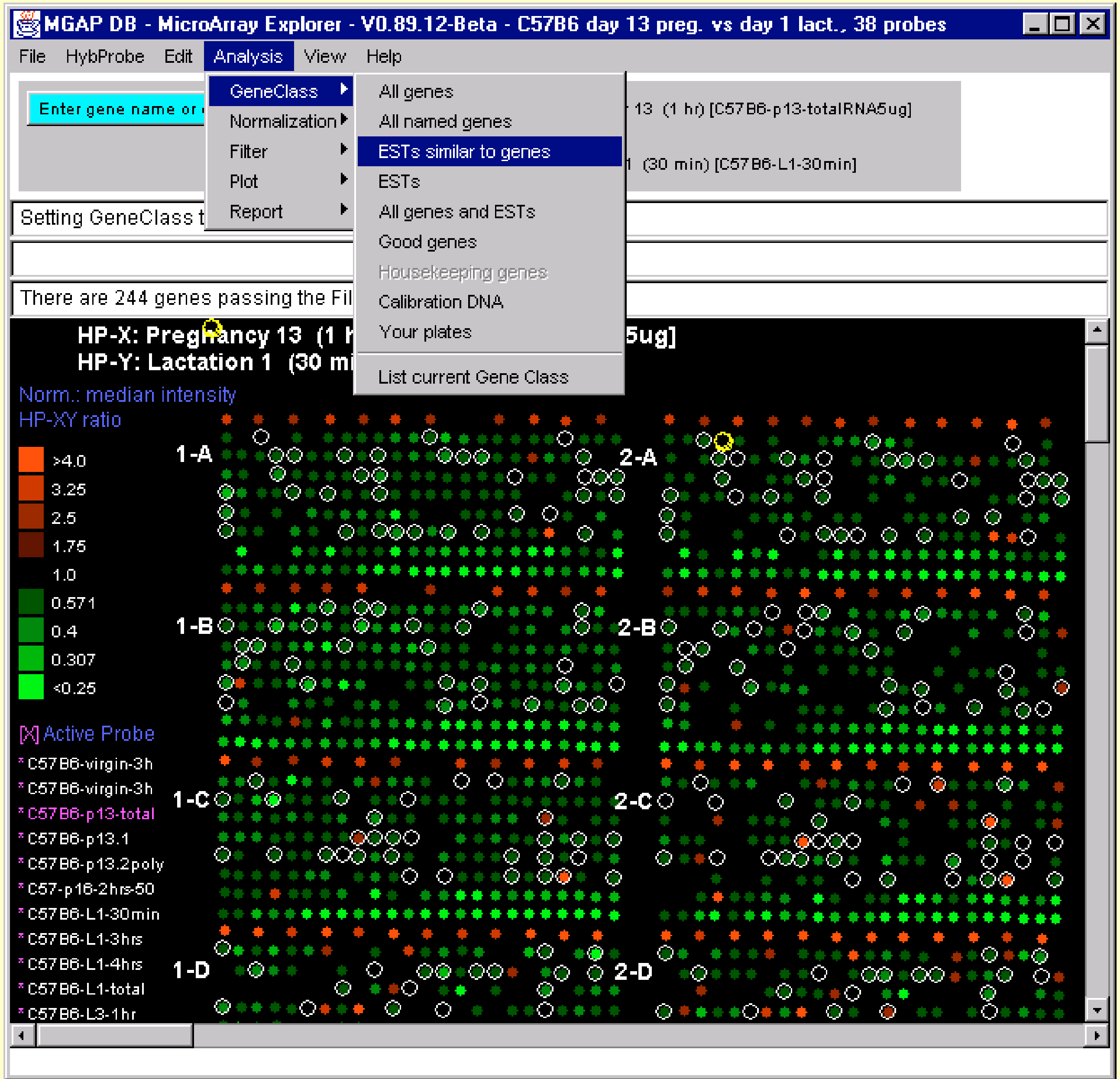
HP-Y: Lactation 1 (30 mi

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

1 (30 min) [C57B6-L1-30min]

II.7c ESTs similar to named genes

- Specify sets of genes for *all ESTs similar to named genes* indicated in the microarray by white circles



Setting GeneClass t

There are 244 genes passing the Fil

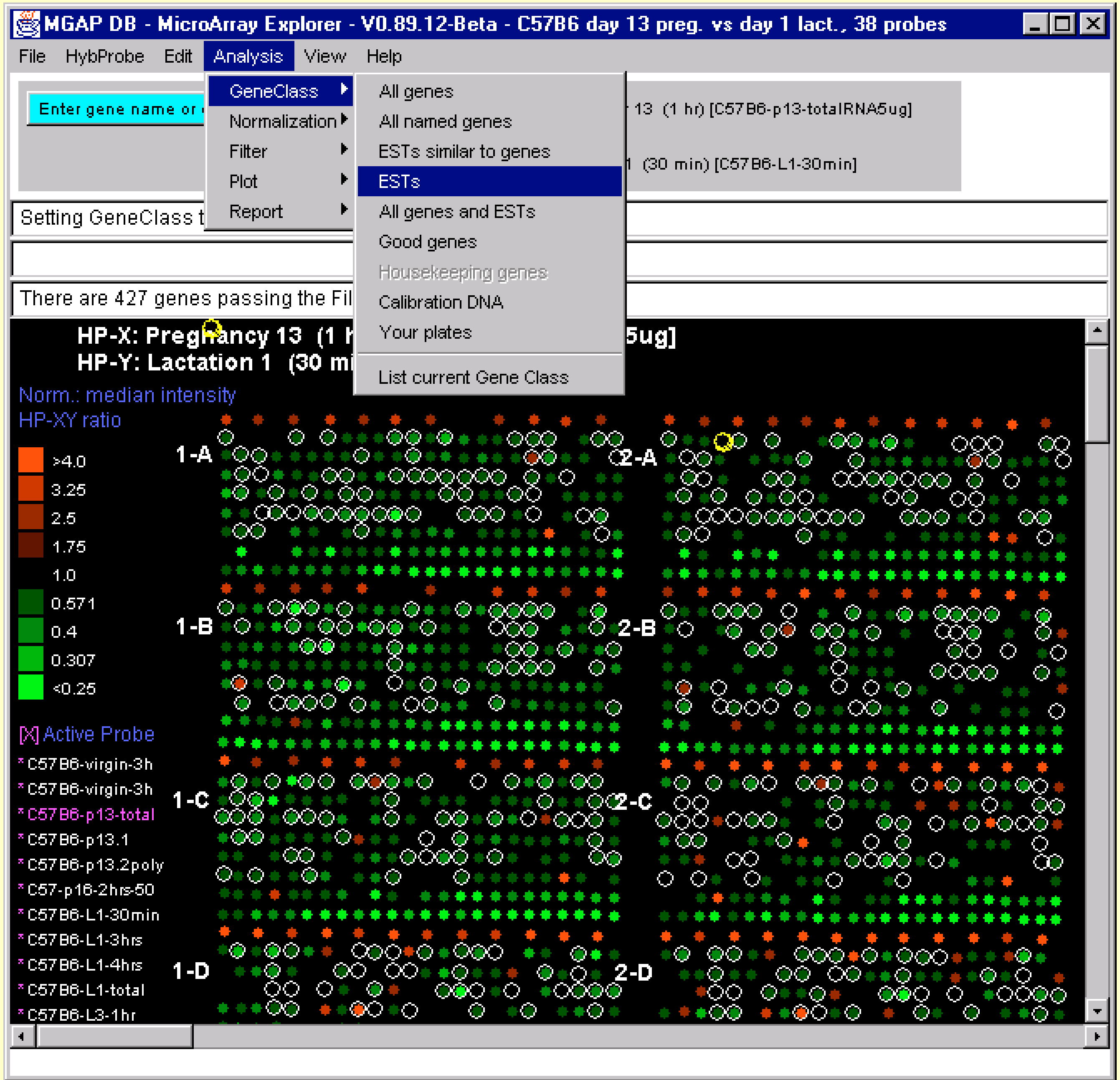
HP-X: Pregnancy 13 (1 h

5ug]

HP-Y: Lactation 1 (30 mi

II.7d Unknown ESTs

- Specify sets of genes for *unknown ESTs* indicated in the microarray by white circles



MGAP DB - MicroArray Explorer - V0.89.12-Beta - C57B6 day 13 preg. vs day 1 lact., 38 probes

Enter gene name or

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

1 (30 min) [C57B6-L1-30min]

Setting GeneClass t

There are 427 genes passing the Fil

HP-X: Pregnancy 13 (1 h 5ug]

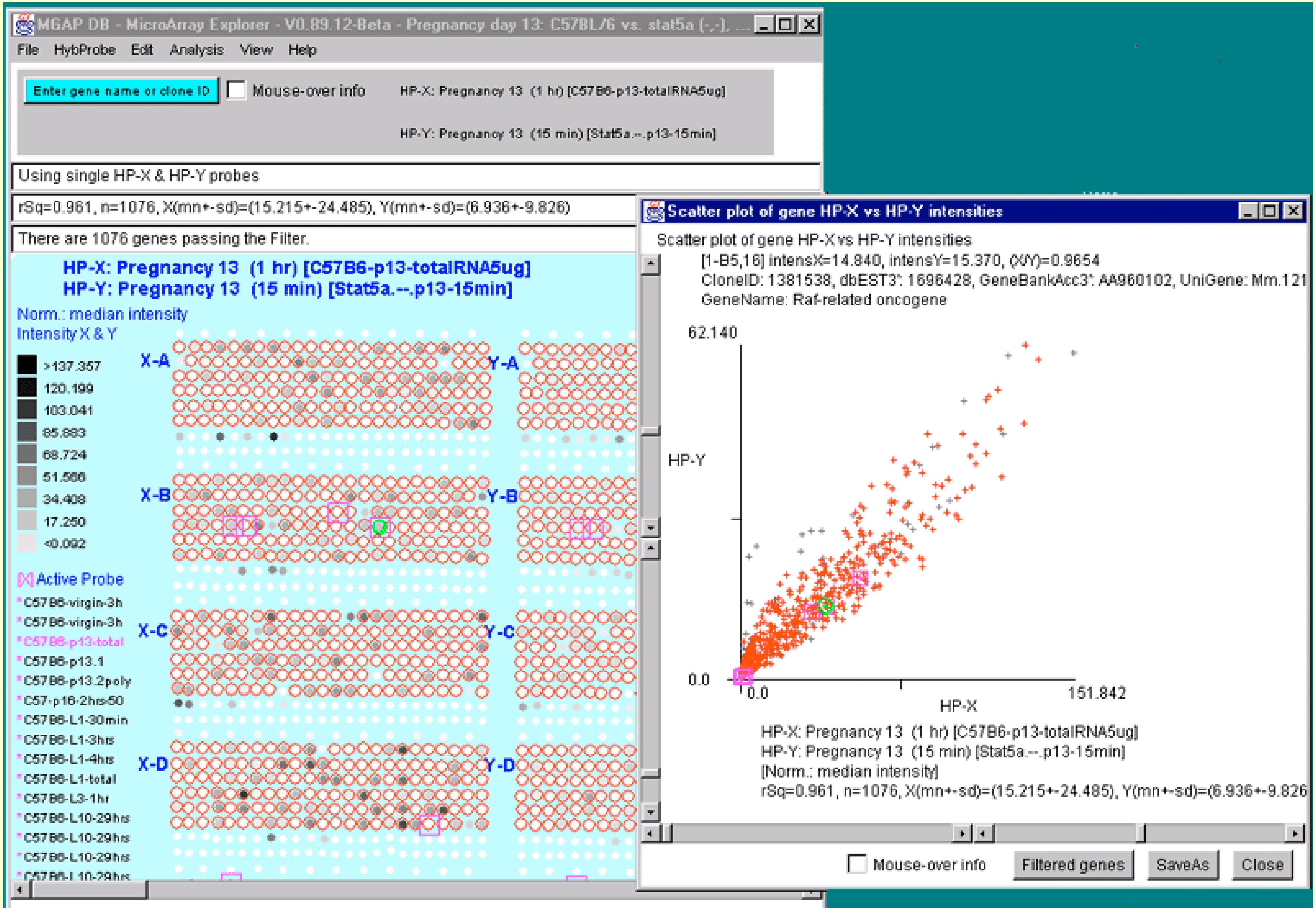
HP-Y: Lactation 1 (30 mi

Norm.: median intensity

HP-XY ratio

II.8a Scatter Plots of Two Conditions

- X-Y scatter plot of ‘sets’ of 2-probes C57B6 vs Stat5a (-,-) 13-day pregnancy in array [MGAP]. Current gene (green circle) & Edited Gene List (magenta squares) in plot



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

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

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

Using single HP-X & HP-Y probes

rSq=0.961, n=1076, X(mn+-sd)=(15.215+-24.485), Y(mn+-sd)=(6.936+-9.826)

There are 1076 genes passing the Filter.

Norm.: median intensity

Scatter plot of gene HP-X vs HP-Y intensities

[1-B5,16] intensX=14.840, intensY=15.370, (X/Y)=0.9654

CloneID: 1381538, dbEST3': 1696428, GeneBankAcc3': AA960102, UniGene: Mm.121

GeneName: Raf-related oncogene

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

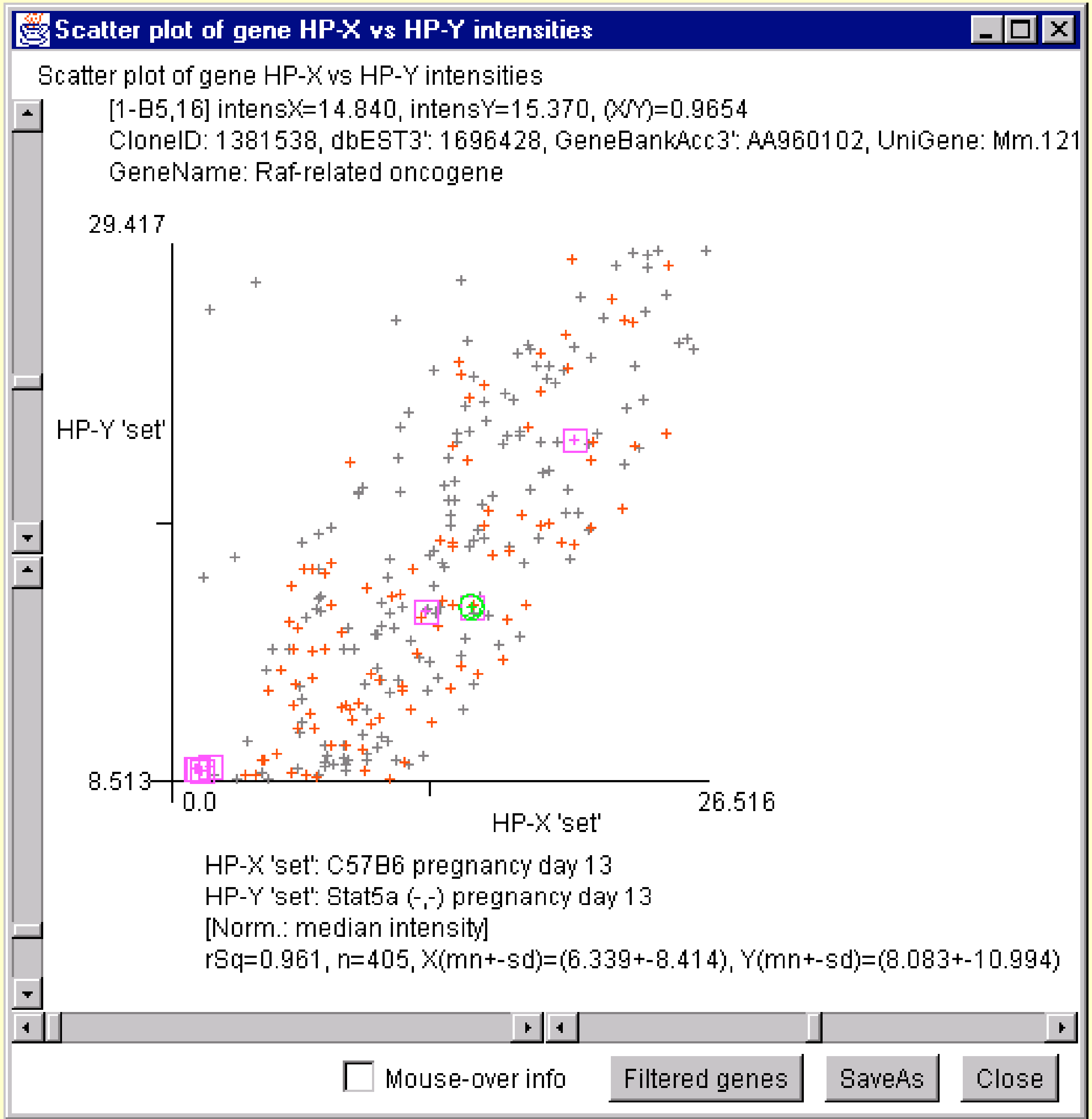
HP-Y: Pregnancy 13 (15 min) [Stat5a.--.p13-15min]

[Norm.: median intensity]

rSq=0.961, n=1076, X(mn+-sd)=(15.215+-24.485), Y(mn+-sd)=(6.936+-9.826

II.8b Zoomed X-Y Scatter Plot (of II.8a)

- Zoomed in on Raf-related oncogene using scrollbars. Genes not passing Filter are grayed out in the plot



Scatter plot of gene HP-X vs HP-Y intensities

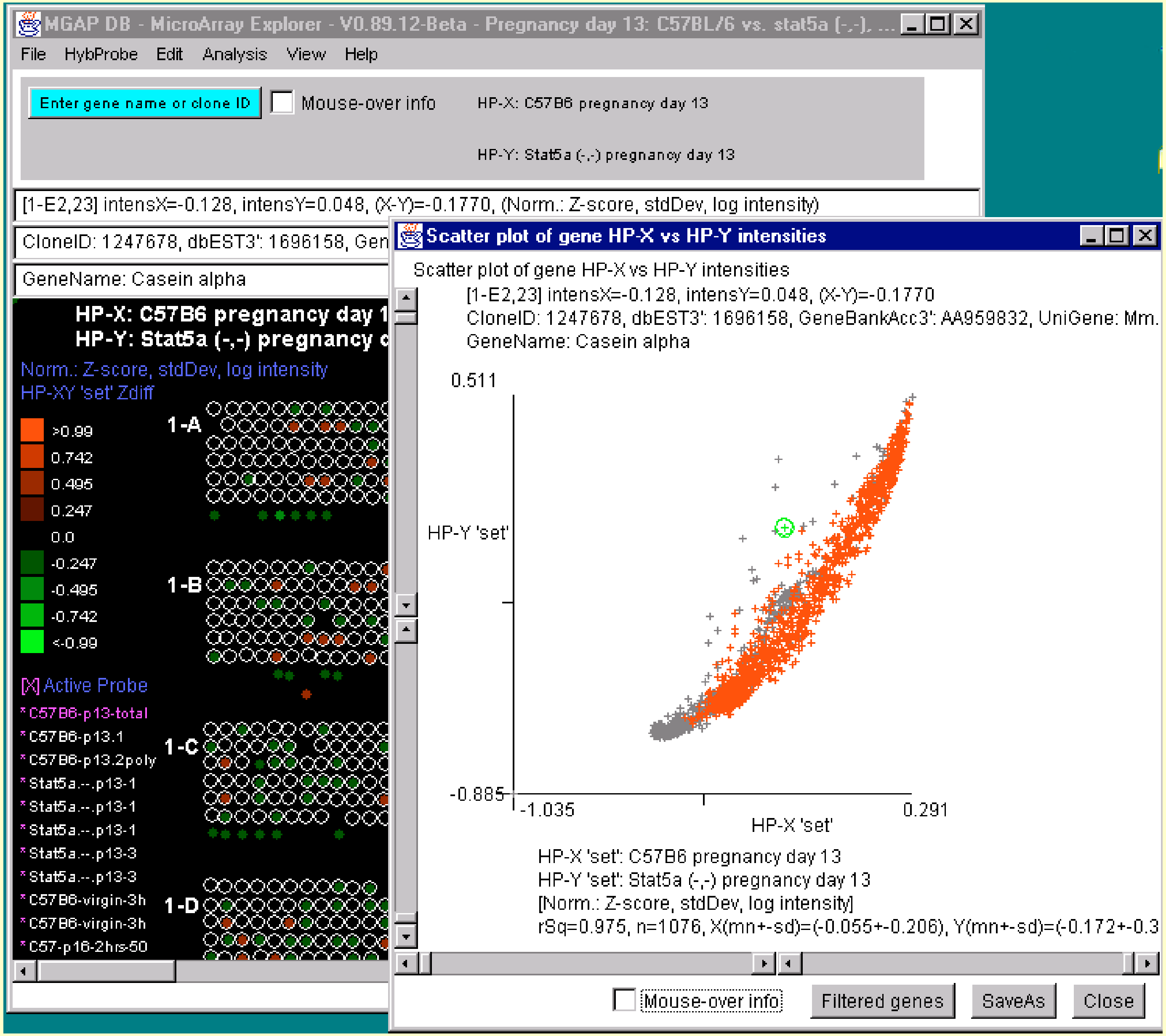
Scatter plot of gene HP-X vs HP-Y intensities

[1-B5,16] intensX=14.840, intensY=15.370, (X/Y)=0.9654
CloneID: 1381538, dbEST3': 1696428, GeneBankAcc3': AA960102, UniGene: Mm.121
GeneName: Raf-related oncogene

HP-X 'set': C57B6 pregnancy day 13
HP-Y 'set': Stat5a (-,-) pregnancy day 13
[Norm.: median intensity]
rSq=0.961, n=405, X(mn+-sd)=(6.339+-8.414), Y(mn+-sd)=(8.083+-10.994)

II.9a Genes Filtered by Gene Class Set

- Genes class subset *named genes and ESTs* in both array & scatter plot normalized by Zscore of log intensity.



[1-E2,23] intensX=-0.128, intensY=0.048, (X-Y)=-0.1770, (Norm.: Z-score, stdDev, log intensity)

CloneID: 1247678, dbEST3': 1696158, Gen

GeneName: Casein alpha

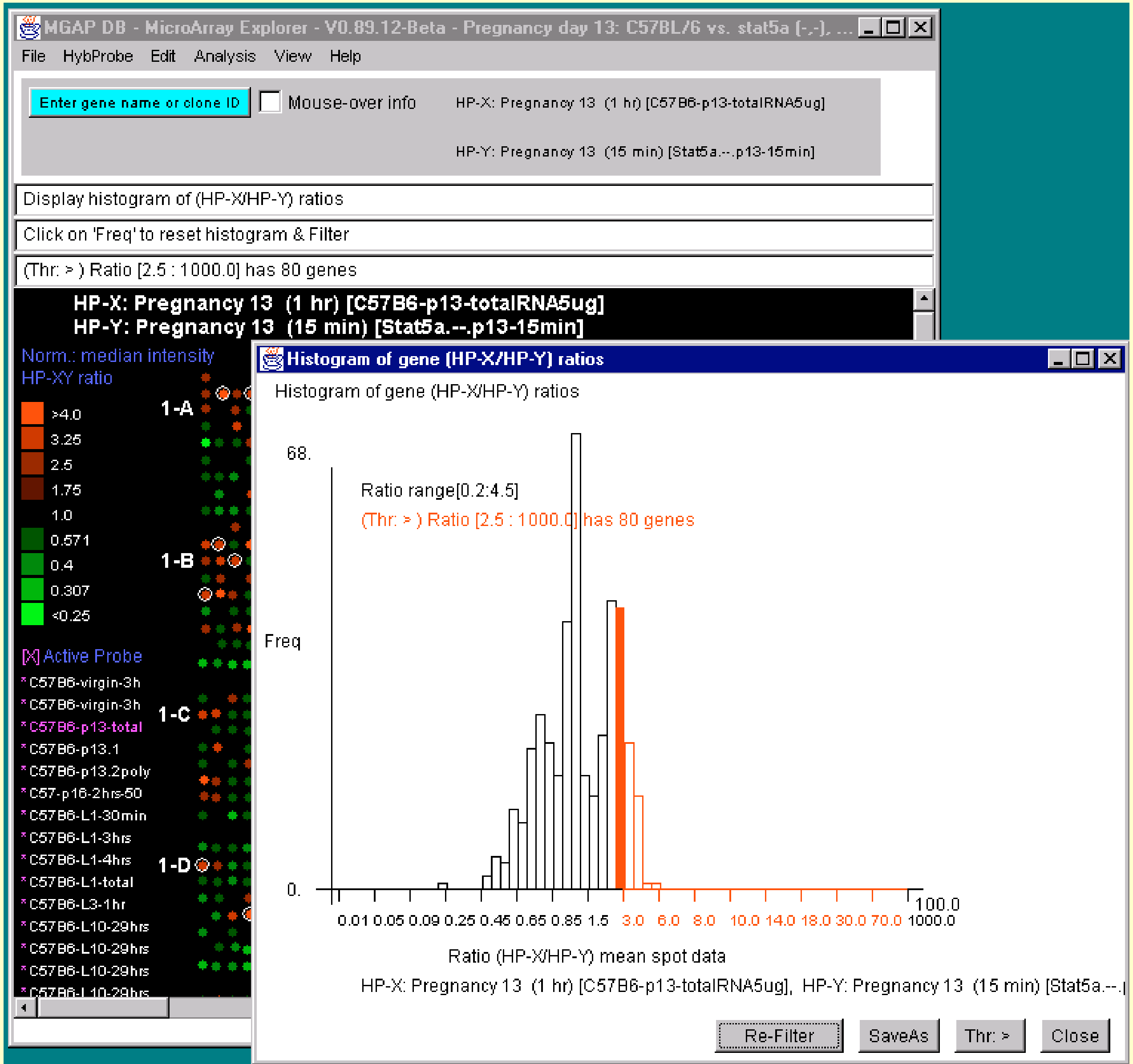
Scatter plot of gene HP-X vs HP-Y intensities

[1-E2,23] intensX=-0.128, intensY=0.048, (X-Y)=-0.1770
CloneID: 1247678, dbEST3': 1696158, GeneBankAcc3': AA959832, UniGene: Mm.
GeneName: Casein alpha

HP-X 'set': C57B6 pregnancy day 13
HP-Y 'set': Stat5a (-,-) pregnancy day 13
[Norm.: Z-score, stdDev, log intensity]
rSq=0.975, n=1076, X(mn+-sd)=(-0.055+-0.206), Y(mn+-sd)=(-0.172+-0.3

II.9b Genes Filtered by Ratio-Histogram Bin

- Genes filtered by HP-X/HP-Y C57B5-preg / Stat5a(-,-) ratio-histogram bin-range [2.5:1000]. Histogram is for all named genes and for ESTs.



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

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

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

Display histogram of (HP-X/HP-Y) ratios

Click on 'Freq' to reset histogram & Filter

(Thr: >) Ratio [2.5 : 1000.0] has 80 genes

HP-X: Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug]
HP-Y: Pregnancy 13 (15 min) [Stat5a.--.p13-15min]

Norm.: median intensity
HP-XY ratio

[X] Active Probe

Histogram of gene (HP-X/HP-Y) ratios

Ratio range[0.2:4.5]

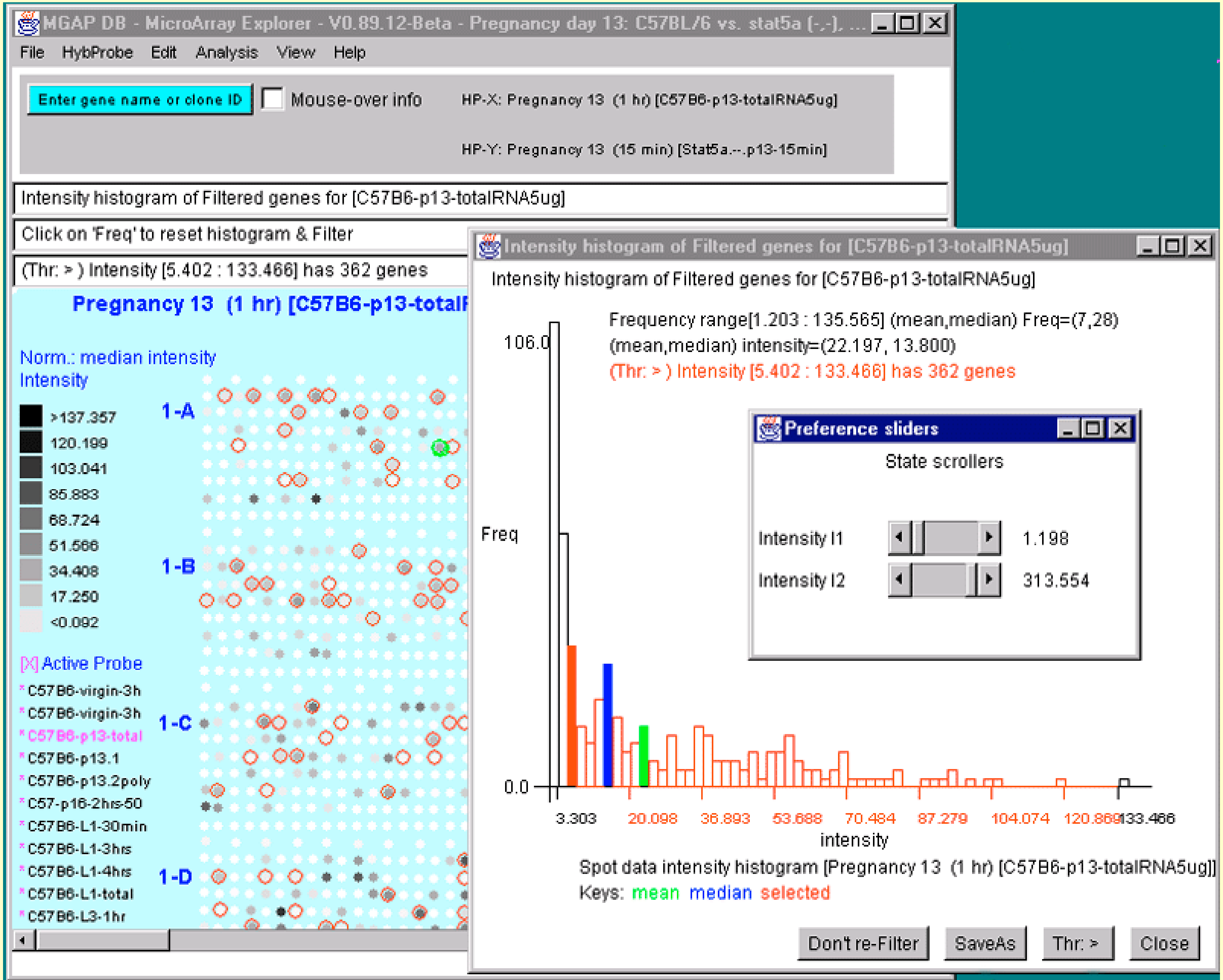
(Thr: >) Ratio [2.5 : 1000.0] has 80 genes

Ratio (HP-X/HP-Y) mean spot data

HP-X: Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug], HP-Y: Pregnancy 13 (15 min) [Stat5a.--.

II.9c Genes Filtered by Intensity-Histogram Bin

- Genes filtered by intensity to remove low signal strength sample genes.



MGAP DB - MicroArray Explorer - V0.89.12-Beta - Pregnancy day 13: C57BL/6 vs. stat5a [-,-], ...

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

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

Intensity histogram of Filtered genes for [C57B6-p13-totalRNA5ug]

Click on 'Freq' to reset histogram & Filter

(Thr: >) Intensity [5.402 : 133.466] has 362 genes

Pregnancy 13 (1 hr) [C57B6-p13-totalF

Norm.: median intensity

Intensity

■	>137.357
■	120.199
■	103.041
■	85.883
■	68.724
■	51.566
■	34.408
■	17.250
□	<0.092

[X] Active Probe

- * C57B6-virgin-3h
- * C57B6-virgin-3h
- * C57B6-p13-total
- * C57B6-p13.1
- * C57B6-p13.2poly
- * C57-p16-2hrs-50
- * C57B6-L1-30min
- * C57B6-L1-3hrs
- * C57B6-L1-4hrs
- * C57B6-L1-total
- * C57B6-L3-1hr

Intensity histogram of Filtered genes for [C57B6-p13-totalRNA5ug]

Intensity histogram of Filtered genes for [C57B6-p13-totalRNA5ug]

Frequency range[1.203 : 135.565] (mean,median) Freq=(7,28)
(mean,median) intensity=(22.197, 13.800)
(Thr: >) Intensity [5.402 : 133.466] has 362 genes

Preference sliders

State scrollers

Intensity I1 1.198

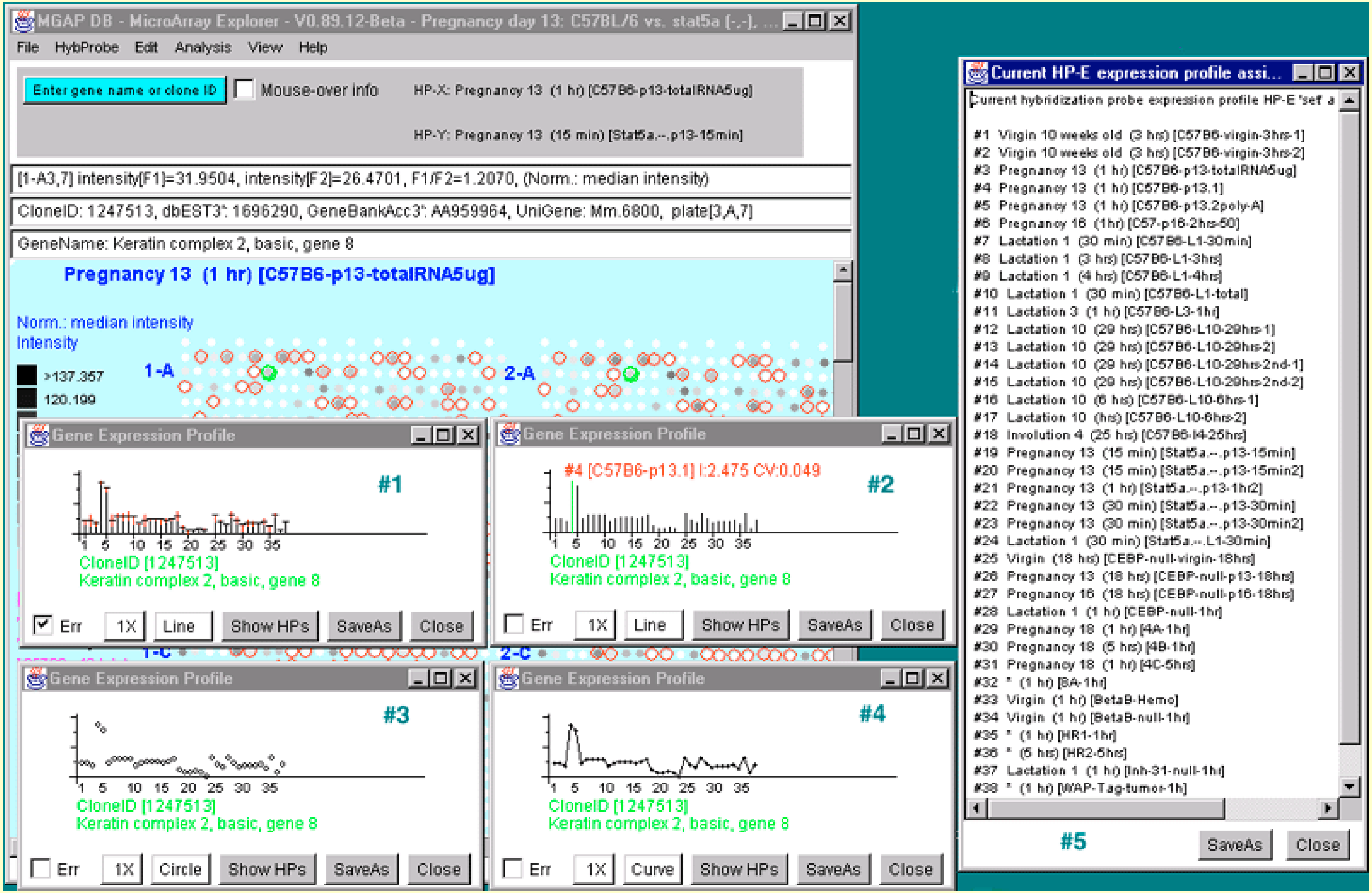
Intensity I2 313.554

Spot data intensity histogram [Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug]]

Keys: mean median selected

II.10a Expression Profile Plots of N-conditions

- Expression profile plot of 38-conditions of current gene (green) . Note numbered list of probes. Intensity data for probe #4 is indicated in red - by clicking on a line in plot



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

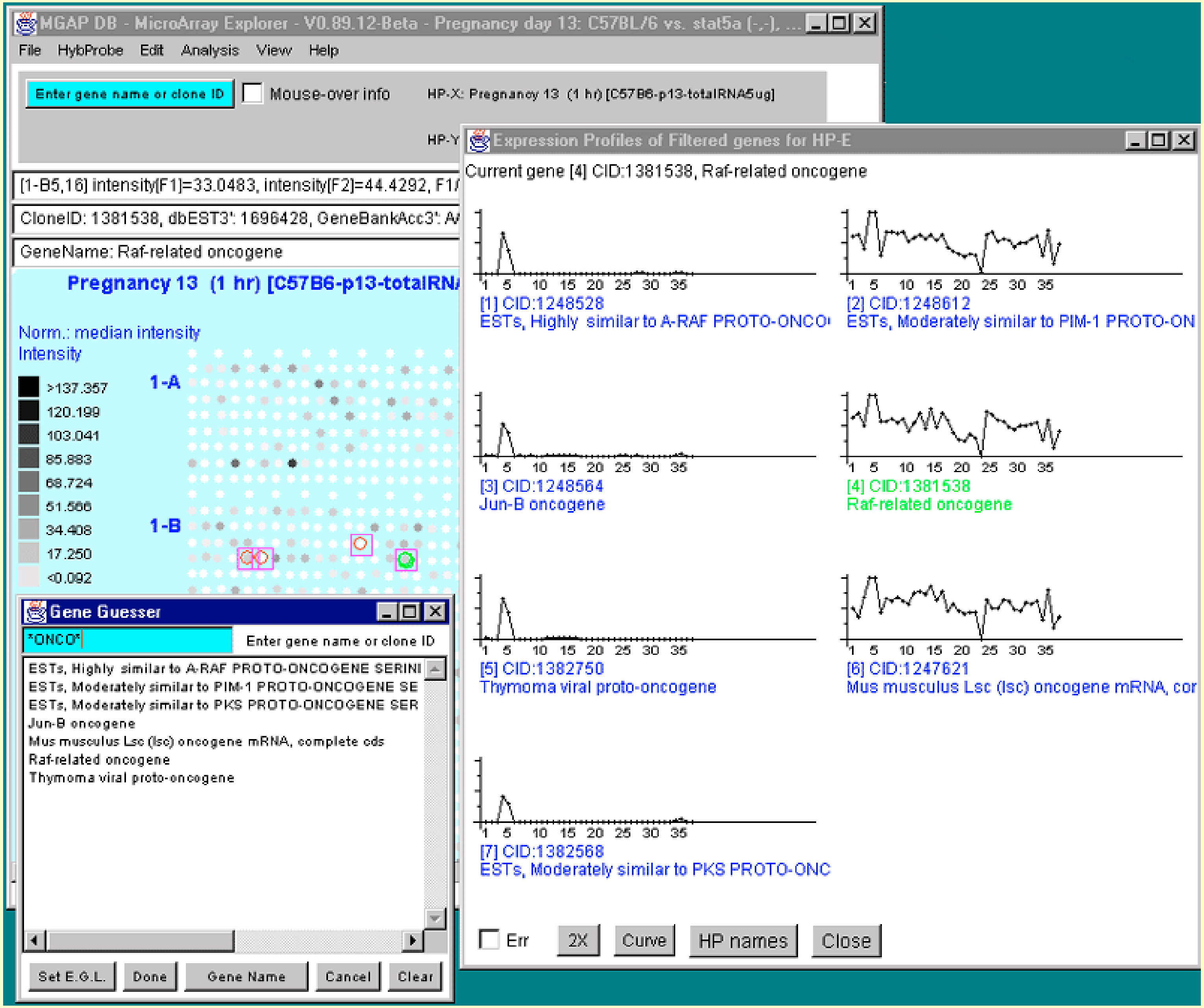
CloneID [1247513]
Keratin complex 2, basic, gene 8

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 16 (1hr) [C57-p16-2hrs-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 (29 hrs) [C57B6-L10-29hrs-1]
#13 Lactation 10 (29 hrs) [C57B6-L10-29hrs-2]
#14 Lactation 10 (29 hrs) [C57B6-L10-29hrs-2nd-1]
#15 Lactation 10 (29 hrs) [C57B6-L10-29hrs-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 16 (18 hrs) [CEBP-null-p16-18hrs]
#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-1h]

II.10b List of Expression Profile Plots

- Scrollable list of EP plots for onco and proto-oncogenes in EGL for MGAP database



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

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

[1-B5,16] intensity[F1]=33.0483, intensity[F2]=44.4292, F1/

CloneID: 1381538, dbEST3': 1696428, GeneBankAcc3': A

GeneName: Raf-related oncogene

Expression Profiles of Filtered genes for HP-E

Current gene [4] CID:1381538, Raf-related oncogene

[1] CID:1248528
ESTs, Highly similar to A-RAF PROTO-ONCO

[2] CID:1248612
ESTs, Moderately similar to PIM-1 PROTO-ON

[3] CID:1248564
Jun-B oncogene

[4] CID:1381538
Raf-related oncogene

[5] CID:1382750
Thymoma viral proto-oncogene

[6] CID:1247621
Mus musculus Lsc (lsc) oncogene mRNA, cor

[7] CID:1382568
ESTs, Moderately similar to PKS PROTO-ONC

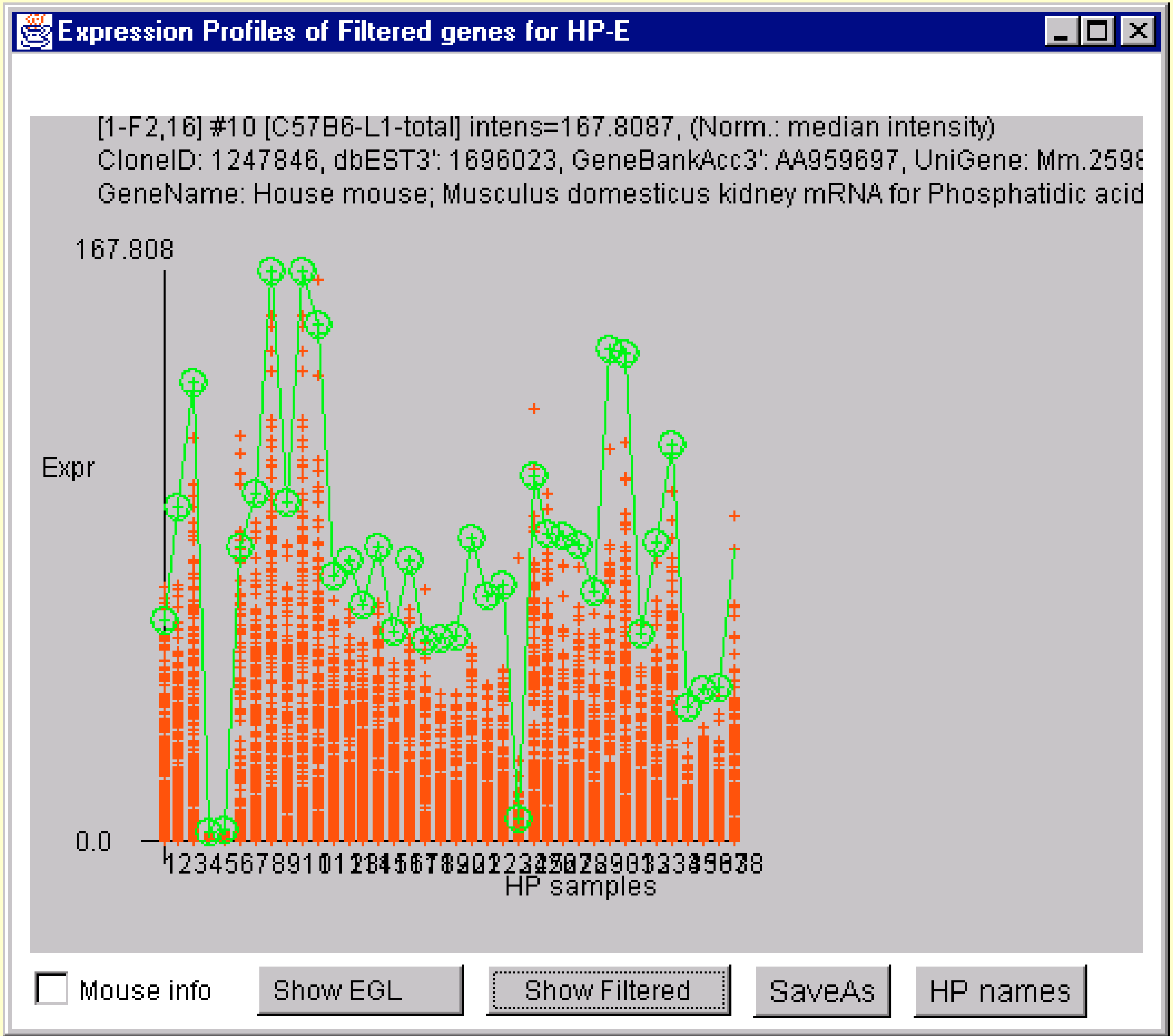
Gene Guesser

ONCO

ESTs, Highly similar to A-RAF PROTO-ONCOGENE SERINI
ESTs, Moderately similar to PIM-1 PROTO-ONCOGENE SE
ESTs, Moderately similar to PKS PROTO-ONCOGENE SER
Jun-B oncogene
Mus musculus Lsc (lsc) oncogene mRNA, complete cds
Raf-related oncogene
Thymoma viral proto-oncogene

II.10.c Expression Profile Overlay Plots

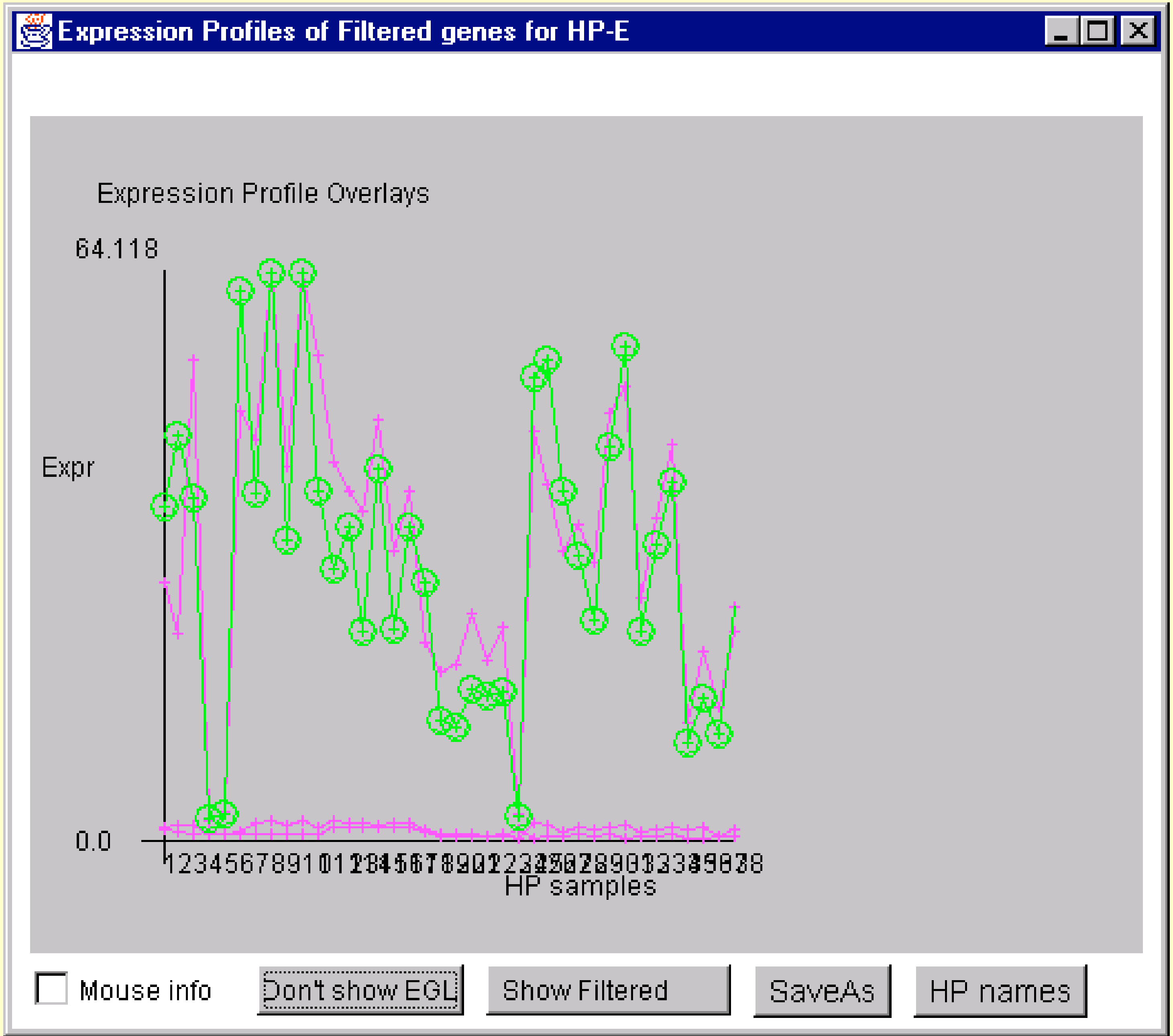
- Overlay EP plots of multiple genes showing current gene for MGAP database



[1-F2,16] #10 [C57B6-L1-total] intens=167.8087, (Norm.: median intensity)
CloneID: 1247846, dbEST3': 1696023, GeneBankAcc3': AA959697, UniGene: Mm.259
GeneName: House mouse; Musculus domesticus kidney mRNA for Phosphatidic acid

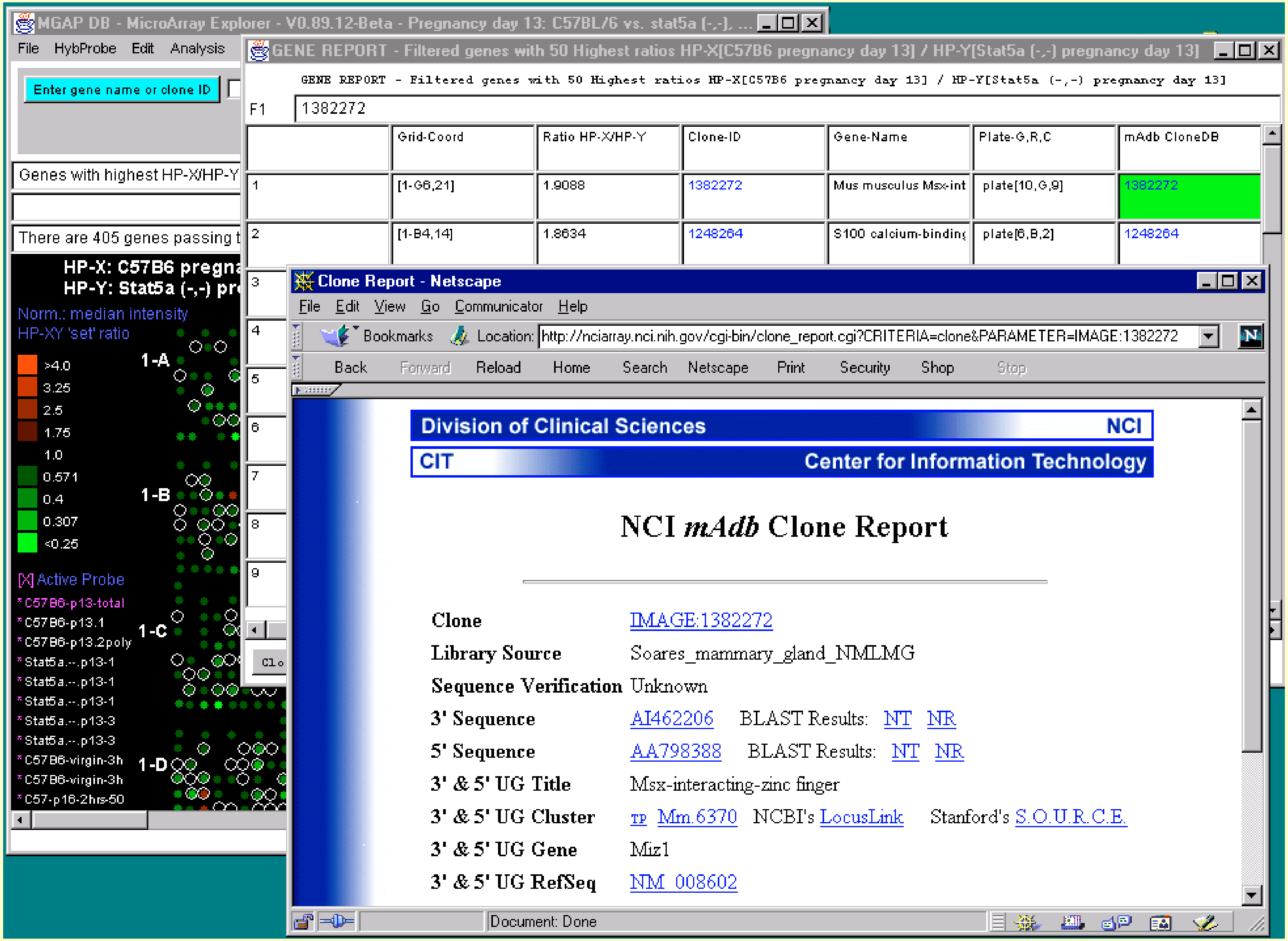
II.10.d Expression Profile Overlay Plots

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



II.11a Scrollable Dynamic Gene Reports

- Scrollable gene report of highest ratio genes & NCI mAdb pop up Web browser page (foreground) of particular gene. Clicking on blue hypertext cell in gene report (middle) invokes pop up web page (NCI mAdb Clone Report shown here)



GENE REPORT - Filtered genes with 50 Highest ratios HP-X[C57B6 pregnancy day 13] / HP-Y[Stat5a (-,-) pregnancy day 13]

F1 1382272

	Grid-Coord	Ratio HP-X/HP-Y	Clone-ID	Gene-Name	Plate-G,R,C	mAdb CloneDB
1	[1-G6,21]	1.9088	1382272	Mus musculus Msx-int	plate[10,G,9]	1382272
2	[1-B4,14]	1.8634	1248264	S100 calcium-binding	plate[6,B,2]	1248264
3						
4						
5						
6						
7						
8						
9						

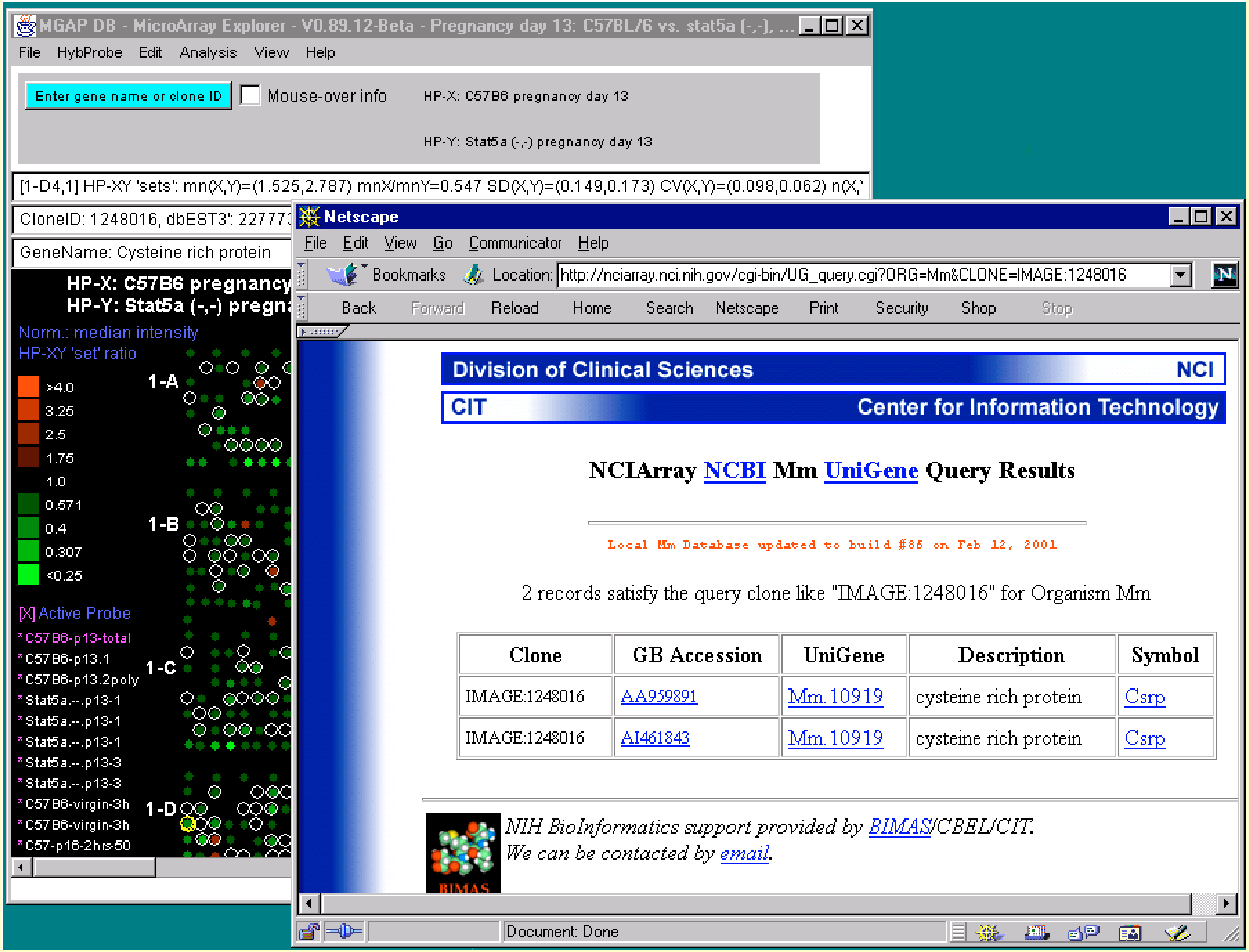
Location: http://nciarray.nci.nih.gov/cgi-bin/clone_report.cgi?CRITERIA=clone&PARAMETER=IMAGE:1382272

NCI *mAdb* Clone Report

Clone	IMAGE:1382272
Library Source	Soares_mammary_gland_NMLMG
Sequence Verification	Unknown
3' Sequence	AI462206 BLAST Results: NT NR
5' Sequence	AA798388 BLAST Results: NT NR
3' & 5' UG Title	Msx-interacting-zinc finger
3' & 5' UG Cluster	TP Mm.6370 NCBI's LocusLink Stanford's S.O.U.R.C.E.
3' & 5' UG Gene	Miz1
3' & 5' UG RefSeq	NM_008602

Document: Done

II.11a.1 Scrollable Dynamic Gene Reports - UniGene Report



[1-D4,1] HP-XY 'sets': mn(X,Y)=(1.525,2.787) mnX/mnY=0.547 SD(X,Y)=(0.149,0.173) CV(X,Y)=(0.098,0.062) n(X,'

CloneID: 1248016, dbEST3': 22777

GeneName: Cysteine rich protein

Location: http://nciarray.nci.nih.gov/cgi-bin/UG_query.cgi?ORG=Mm&CLONE=IMAGE:1248016

NCIArray NCBI Mm UniGene Query Results

Local Mm Database updated to build #86 on Feb 12, 2001

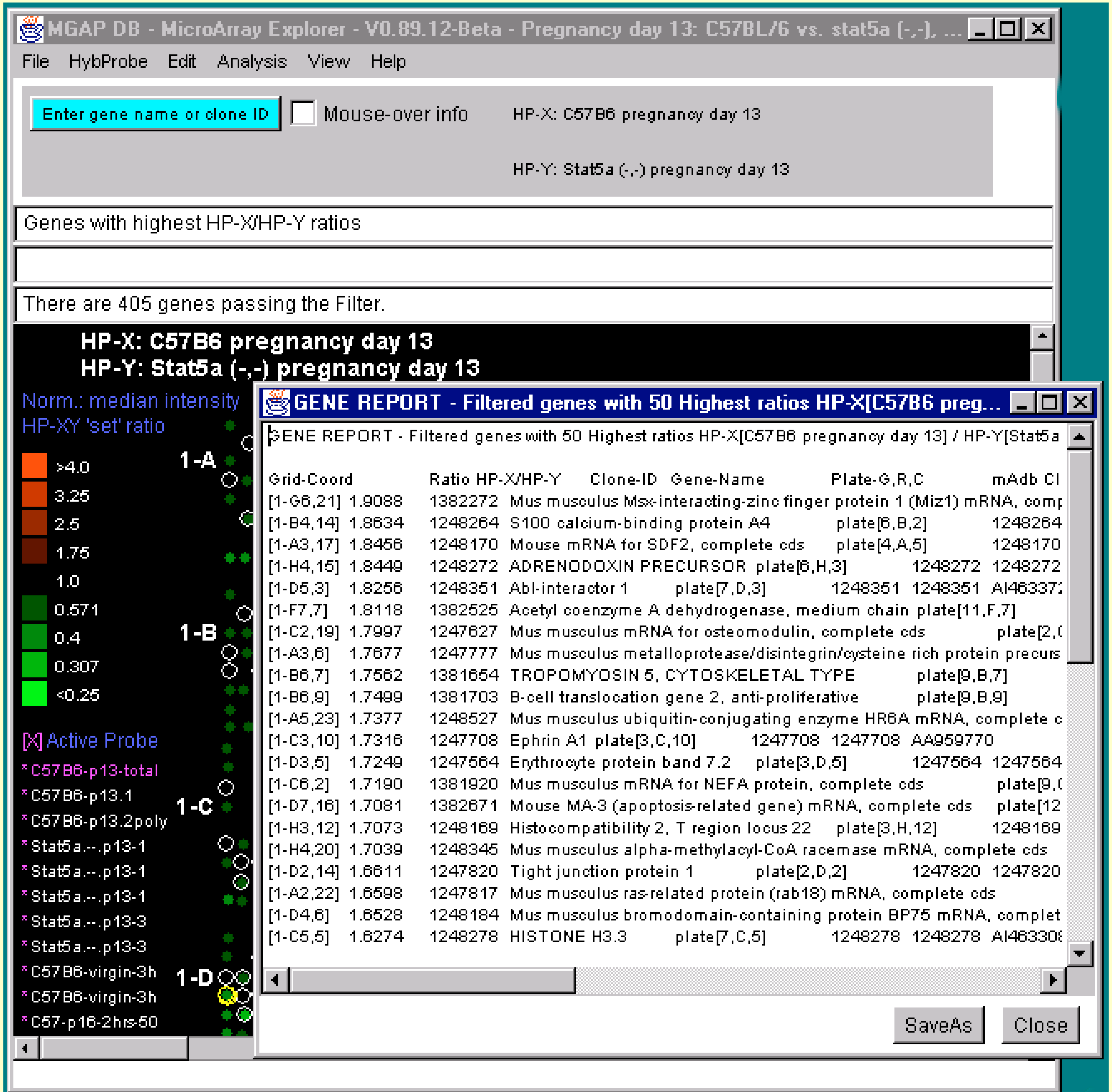
2 records satisfy the query clone like "IMAGE:1248016" for Organism Mm

Clone	GB Accession	UniGene	Description	Symbol
IMAGE:1248016	AA959891	Mm.10919	cysteine rich protein	Csrp
IMAGE:1248016	AI461843	Mm.10919	cysteine rich protein	Csrp

NIH BioInformatics support provided by BIMAS/CBEL/CIT.
We can be contacted by email.

II.11b Gene Reports are Exportable to Excel

- Tab-delimited gene reports are exportable to Excel using cut & paste or SaveAs DB



Genes with highest HP-X/HP-Y ratios

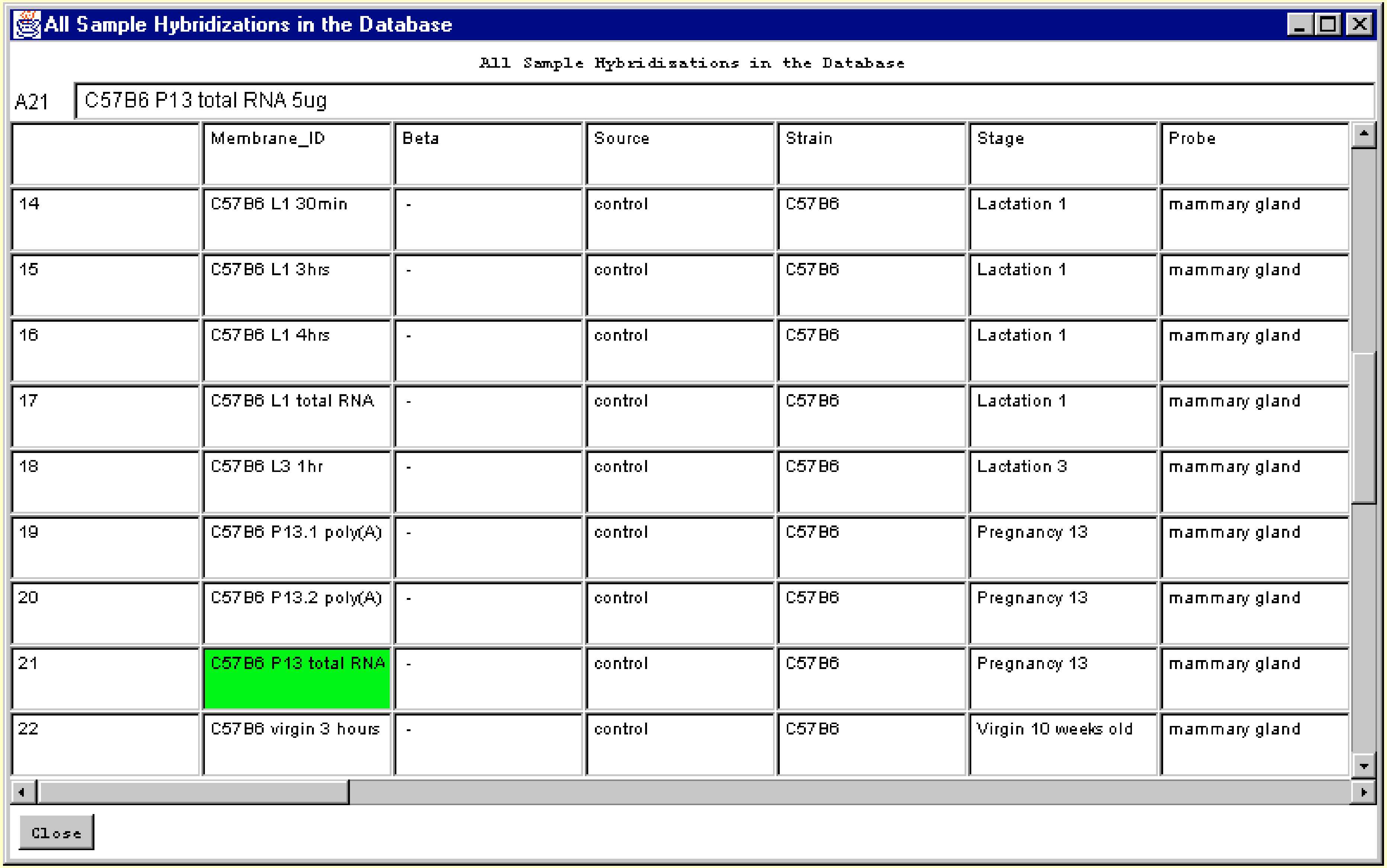
There are 405 genes passing the Filter.

```
GENE REPORT - Filtered genes with 50 Highest ratios HP-X[C57B6 pregnancy day 13] / HP-Y[Stat5a

Grid-Coord          Ratio HP-X/HP-Y     Clone-ID   Gene-Name            Plate-G,R,C           mAdb Cl
[1-G6,21] 1.9088    1382272  Mus musculus Msx-interacting-zinc finger protein 1 (Miz1) mRNA, comp
[1-B4,14] 1.8634    1248264  S100 calcium-binding protein A4           plate[6,B,2]        1248264
[1-A3,17] 1.8456    1248170  Mouse mRNA for SDF2, complete cds      plate[4,A,5]        1248170
[1-H4,15] 1.8449    1248272  ADRENODOXIN PRECURSOR  plate[6,H,3]            1248272  1248272
[1-D5,3]   1.8256    1248351  Abl-interactor 1          plate[7,D,3]            1248351  1248351  AI46337
[1-F7,7]   1.8118    1382525  Acetyl coenzyme A dehydrogenase, medium chain  plate[11,F,7]
[1-C2,19] 1.7997    1247627  Mus musculus mRNA for osteomodulin, complete cds            plate[2,
[1-A3,6]   1.7677    1247777  Mus musculus metalloprotease/disintegrin/cysteine rich protein precurs
[1-B6,7]   1.7562    1381654  TROPOMYOSIN 5, CYTOSKELETAL TYPE           plate[9,B,7]
[1-B6,9]   1.7499    1381703  B-cell translocation gene 2, anti-proliferative         plate[9,B,9]
[1-A5,23] 1.7377    1248527  Mus musculus ubiquitin-conjugating enzyme HR6A mRNA, complete c
[1-C3,10] 1.7316    1247708  Ephrin A1  plate[3,C,10]          1247708  1247708  AA959770
[1-D3,5]   1.7249    1247564  Erythrocyte protein band 7.2     plate[3,D,5]            1247564  1247564
[1-C6,2]   1.7190    1381920  Mus musculus mRNA for NEFA protein, complete cds            plate[9,
[1-D7,16] 1.7081    1382671  Mouse MA-3 (apoptosis-related gene) mRNA, complete cds     plate[12
[1-H3,12] 1.7073    1248169  Histocompatibility 2, T region locus 22      plate[3,H,12]        1248169
[1-H4,20] 1.7039    1248345  Mus musculus alpha-methylacyl-CoA racemase mRNA, complete cds
[1-D2,14] 1.6611    1247820  Tight junction protein 1           plate[2,D,2]            1247820  1247820
[1-A2,22] 1.6598    1247817  Mus musculus ras-related protein (rab18) mRNA, complete cds
[1-D4,6]   1.6528    1248184  Mus musculus bromodomain-containing protein BP75 mRNA, complet
[1-C5,5]   1.6274    1248278  HISTONE H3.3         plate[7,C,5]            1248278  1248278  AI46330
```

II.11c Sample Information Array Reports

- Details are available on all hybridized array samples



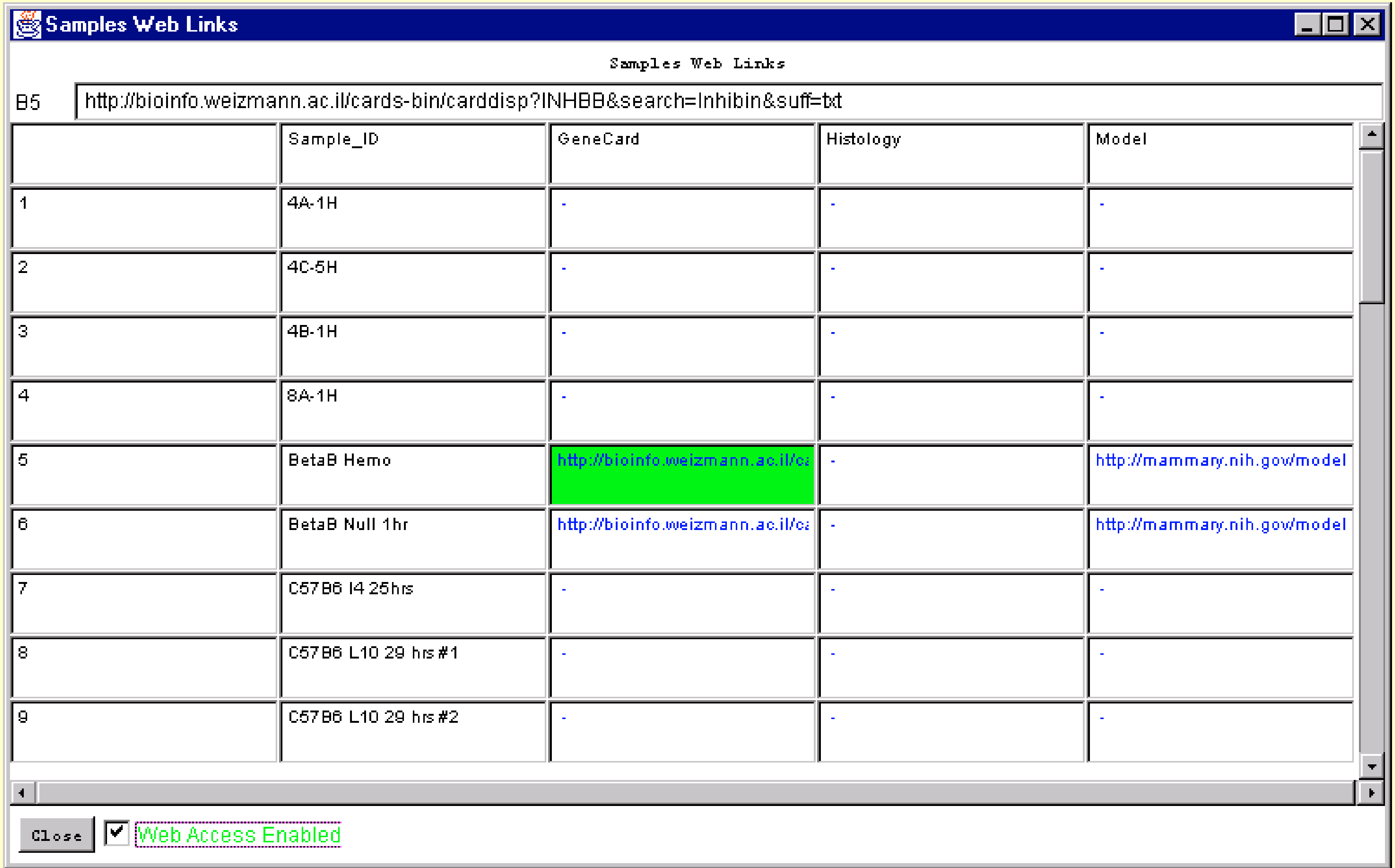
All Sample Hybridizations in the Database

A21 C57B6 P13 total RNA 5ug

	Membrane_ID	Beta	Source	Strain	Stage	Probe
14	C57B6 L1 30min	-	control	C57B6	Lactation 1	mammary gland
15	C57B6 L1 3hrs	-	control	C57B6	Lactation 1	mammary gland
16	C57B6 L1 4hrs	-	control	C57B6	Lactation 1	mammary gland
17	C57B6 L1 total RNA	-	control	C57B6	Lactation 1	mammary gland
18	C57B6 L3 1hr	-	control	C57B6	Lactation 3	mammary gland
19	C57B6 P13.1 poly(A)	-	control	C57B6	Pregnancy 13	mammary gland
20	C57B6 P13.2 poly(A)	-	control	C57B6	Pregnancy 13	mammary gland
21	C57B6 P13 total RNA	-	control	C57B6	Pregnancy 13	mammary gland
22	C57B6 virgin 3 hours	-	control	C57B6	Virgin 10 weeks old	mammary gland

II.11d Sample Web links Array Reports

- Hyper-links to Web databases describing the hybridized samples popup Web browser (customizable for specific database projects)



Samples Web Links

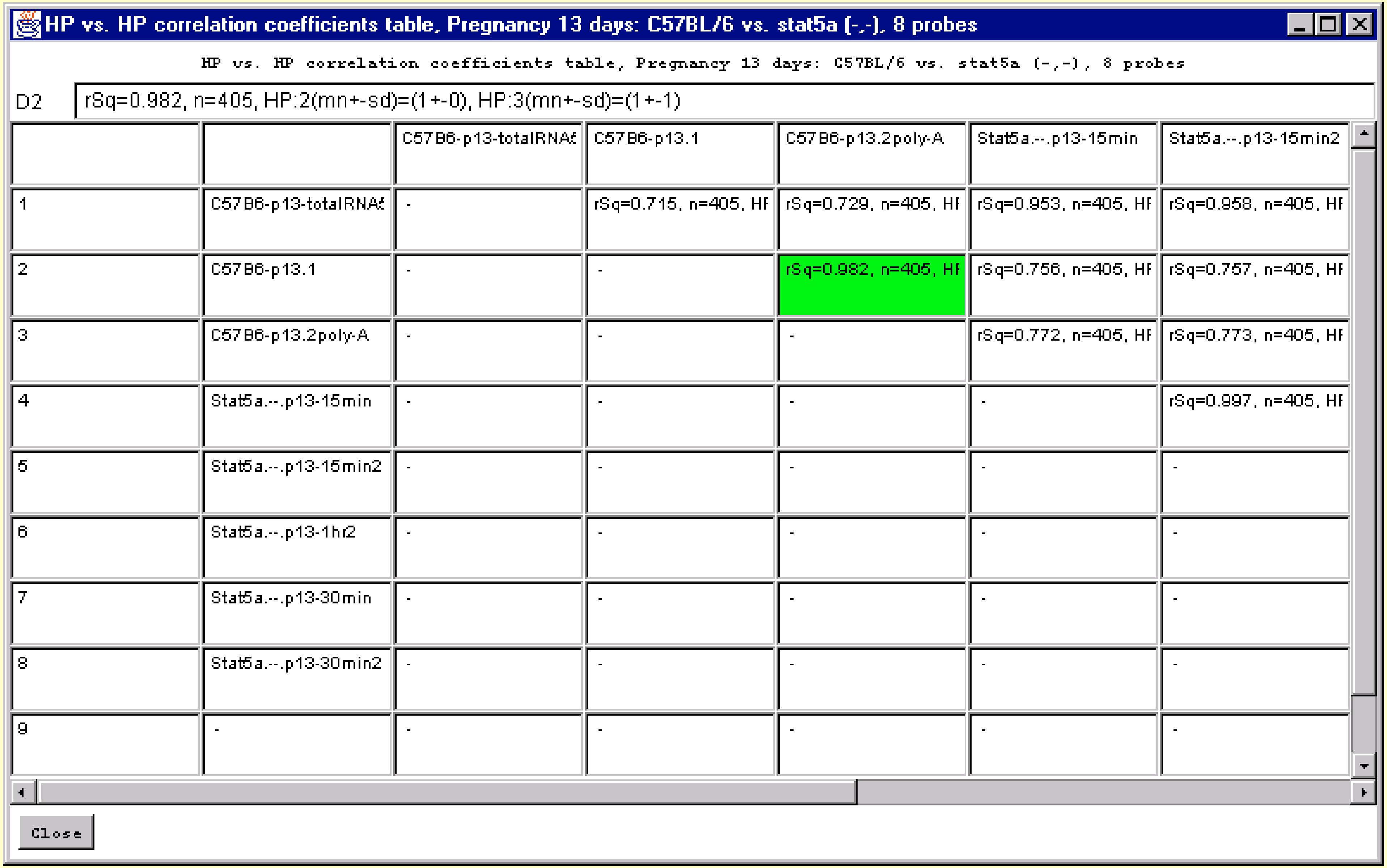
B5 http://bioinfo.weizmann.ac.il/cards-bin/carddisp?INHBB&search=Inhibin&suff=txt

	Sample_ID	GeneCard	Histology	Model
1	4A-1H	-	-	-
2	4C-5H	-	-	-
3	4B-1H	-	-	-
4	8A-1H	-	-	-
5	BetaB Hemo	🟩 http://bioinfo.weizmann.ac.il/ca	-	http://mammary.nih.gov/model
6	BetaB Null 1hr	http://bioinfo.weizmann.ac.il/ca	-	http://mammary.nih.gov/model
7	C57B6 I4 25hrs	-	-	-
8	C57B6 L10 29 hrs #1	-	-	-
9	C57B6 L10 29 hrs #2	-	-	-

☑ Web Access Enabled

II.11e Samples Correlation Reports

- Sample vs. Sample correlation coefficient reports for set of currently Filtered genes



HP vs. HP correlation coefficients table, Pregnancy 13 days: C57BL/6 vs. stat5a (-,-), 8 probes

HP vs. HP correlation coefficients table, Pregnancy 13 days: C57BL/6 vs. stat5a (-,-), 8 probes

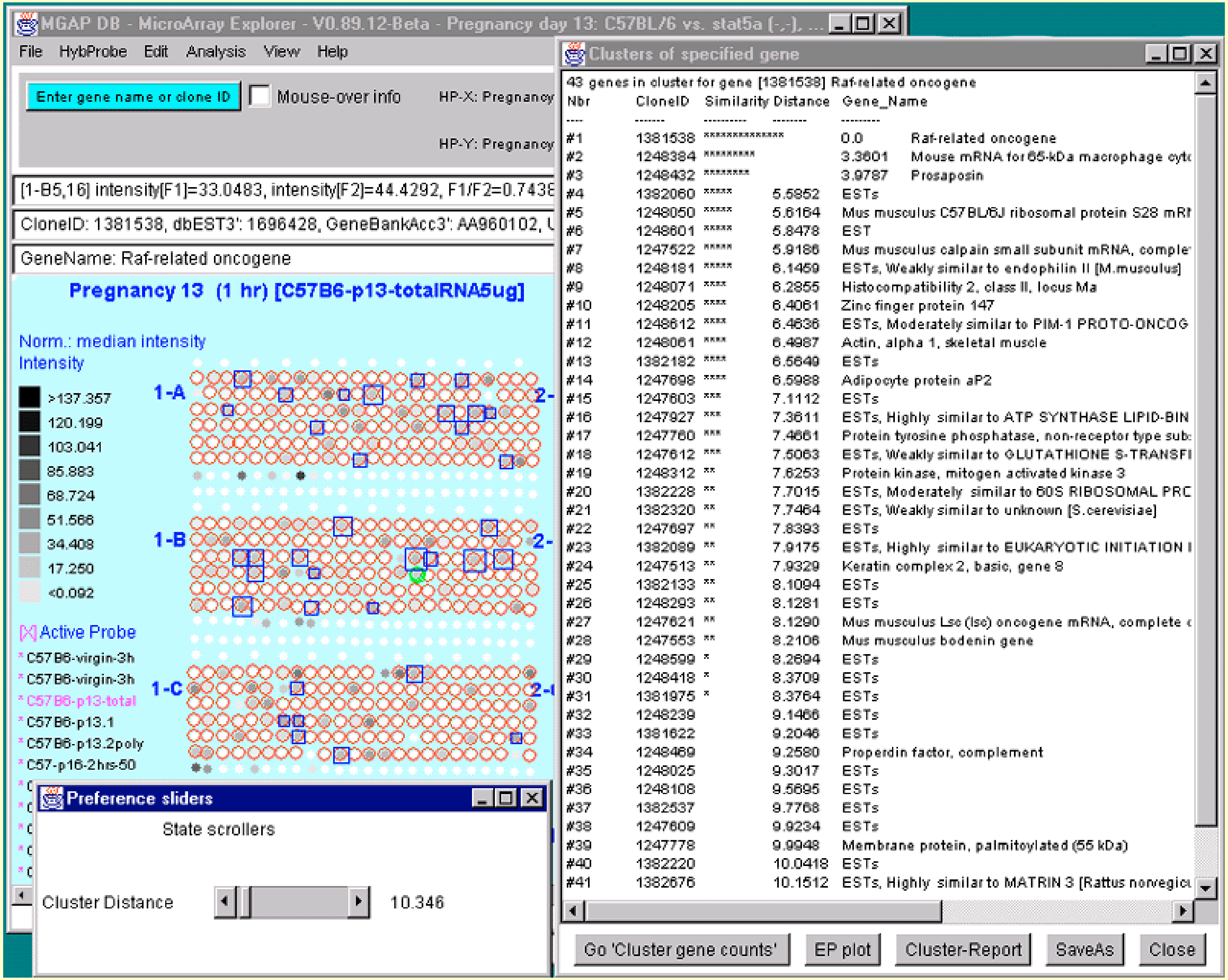
D2 rSq=0.982, n=405, HP:2(mn+-sd)=(1+-0), HP:3(mn+-sd)=(1+-1)

		C57B6-p13-totalRNA	C57B6-p13.1	C57B6-p13.2poly-A	Stat5a.--.p13-15min	Stat5a.--.p13-15min2
1	C57B6-p13-totalRNA	-	rSq=0.715, n=405, H	rSq=0.729, n=405, H	rSq=0.953, n=405, H	rSq=0.958, n=405, H
2	C57B6-p13.1	-	-	rSq=0.982, n=405, H	rSq=0.756, n=405, H	rSq=0.757, n=405, H
3	C57B6-p13.2poly-A	-	-	-	rSq=0.772, n=405, H	rSq=0.773, n=405, H
4	Stat5a.--.p13-15min	-	-	-	-	rSq=0.997, n=405, H
5	Stat5a.--.p13-15min2	-	-	-	-	-
6	Stat5a.--.p13-1hr2	-	-	-	-	-
7	Stat5a.--.p13-30min	-	-	-	-	-
8	Stat5a.--.p13-30min2	-	-	-	-	-
9	-	-	-	-	-	-

Clustering Methods: (4 methods)

II.12a Finding Genes With Similar Expression

- Genes that clustered to Raf-related oncogene with similar expression patterns



[1-B5,16] intensity[F1]=33.0483, intensity[F2]=44.4292, F1/F2=0.743

CloneID: 1381538, dbEST3': 1696428, GeneBankAcc3': AA960102,

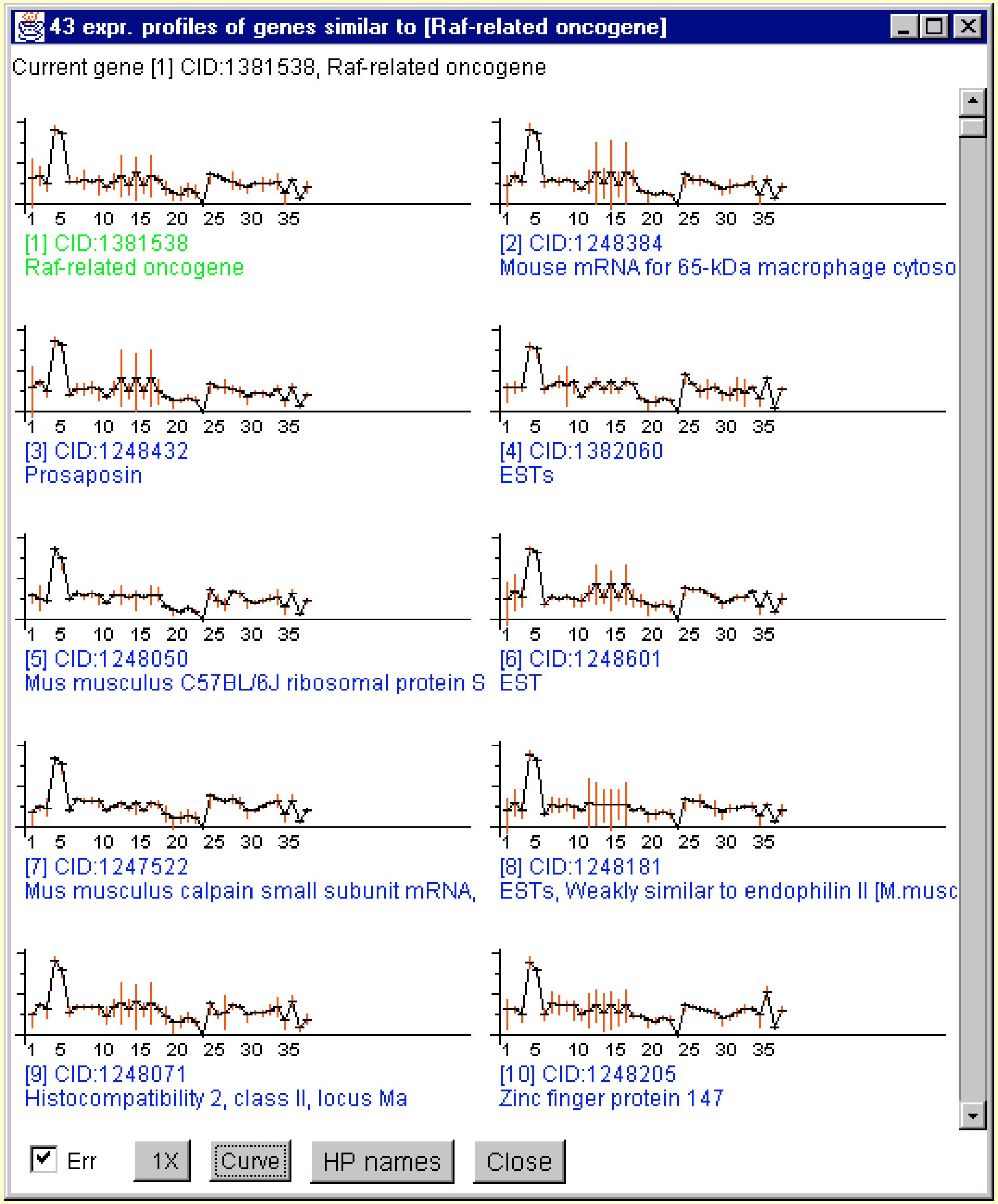
GeneName: Raf-related oncogene

Cluster Distance 10.346

```
43 genes in cluster for gene [1381538] Raf-related oncogene
Nbr    CloneID  Similarity Distance  Gene_Name
----   -------  ---------- --------  ---------
#1     1381538  **************       0.0      Raf-related oncogene
#2     1248384  **********           3.3601   Mouse mRNA for 65-kDa macrophage cyto
#3     1248432  *********            3.9787   Prosaposin
#4     1382060  *****      5.5852   ESTs
#5     1248050  *****      5.6164   Mus musculus C57BL/6J ribosomal protein S28 mRI
#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  ****       6.4061   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    1247697  **         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 (lsc) oncogene mRNA, complete
#28    1247553  **         8.2106   Mus musculus bodenin gene
#29    1248599  *          8.2694   ESTs
#30    1248418  *          8.3709   ESTs
#31    1381975  *          8.3764   ESTs
#32    1248239             9.1466   ESTs
#33    1381622             9.2046   ESTs
#34    1248469             9.2580   Properdin factor, complement
#35    1248025             9.3017   ESTs
#36    1248108             9.5695   ESTs
#37    1382537             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 norvegic
```

II.12b EP Plots for Similar Genes

- Sorted list of EP plots of similar genes that clustered to Raf-related oncogene



43 expr. profiles of genes similar to [Raf-related oncogene]

Current gene [1] CID:1381538, Raf-related oncogene

[1] CID:1381538
Raf-related oncogene

[2] CID:1248384
Mouse mRNA for 65-kDa macrophage cytoso

[3] CID:1248432
Prosaposin

[4] CID:1382060
ESTs

[5] CID:1248050
Mus musculus C57BL/6J ribosomal protein S

[6] CID:1248601
EST

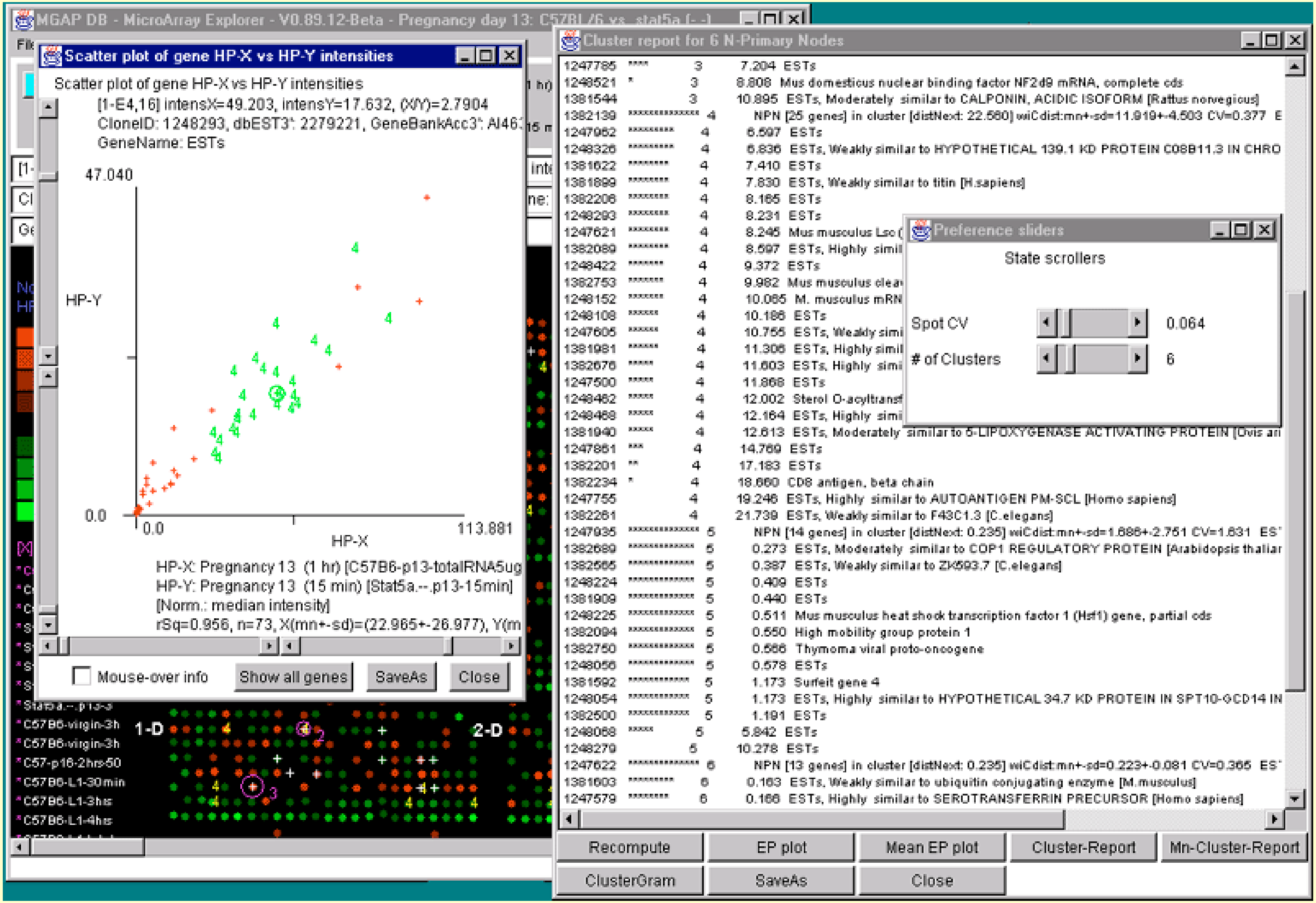
[7] CID:1247522
Mus musculus calpain small subunit mRNA,

[8] CID:1248181
ESTs, Weakly similar to endophilin II [M.musc

[9] CID:1248071
Histocompatibility 2, class II, locus Ma

[10] CID:1248205
Zinc finger protein 147

II.12c Finding K-Clusters of Genes with Similar Expression Patterns (similar to K-means)



Scatter plot of gene HP-X vs HP-Y intensities
[1-E4,16] intensX=49.203, intensY=17.632, (X/Y)=2.7904
CloneID: 1248293, dbEST3': 2279221, GeneBankAcc3': AI46
GeneName: ESTs

HP-X: Pregnancy 13 (1 hr) [C57B6-p13-totalRNA5ug
HP-Y: Pregnancy 13 (15 min) [Stat5a.--.p13-15min]
[Norm.: median intensity]
rSq=0.956, n=73, X(mn+-sd)=(22.965+-26.977), Y(m

```
1247785 *****           3      7.204  ESTs
1248521 *               3      8.808  Mus domesticus nuclear binding factor NF2d9 mRNA, complete cds
1381544                 3     10.895  ESTs, Moderately  similar to CALPONIN, ACIDIC ISOFORM [Rattus norvegious]
1382139 ************** 4        NPN [25 genes] in cluster [distNext: 22.580] wiCdist:mn+-sd=11.919+-4.503 CV=0.377  E
1247962 *********      4       6.597  ESTs
1248326 *********      4       6.836  ESTs, Weakly similar to HYPOTHETICAL 139.1 KD PROTEIN C08B11.3 IN CHRO
1381622 ********       4       7.410  ESTs
1381699 ********       4       7.830  ESTs, Weakly similar to titin [H.sapiens]
1382206 ********       4       8.165  ESTs
1248293 ********       4       8.231  ESTs
1247621 ********       4       8.245  Mus musculus Lsc (
1382089 ********       4       8.597  ESTs, Highly  simil
1248422 *******        4       9.372  ESTs
1382753 *******        4       9.982  Mus musculus clea
1248152 *******        4      10.065  M. musculus mRN
1248108 ******         4      10.186  ESTs
1247605 ******         4      10.755  ESTs, Weakly simi
1381981 ******         4      11.306  ESTs, Highly simil
1382676 *****          4      11.603  ESTs, Highly  simi
1247500 *****          4      11.868  ESTs
1248462 *****          4      12.002  Sterol O-acyltransf
1248468 *****          4      12.164  ESTs, Highly  simi
1381940 *****          4      12.613  ESTs, Moderately  similar to 5-LIPOXYGENASE ACTIVATING PROTEIN [Ovis ari
1247861 ***           4      14.769  ESTs
1382201 **            4      17.183  ESTs
1382234 *             4      18.660  CD8 antigen, beta chain
1247755               4      19.246  ESTs, Highly  similar to AUTOANTIGEN PM-SCL [Homo sapiens]
1382261               4      21.739  ESTs, Weakly similar to F43C1.3 [C.elegans]
1247935 ************** 5        NPN [14 genes] in cluster [distNext: 0.235] wiCdist:mn+-sd=1.686+-2.761 CV=1.631  ES
1382689 *************  5        0.273  ESTs, Moderately  similar to COP1 REGULATORY PROTEIN [Arabidopsis thaliar
1382585 *************  5        0.387  ESTs, Weakly similar to ZK593.7 [C.elegans]
1248224 *************  5        0.409  ESTs
1381909 *************  5        0.440  ESTs
1248225 *************  5        0.511  Mus musculus heat shock transcription factor 1 (Hsf1) gene, partial cds
1382094 *************  5        0.550  High mobility group protein 1
1382750 *************  5        0.566  Thymoma viral proto-oncogene
1248056 *************  5        0.578  ESTs
1381592 ************   5        1.173  Surfeit gene 4
1248054 ************   5        1.173  ESTs, Highly  similar to HYPOTHETICAL 34.7 KD PROTEIN IN SPT10-GCD14 IN
1382500 ************   5        1.191  ESTs
1248068 *****          5      5.842  ESTs
1248279               5     10.278  ESTs
1247622 ************** 6        NPN [13 genes] in cluster [distNext: 0.235] wiCdist:mn+-sd=0.223+-0.081 CV=0.365  ES
1381603 *********      6       0.163  ESTs, Weakly similar to ubiquitin conjugating enzyme [M.musculus]
1247579 ********       6       0.166  ESTs, Highly  similar to SEROTRANSFERRIN PRECURSOR [Homo sapiens]
```

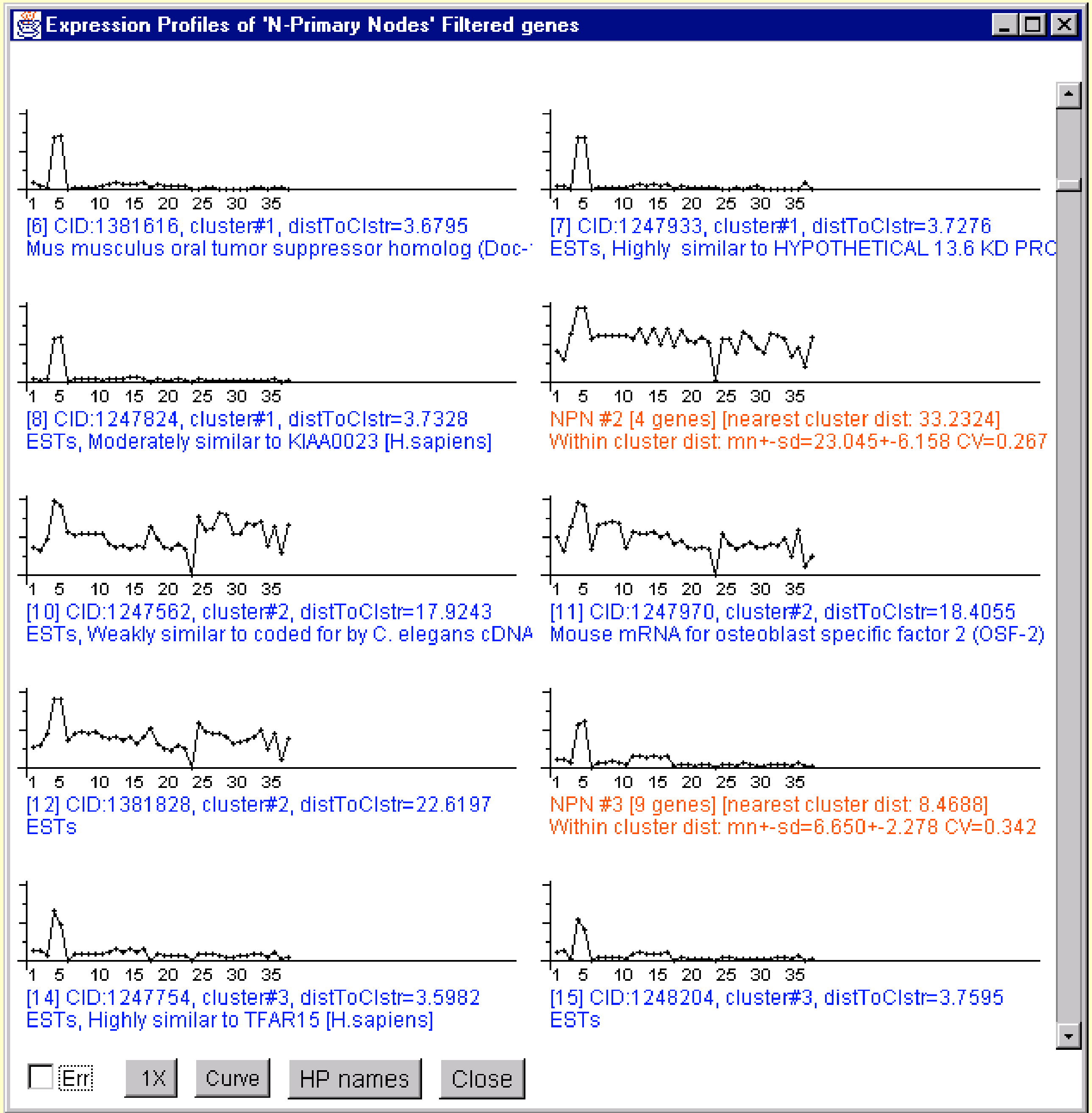
State scrollers

Spot CV 0.064

# of Clusters 6

II.12d Expression Profiles of Clusters

- Scrollable list of EP plots showing genes from clusters #1, #2, #3 (from figure II.12c)



[6] CID:1381616, cluster#1, distToClstr=3.6795
Mus musculus oral tumor suppressor homolog (Doc-

[7] CID:1247933, cluster#1, distToClstr=3.7276
ESTs, Highly similar to HYPOTHETICAL 13.6 KD PRO

[8] CID:1247824, cluster#1, distToClstr=3.7328
ESTs, Moderately similar to KIAA0023 [H.sapiens]

NPN #2 [4 genes] [nearest cluster dist: 33.2324]
Within cluster dist: mn+-sd=23.045+-6.158 CV=0.267

[10] CID:1247562, cluster#2, distToClstr=17.9243
ESTs, Weakly similar to coded for by C. elegans cDNA

[11] CID:1247970, cluster#2, distToClstr=18.4055
Mouse mRNA for osteoblast specific factor 2 (OSF-2)

[12] CID:1381828, cluster#2, distToClstr=22.6197
ESTs

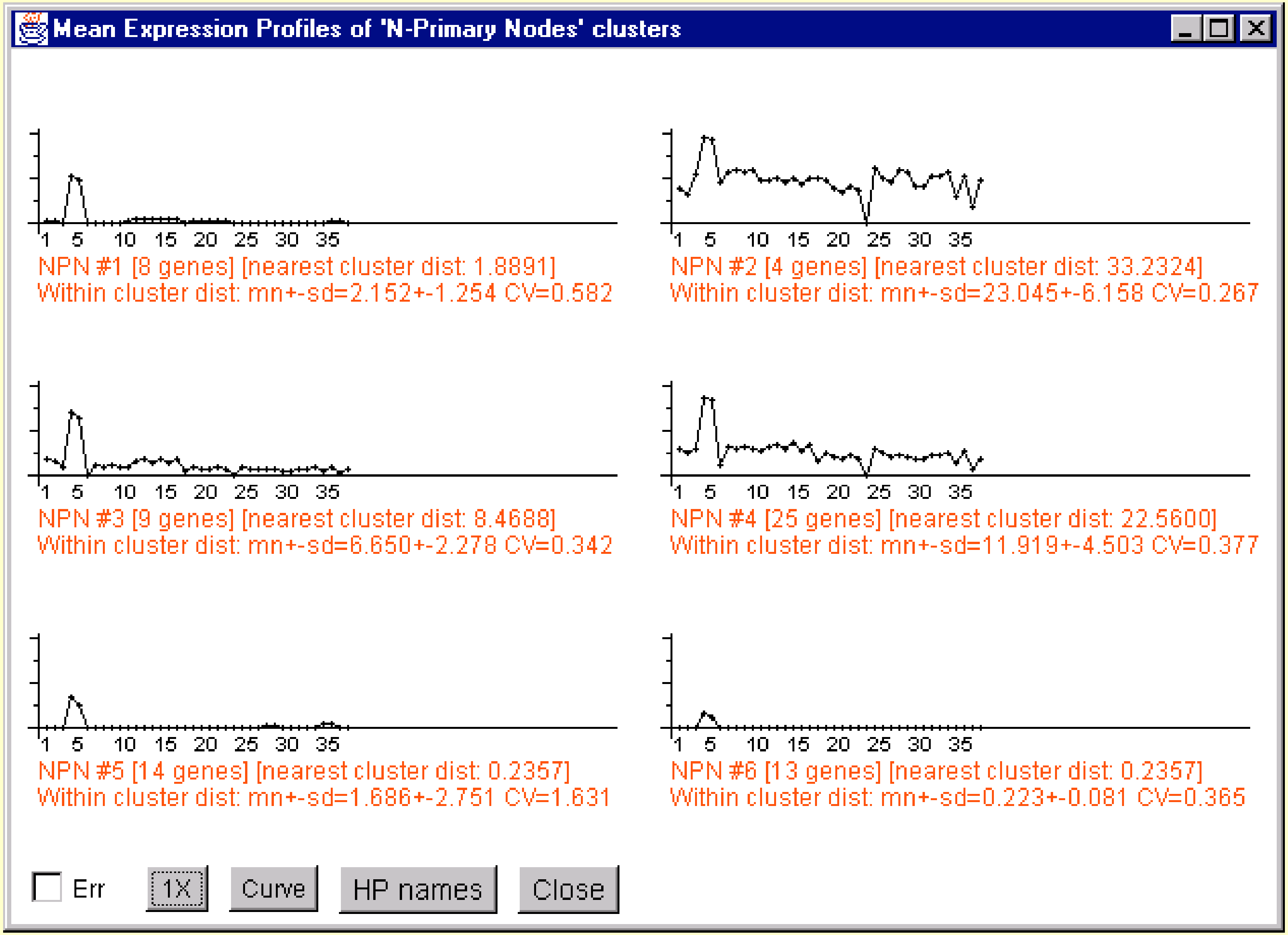
NPN #3 [9 genes] [nearest cluster dist: 8.4688]
Within cluster dist: mn+-sd=6.650+-2.278 CV=0.342

[14] CID:1247754, cluster#3, distToClstr=3.5982
ESTs, Highly similar to TFAR15 [H.sapiens]

[15] CID:1248204, cluster#3, distToClstr=3.7595
ESTs

II.12e Mean Expression Profile Plots of Clusters

- Mean clusters and their statistics (from figure II.12c). Error bars are standard-deviation of genes' intensities in each cluster



NPN #1 [8 genes] [nearest cluster dist: 1.8891]
Within cluster dist: mn+-sd=2.152+-1.254 CV=0.582

NPN #2 [4 genes] [nearest cluster dist: 33.2324]
Within cluster dist: mn+-sd=23.045+-6.158 CV=0.267

NPN #3 [9 genes] [nearest cluster dist: 8.4688]
Within cluster dist: mn+-sd=6.650+-2.278 CV=0.342

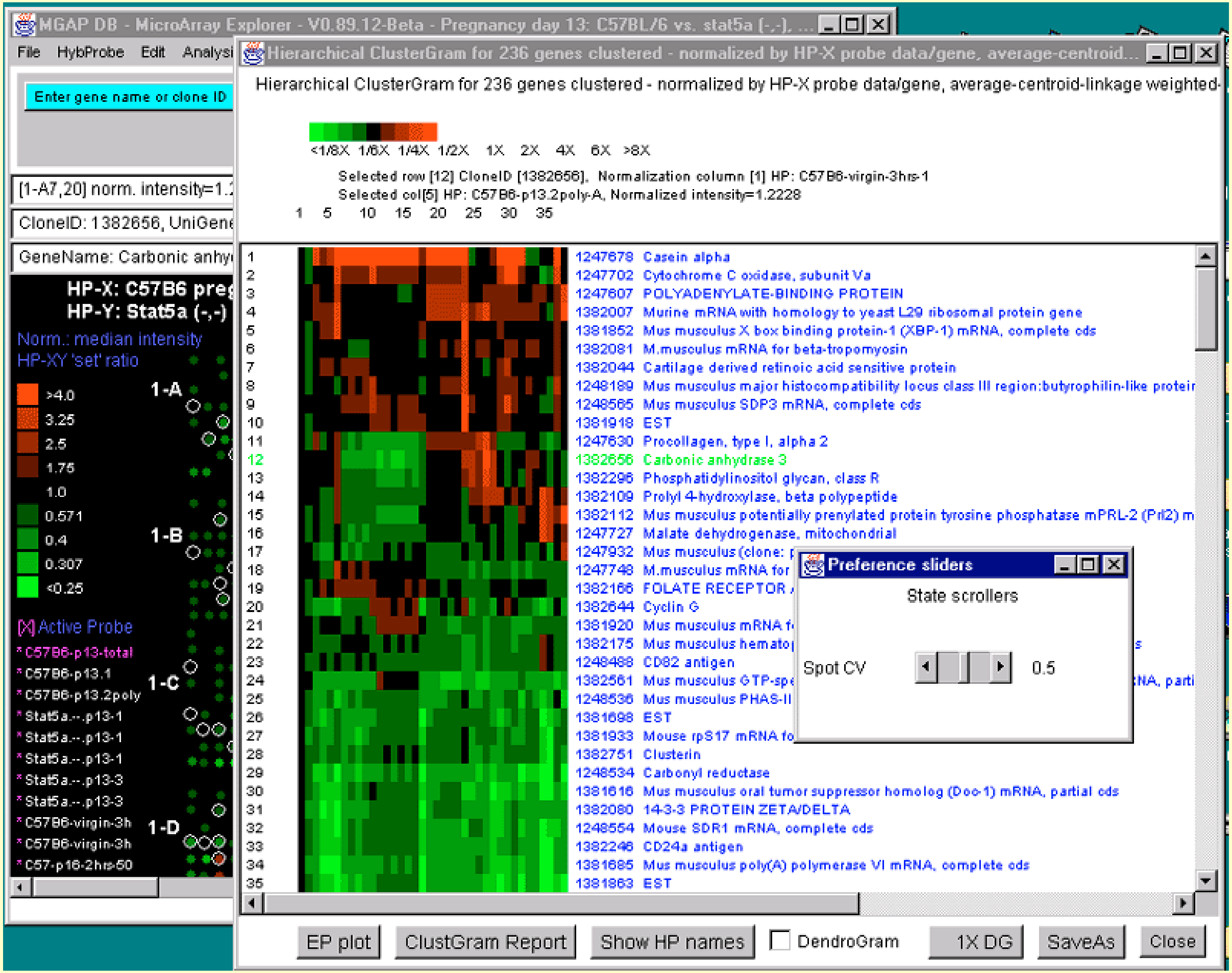
NPN #4 [25 genes] [nearest cluster dist: 22.5600]
Within cluster dist: mn+-sd=11.919+-4.503 CV=0.377

NPN #5 [14 genes] [nearest cluster dist: 0.2357]
Within cluster dist: mn+-sd=1.686+-2.751 CV=1.631

NPN #6 [13 genes] [nearest cluster dist: 0.2357]
Within cluster dist: mn+-sd=0.223+-0.081 CV=0.365

II.13a Hierarchical Clustering

ClusterGrams of Expression Profiles



[1-A7,20] norm. intensity=1.2

CloneID: 1382656, UniGene

GeneName: Carbonic anhy

- C57B6-p13-total
- C57B6-p13.1
- C57B6-p13.2poly
- Stat5a.--.p13-1
- Stat5a.--.p13-1
- Stat5a.--.p13-1
- Stat5a.--.p13-3
- Stat5a.--.p13-3
- C57B6-virgin-3h
- C57B6-virgin-3h
- C57-p16-2hrs-50

Hierarchical ClusterGram for 236 genes clustered - normalized by HP-X probe data/gene, average-centroid-linkage weighted-

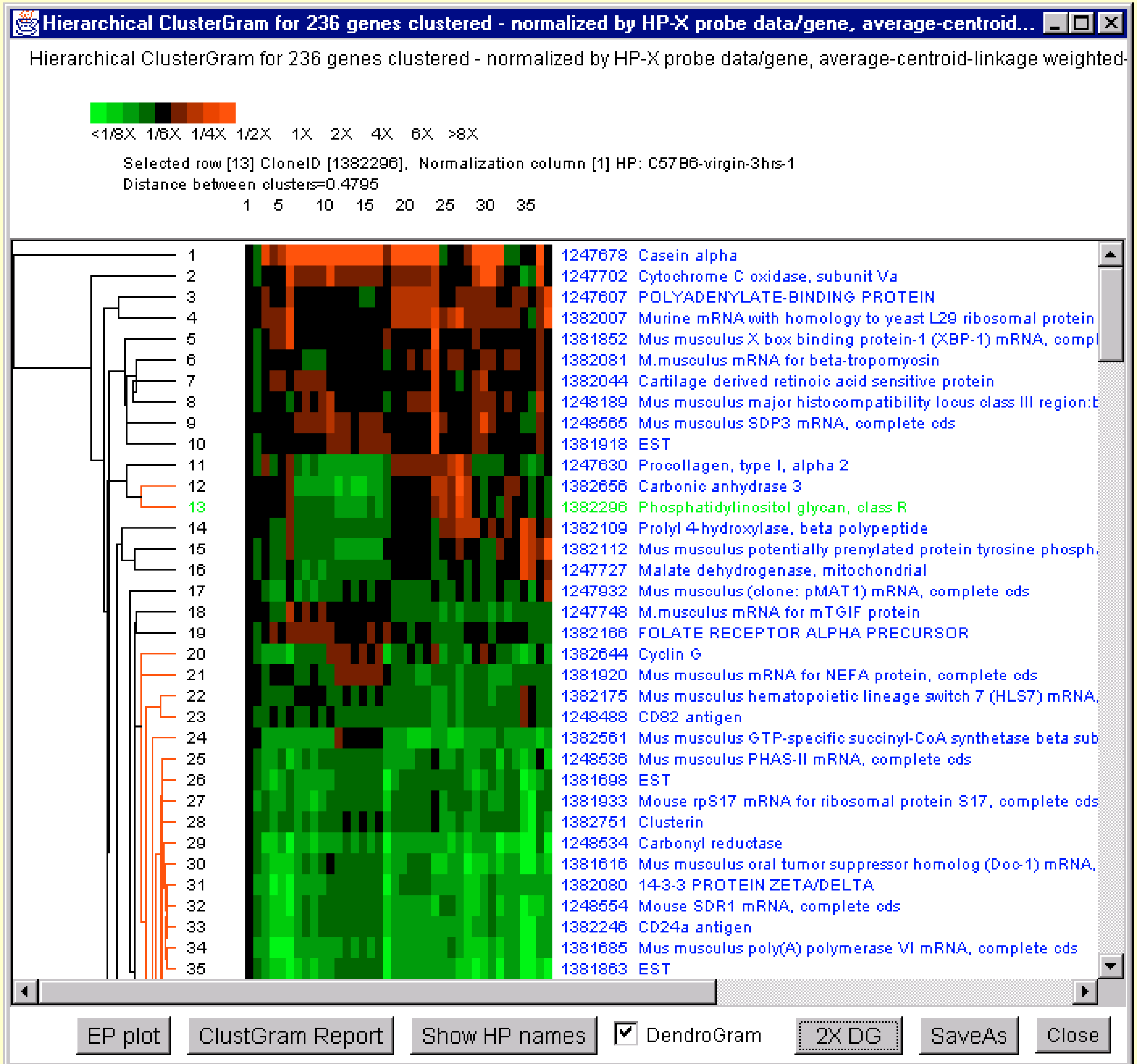
Selected row [12] CloneID [1382656], Normalization column [1] HP: C57B6-virgin-3hrs-1
Selected col[5] HP: C57B6-p13.2poly-A, Normalized intensity=1.2228

```
1   1247678  Casein alpha
2   1247702  Cytochrome C oxidase, subunit Va
3   1247607  POLYADENYLATE-BINDING PROTEIN
4   1382007  Murine mRNA with homology to yeast L29 ribosomal protein gene
5   1381852  Mus musculus X box binding protein-1 (XBP-1) mRNA, complete cds
6   1382081  M.musculus mRNA for beta-tropomyosin
7   1382044  Cartilage derived retinoic acid sensitive protein
8   1248189  Mus musculus major histocompatibility locus class III region:butyrophilin-like protein
9   1248565  Mus musculus SDP3 mRNA, complete cds
10  1381918  EST
11  1247630  Procollagen, type I, alpha 2
12  1382656  Carbonic anhydrase 3
13  1382296  Phosphatidylinositol glycan, class R
14  1382109  Prolyl 4-hydroxylase, beta polypeptide
15  1382112  Mus musculus potentially prenylated protein tyrosine phosphatase mPRL-2 (Prl2) m
16  1247727  Malate dehydrogenase, mitochondrial
17  1247932  Mus musculus (clone: p
18  1247748  M.musculus mRNA for
19  1382166  FOLATE RECEPTOR A
20  1382644  Cyclin G
21  1381920  Mus musculus mRNA f
22  1382175  Mus musculus hemato                                                          s
23  1248488  CD82 antigen
24  1382561  Mus musculus GTP-spe                                                         NA, parti
25  1248536  Mus musculus PHAS-II
26  1381698  EST
27  1381933  Mouse rpS17 mRNA fo
28  1382751  Clusterin
29  1248534  Carbonyl reductase
30  1381616  Mus musculus oral tumor suppressor homolog (Doc-1) mRNA, partial cds
31  1382080  14-3-3 PROTEIN ZETA/DELTA
32  1248554  Mouse SDR1 mRNA, complete cds
33  1382246  CD24a antigen
34  1381685  Mus musculus poly(A) polymerase VI mRNA, complete cds
35  1381863  EST
```

Spot CV 0.5

II.13b Hierarchical Clustering Dendrogram

- Clusters less than cluster distance from each other are shown in red (from figure II.12f)



Hierarchical ClusterGram for 236 genes clustered - normalized by HP-X probe data/gene, average-centroid...

Hierarchical ClusterGram for 236 genes clustered - normalized by HP-X probe data/gene, average-centroid-linkage weighted-

Selected row [13] CloneID [1382296], Normalization column [1] HP: C57B6-virgin-3hrs-1
Distance between clusters=0.4795

1	1247678	Casein alpha
2	1247702	Cytochrome C oxidase, subunit Va
3	1247607	POLYADENYLATE-BINDING PROTEIN
4	1382007	Murine mRNA with homology to yeast L29 ribosomal protein
5	1381852	Mus musculus X box binding protein-1 (XBP-1) mRNA, compl
6	1382081	M.musculus mRNA for beta-tropomyosin
7	1382044	Cartilage derived retinoic acid sensitive protein
8	1248189	Mus musculus major histocompatibility locus class III region:b
9	1248565	Mus musculus SDP3 mRNA, complete cds
10	1381918	EST
11	1247630	Procollagen, type I, alpha 2
12	1382656	Carbonic anhydrase 3
13	1382296	Phosphatidylinositol glycan, class R
14	1382109	Prolyl 4-hydroxylase, beta polypeptide
15	1382112	Mus musculus potentially prenylated protein tyrosine phosph,
16	1247727	Malate dehydrogenase, mitochondrial
17	1247932	Mus musculus (clone: pMAT1) mRNA, complete cds
18	1247748	M.musculus mRNA for mTGIF protein
19	1382166	FOLATE RECEPTOR ALPHA PRECURSOR
20	1382644	Cyclin G
21	1381920	Mus musculus mRNA for NEFA protein, complete cds
22	1382175	Mus musculus hematopoietic lineage switch 7 (HLS7) mRNA,
23	1248488	CD82 antigen
24	1382561	Mus musculus GTP-specific succinyl-CoA synthetase beta sub
25	1248536	Mus musculus PHAS-II mRNA, complete cds
26	1381698	EST
27	1381933	Mouse rpS17 mRNA for ribosomal protein S17, complete cds
28	1382751	Clusterin
29	1248534	Carbonyl reductase
30	1381616	Mus musculus oral tumor suppressor homolog (Doc-1) mRNA,
31	1382080	14-3-3 PROTEIN ZETA/DELTA
32	1248554	Mouse SDR1 mRNA, complete cds
33	1382246	CD24a antigen
34	1381685	Mus musculus poly(A) polymerase VI mRNA, complete cds
35	1381863	EST

Summary of MAExplorer

- MAExplorer is used as a stand-alone application or as applet over the Web
- Accepts different array geometries, spot supports, ^{33}P or Cy3/Cy5 labeling, scanners
- Analyzes multiple probes, X-Y replicate sets, expression profiles, replicate spots
- Provides direct manipulation of array pseudo images, scatter-plots, histograms, clustergrams, dendrograms, silhouette plots, spreadsheets
- Data filters genes by gene subsets, spot intensities and ratios, and statistical tests, etc.
- Set operations on gene subsets help manage search results
- Uses active Web links to genomic, histology and model Web databases
- Generates reports as Web-accessible spreadsheets or exportable to Excel
- Users may save their data-mining session state locally for later use or sharing
- Building tools to import commercial and academic quantified micro array data
- MAExplorer used to identify genes in MGAP DB preferentially expressed during lactation. Results verified using northern blots (NIDDK), ***Nucleic Acids Res.* 28:4452-4459 (2000).**
- Online documentation (manual, tutorials, examples, etc.) is available on Web site

Some MAExplorer URL References

- **Home Page (includes the following and other links)**
  http://www.lecb.ncifcrf.gov/MAExplorer/

- **Reference Manual (including tutorials, and use with other arrays sections)**
  http://www.lecb.ncifcrf.gov/MAExplorer/hmaeHelp.html (online)
  http://www.lecb.ncifcrf.gov/MAExplorer/MaeRefMan.zip (download)

- **Overview of MAExplorer**
  http://www.lecb.ncifcrf.gov/MAExplorer/PDF/Overview-MAE.pdf

- **Examples of data mining with MAExplorer**
  http://www.lecb.ncifcrf.gov/MAExplorer/Examples-MAE-session.pdf

- **Using with mAdb with MAExplorer**
  http://www.lecb.ncifcrf.gov/MAExplorer/Using-mAdb-with-MAExplorer.pdf

- ***Nucleic Acids Res*. (2000) 28:4452 paper**
  http://www.lecb.ncifcrf.gov/MAExplorer/lemkin-NAR-2000-Vol28-pp4452.pdf

- **Download MAExplorer (includes 38 samples from MGAP DB)**
  http://www.lecb.ncifcrf.gov/MAExplorer/hmaeInstall.html

Using MAExplorer with mAdb data

- The **NCI/CIT mAdb Web microarray database server** is an array data repository and analysis facility for microarrays created in conjunction with the NCI-ATC facility.

  http://nciarray.nci.nih.gov/

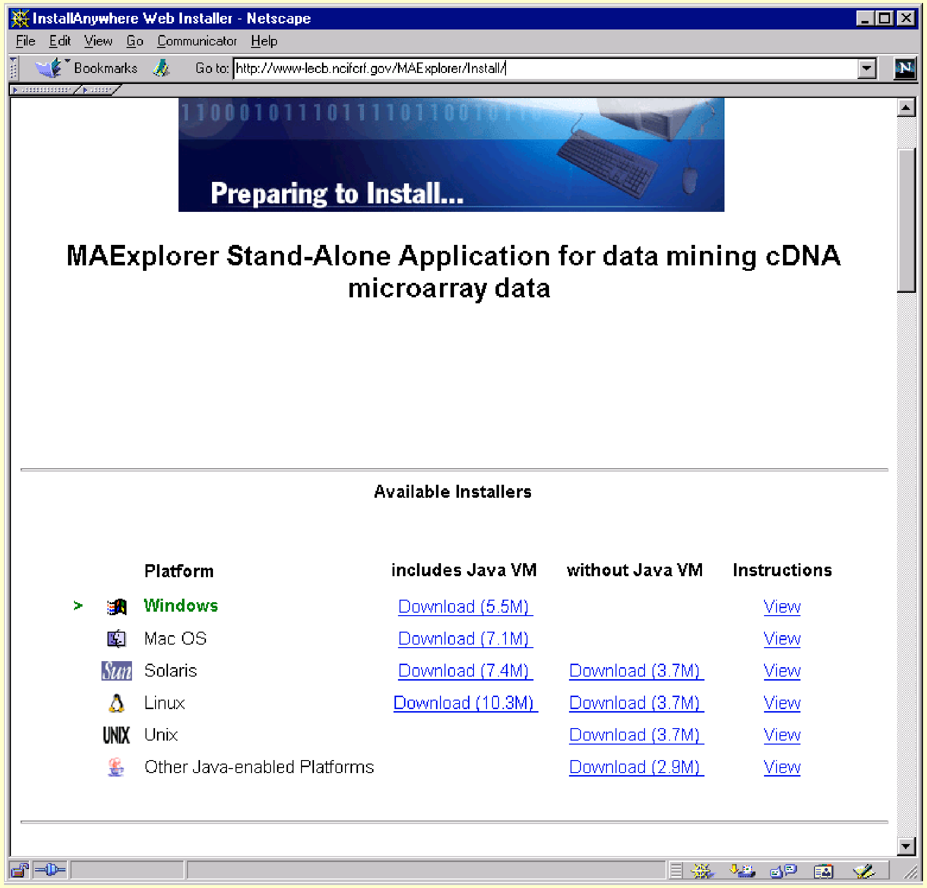
- It can create a set of data files, downloaded as a Zip file from the mAdb, in a format compatible with MAExplorer

- Section III describes the procedure for **downloading MAExplorer.** You should periodically check the MAExplorer Web site to see if there is a major revision that you might want to download

- Section IV describes the procedure for **downloading a mAdb data set** and starting MAExplorer on that data.

- **Help desk for MAExplorer :** *mae@ncifcrf.gov*

III. Installing MicroArray Explorer on Your Computer



MAExplorer Stand-Alone Application for data mining cDNA microarray data

Available Installers

Platform	includes Java VM	without Java VM	Instructions
Windows	Download (5.5M)		View
Mac OS	Download (7.1M)		View
Solaris	Download (7.4M)	Download (3.7M)	View
Linux	Download (10.3M)	Download (3.7M)	View
Unix		Download (3.7M)	View
Other Java-enabled Platforms		Download (2.9M)	View

Outline

1. MAExplorer home page
2. Download installer to your computer
3. Run the installer
4. Test it on MGAP sample database

III. Procedure to download & install MAExplorer

- 1. Go to **http://www.lecb.ncifcrf.gov/MAExplorer** with your Web browser.

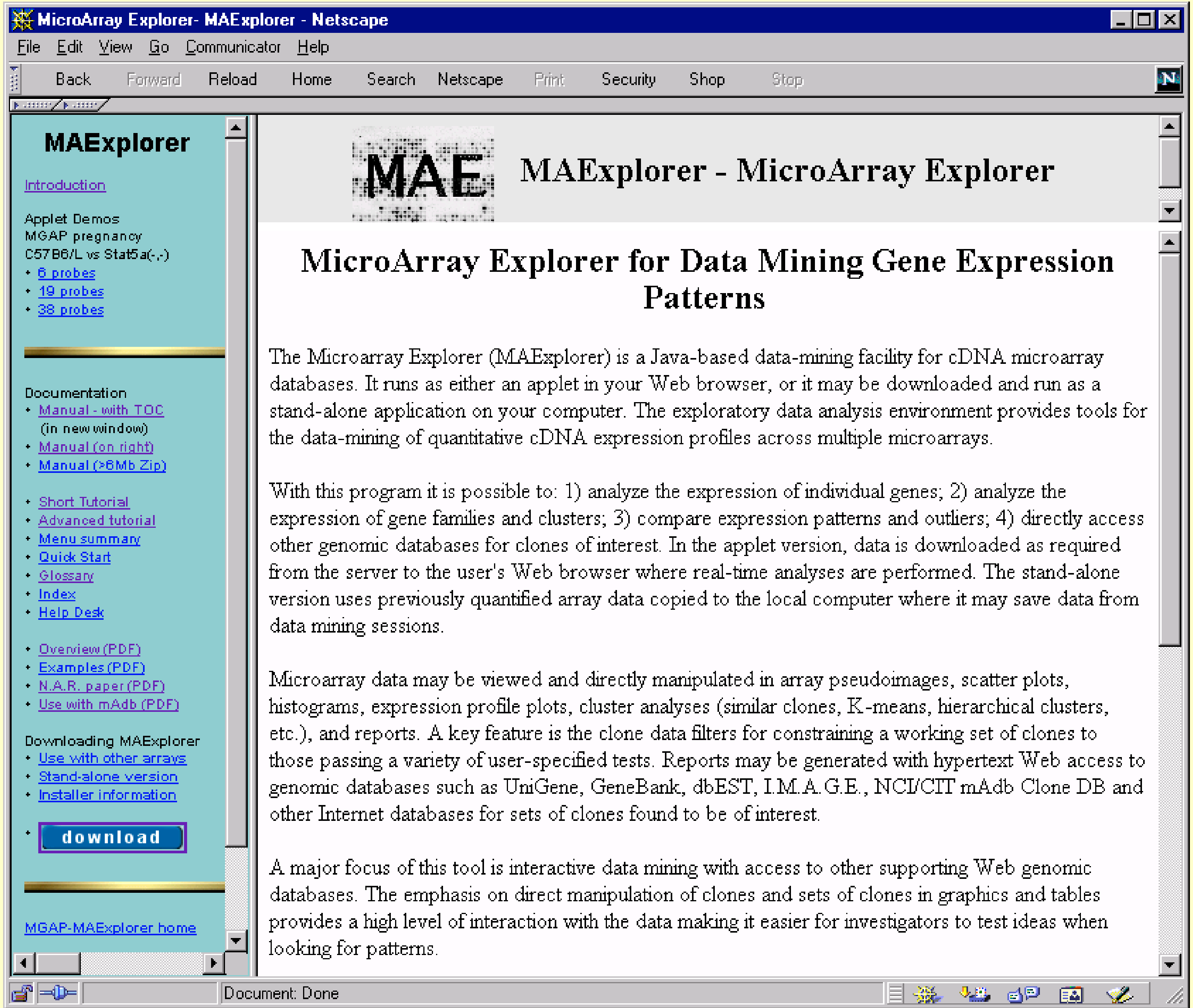
- 2. Select **Download** to start the install process. It uses the InstallAnywhere™ program. You have a choice of:

- 3.1 Allowing InstallAnywhere™ to select the installer and request where you want to install it (eg. in Windows this would be C:\Program Files\MAExplorer), or

- 3.2 You may download the installer file and select where you want to install it.
- A) Find your computer **Platform** in the list. Click on the corresponding **Download** word and save the installer on your computer.
- B) Go to **View** for your platform in the same download Web page to see how to finish the installation for your particular platform.
- C) Now install MAExplorer on your computer in the location you desire.

- 4. You are ready to use MAExplorer. In Windows Start menu, click on MAExplorer. After it starts, select "Open file DB" in the File | Database menu.

III.1 MAExplorer home page - press "download"
http://www.lecb.ncifcrf.gov/MAExplorer



MAExplorer

Introduction

Applet Demos
MGAP pregnancy
C57B6/L vs Stat5a(-,-)
- 6 probes
- 19 probes
- 38 probes

Documentation
- Manual - with TOC
  (in new window)
- Manual (on right)
- Manual (>6Mb Zip)

- Short Tutorial
- Advanced tutorial
- Menu summary
- Quick Start
- Glossary
- Index
- Help Desk

- Overview (PDF)
- Examples (PDF)
- N.A.R. paper (PDF)
- Use with mAdb (PDF)

Downloading MAExplorer
- Use with other arrays
- Stand-alone version
- Installer information

- download

MGAP-MAExplorer home

MAExplorer - MicroArray Explorer

MicroArray Explorer for Data Mining Gene Expression Patterns

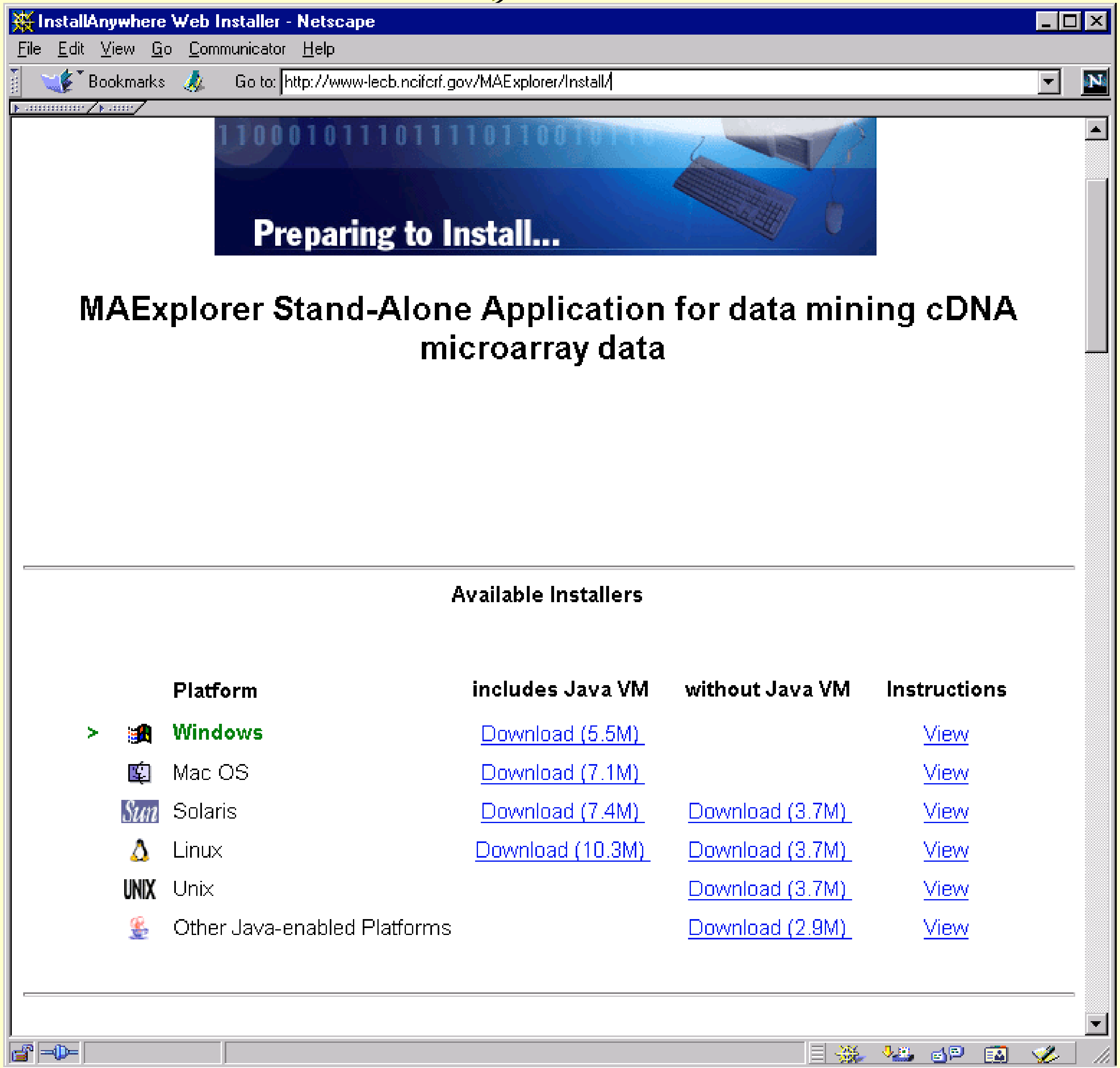
The Microarray Explorer (MAExplorer) is a Java-based data-mining facility for cDNA microarray databases. It runs as either an applet in your Web browser, or it may be downloaded and run as a stand-alone application on your computer. The exploratory data analysis environment provides tools for the data-mining 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 pseudoimages, 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, GeneBank, 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.

III.2 Download Stand-alone version Web page - find your "Platform", then select "Download"



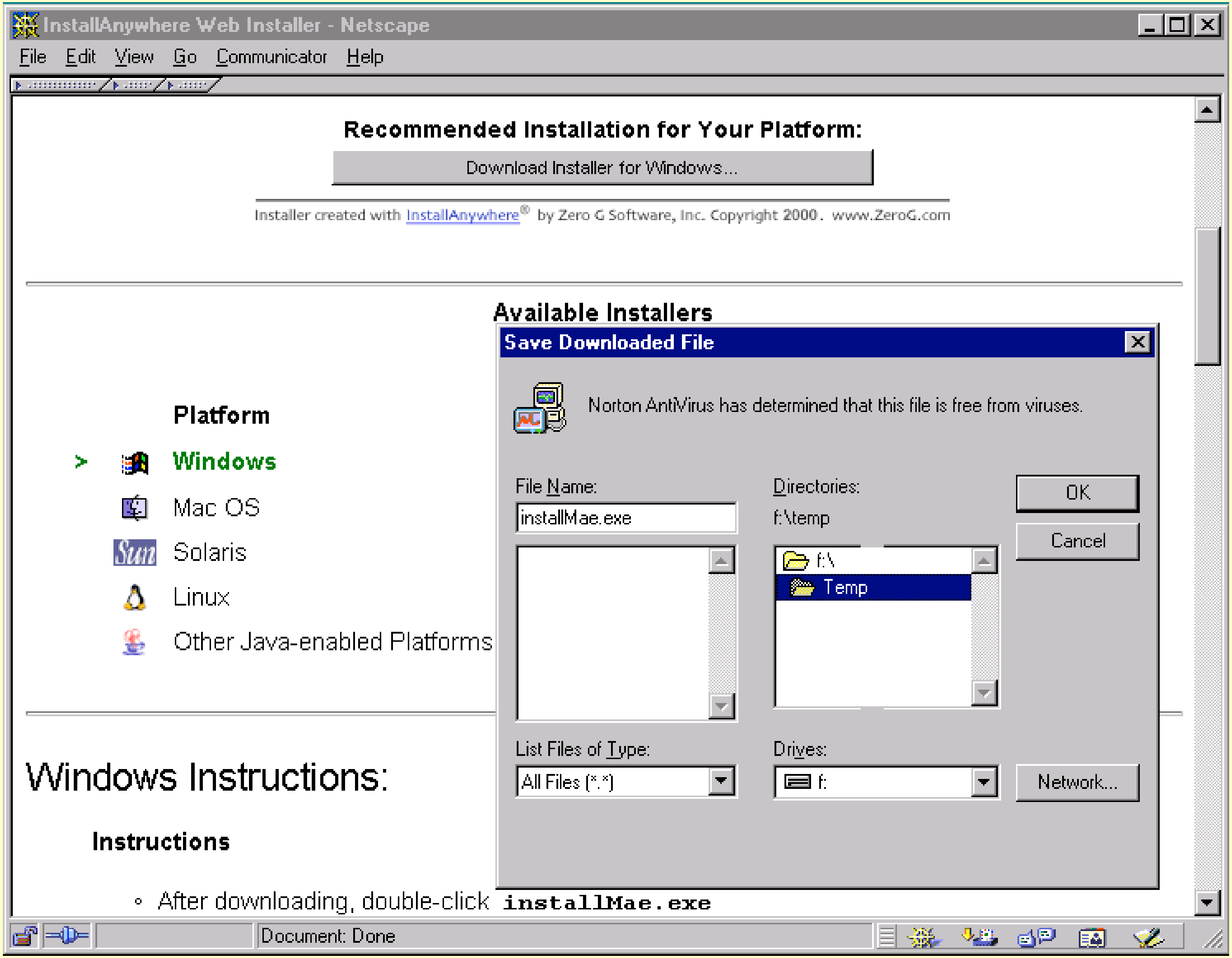
MAExplorer Stand-Alone Application for data mining cDNA microarray data

---

Available Installers

	Platform	includes Java VM	without Java VM	Instructions
>	**Windows**	Download (5.5M)		View
	Mac OS	Download (7.1M)		View
	Solaris	Download (7.4M)	Download (3.7M)	View
	Linux	Download (10.3M)	Download (3.7M)	View
	Unix		Download (3.7M)	View
	Other Java-enabled Platforms		Download (2.9M)	View

III.3 Save the installer on your local computer



Recommended Installation for Your Platform:

Download Installer for Windows...

Installer created with InstallAnywhere® by Zero G Software, Inc. Copyright 2000. www.ZeroG.com

Available Installers

Platform

> **Windows**
Mac OS
Solaris
Linux
Other Java-enabled Platforms

Windows Instructions:

Instructions

- After downloading, double-click `installMae.exe`

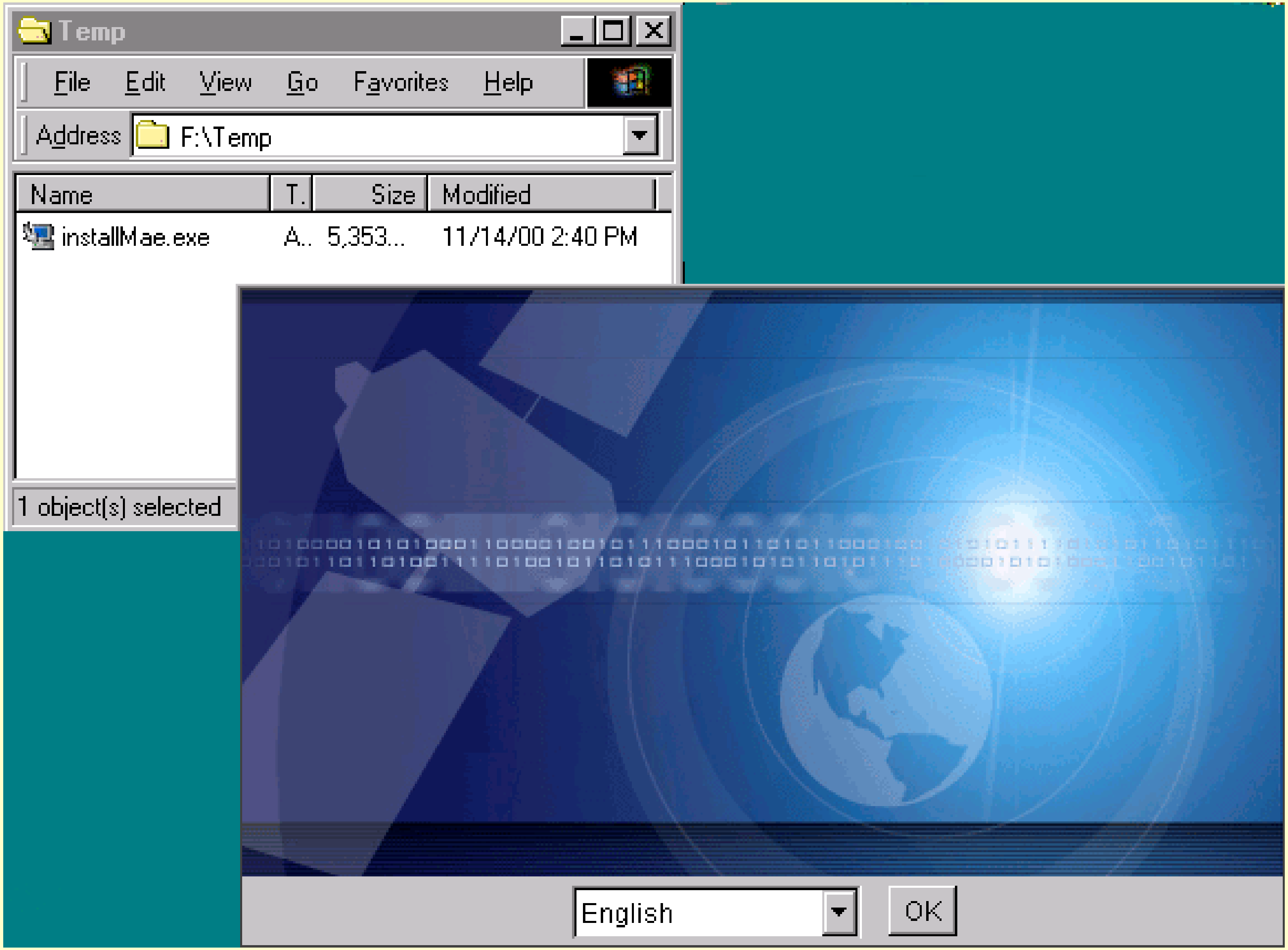
Save Downloaded File

Norton AntiVirus has determined that this file is free from viruses.

File Name: installMae.exe
Directories: f:\temp
List Files of Type: All Files (*.*)
Drives: f:

Document: Done

III.4 Start the installer - e.g. in Windows, click on installMAE.exe. Then answer questions, “OK” etc.

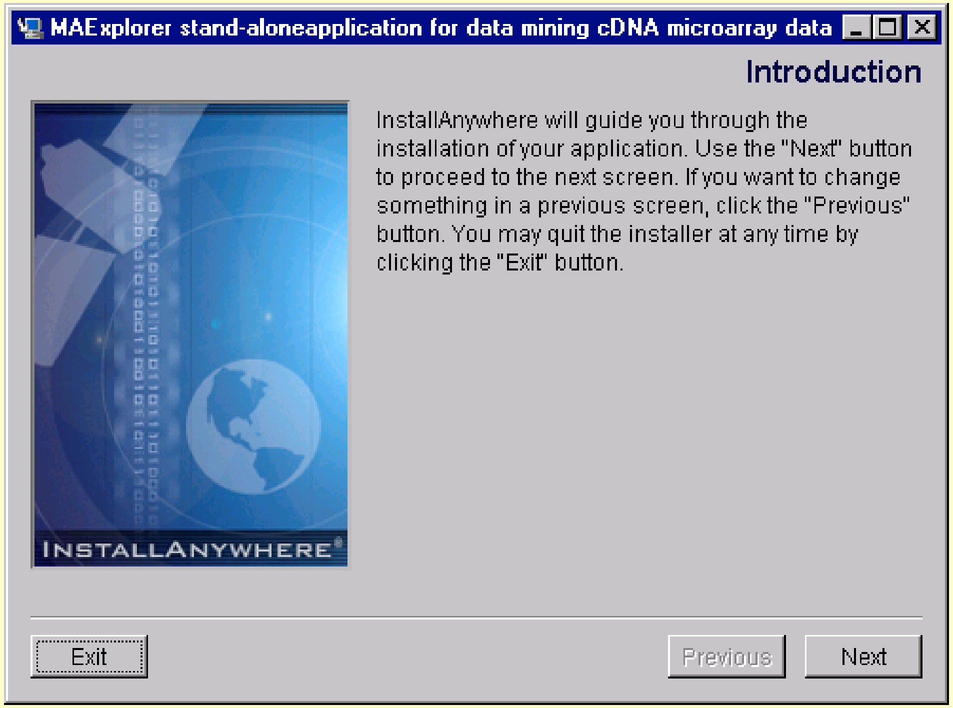


Address: F:\Temp

Name	T.	Size	Modified
installMae.exe	A..	5,353...	11/14/00 2:40 PM

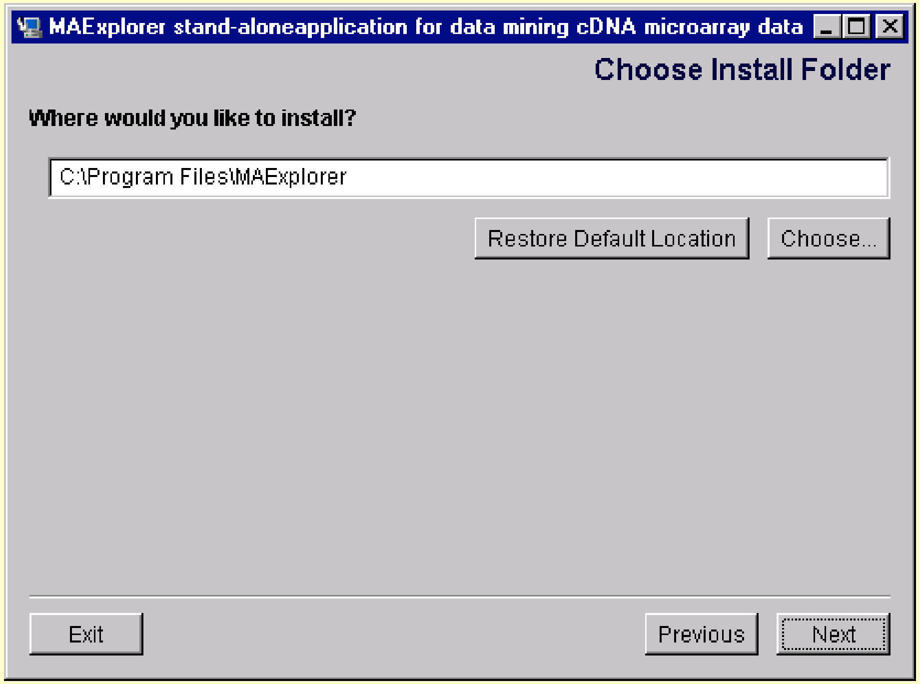
1 object(s) selected

III.5 Sucessive steps during installation of MAExplorer - press "Next"



Introduction

InstallAnywhere will guide you through the installation of your application. Use the "Next" button to proceed to the next screen. If you want to change something in a previous screen, click the "Previous" button. You may quit the installer at any time by clicking the "Exit" button.

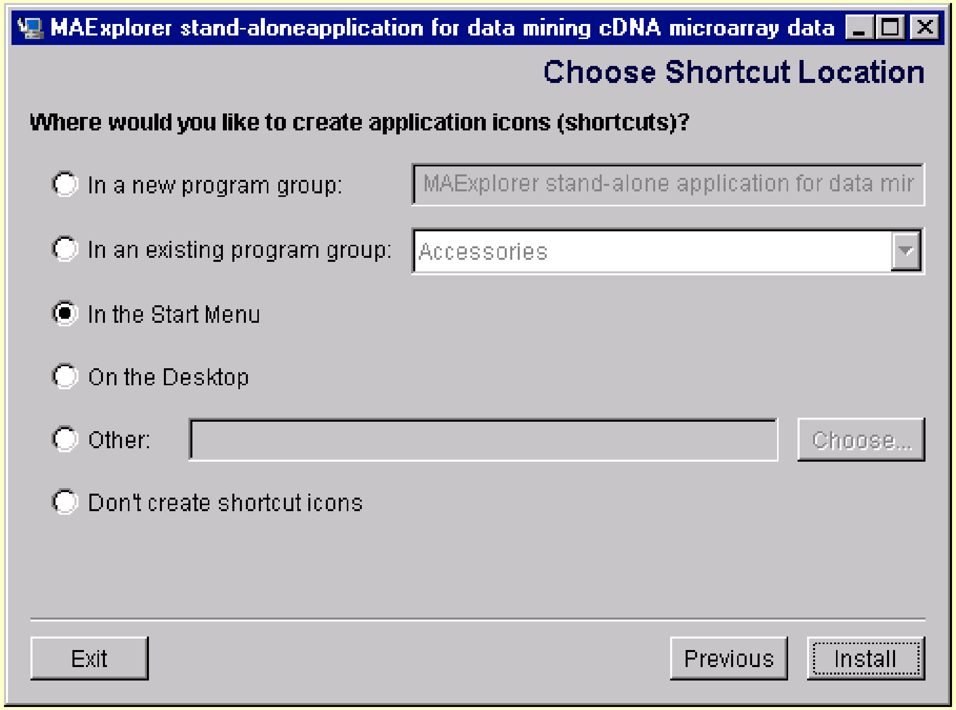


Choose Install Folder

Where would you like to install?

C:\Program Files\MAExplorer

III.6 Finish installation of MAExplorer: A) press “Install”, B) press “Done”



MAExplorer stand-aloneapplication for data mining cDNA microarray data

Choose Shortcut Location

Where would you like to create application icons (shortcuts)?

○ In a new program group: MAExplorer stand-alone application for data mir

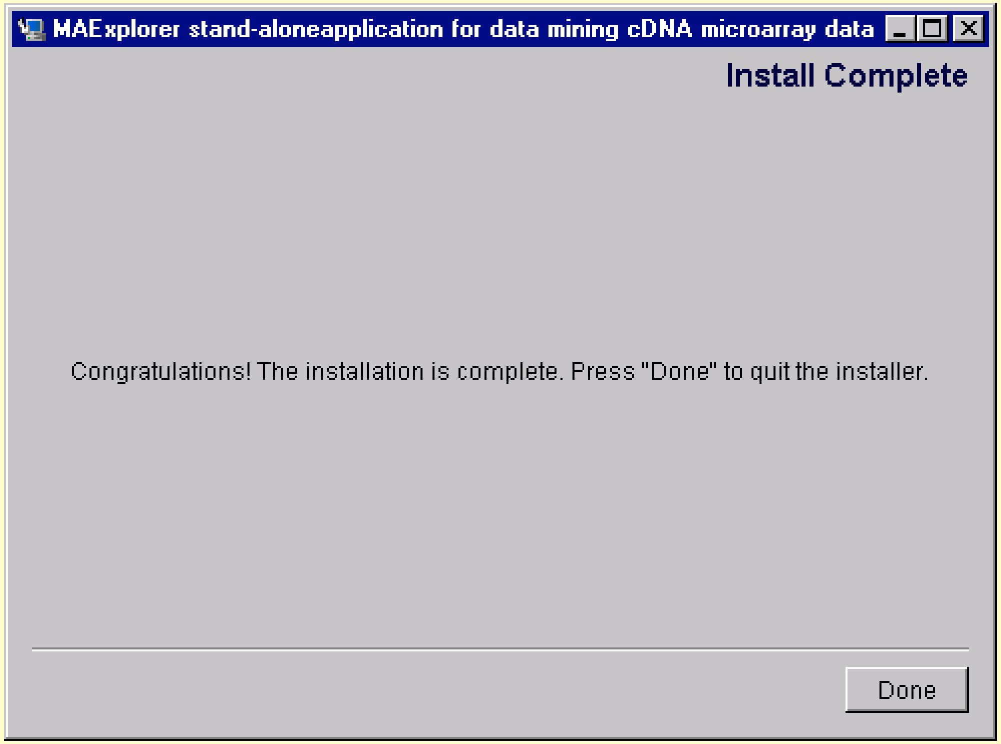
○ In an existing program group: Accessories

● In the Start Menu

○ On the Desktop

○ Other:

○ Don't create shortcut icons

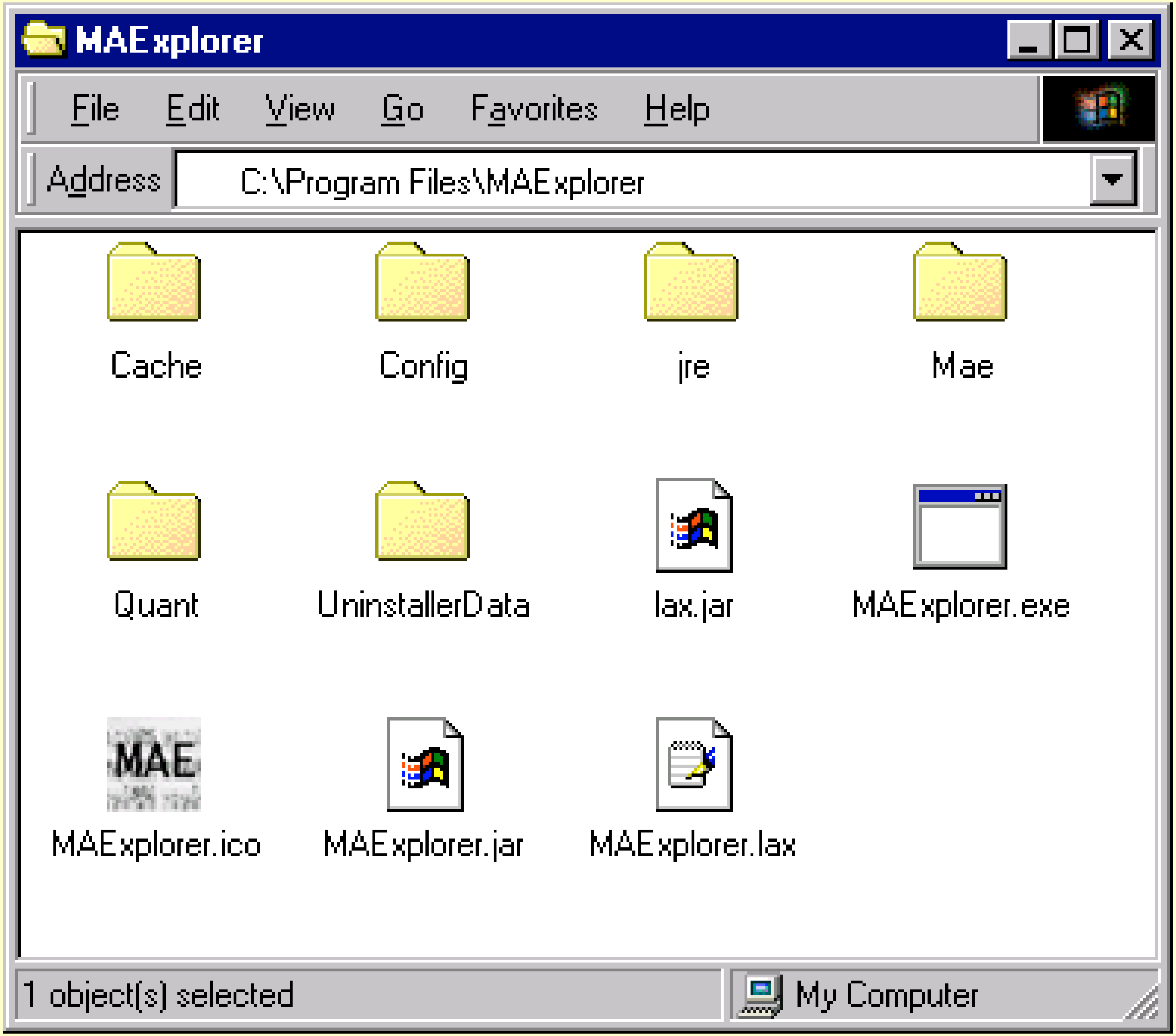


MAExplorer stand-aloneapplication for data mining cDNA microarray data

Install Complete

Congratulations! The installation is complete. Press "Done" to quit the installer.

III.7 Directory structure of downloaded files



C:\Program Files\MAExplorer

Cache

Config

jre

Mae

Quant

UninstallerData

lax.jar

MAExplorer.exe

MAExplorer.ico

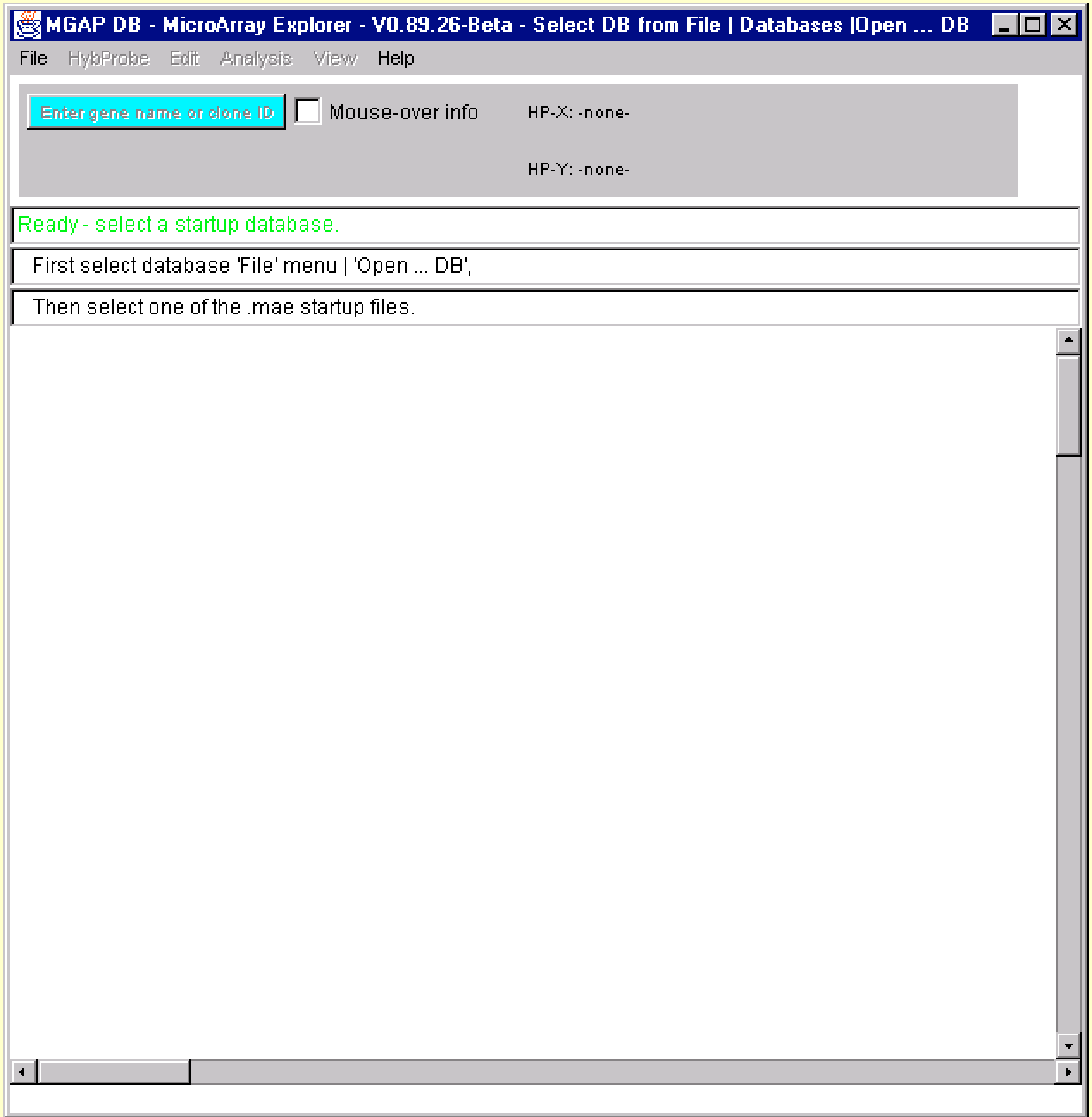
MAExplorer.jar

MAExplorer.lax

1 object(s) selected

My Computer

III.8 Start MAExplorer from Windows PC "Start" menu. Initially starts with empty database



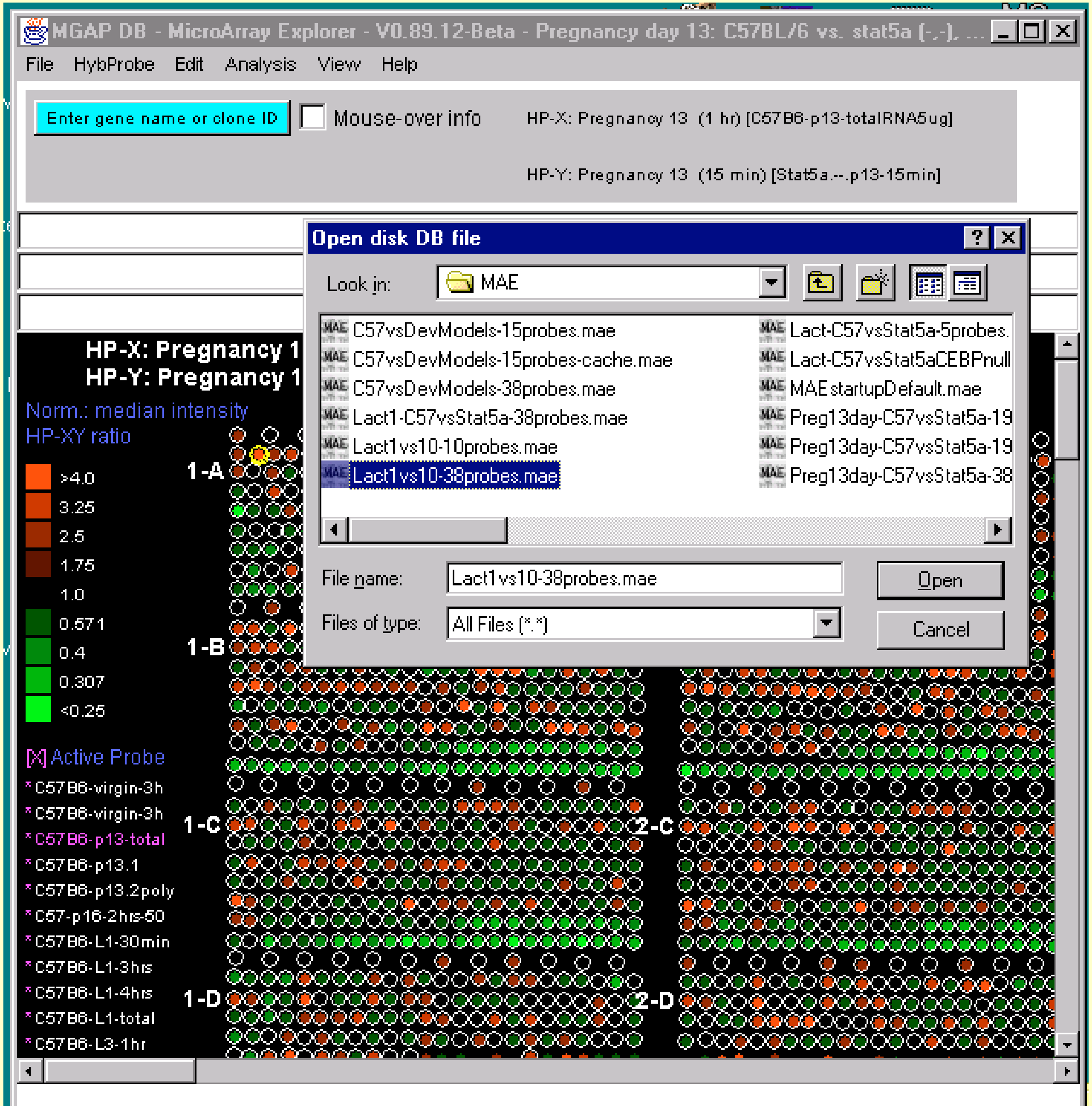
Ready - select a startup database.

First select database 'File' menu | 'Open ... DB',

Then select one of the .mae startup files.

III.9 Open demo (MGAP) database from local disk

- Browse demo project for startup database. Select File menu, then Open file DB



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

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

HP-X: Pregnancy 1

HP-Y: Pregnancy 1

Norm.: median intensity

HP-XY ratio

🟥	>4.0
🟥	3.25
🟥	2.5
🟥	1.75
	1.0
🟩	0.571
🟩	0.4
🟩	0.307
🟩	<0.25

[X] Active Probe

- C57B6-virgin-3h
- C57B6-virgin-3h
- C57B6-p13-total
- C57B6-p13.1
- C57B6-p13.2poly
- C57-p16-2hrs-50
- C57B6-L1-30min
- C57B6-L1-3hrs
- C57B6-L1-4hrs
- C57B6-L1-total
- C57B6-L3-1hr

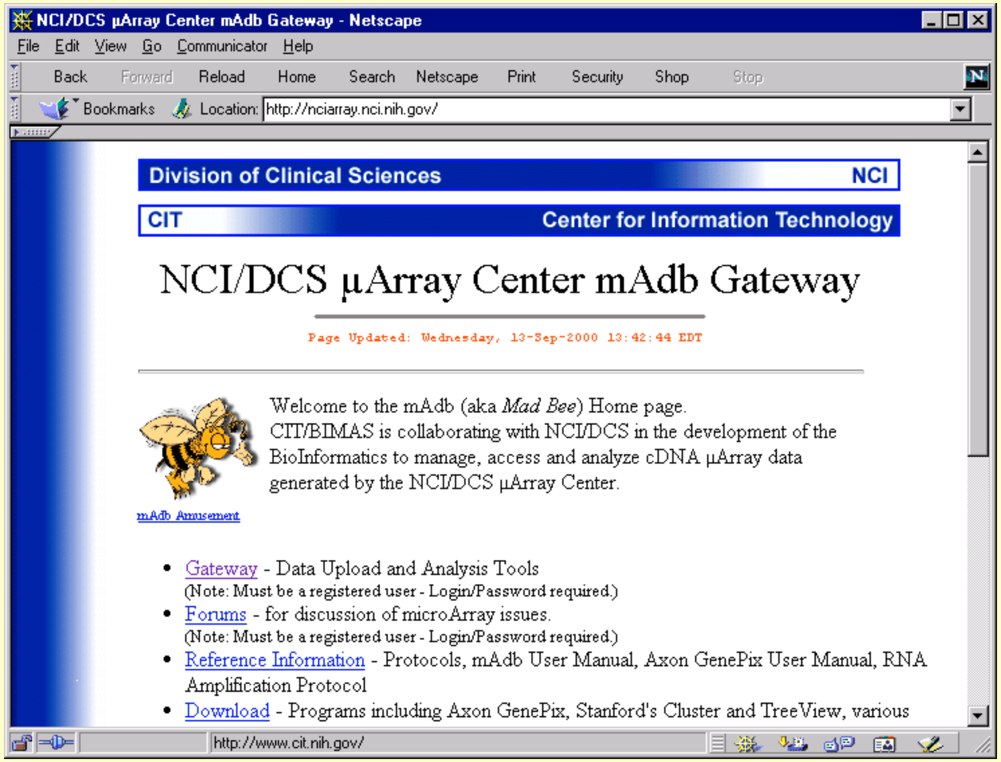
Look in: MAE

C57vsDevModels-15probes.mae
C57vsDevModels-15probes-cache.mae
C57vsDevModels-38probes.mae
Lact1-C57vsStat5a-38probes.mae
Lact1vs10-10probes.mae
Lact1vs10-38probes.mae
Lact-C57vsStat5a-5probes.
Lact-C57vsStat5aCEBPnull
MAEstartupDefault.mae
Preg13day-C57vsStat5a-19
Preg13day-C57vsStat5a-19
Preg13day-C57vsStat5a-38

File name: Lact1vs10-38probes.mae

Files of type: All Files (*.*)

IV. Using NCI/CIT mAdb data with MicroArray Explorer



http://nciarray.nci.nih.gov/

Division of Clinical Sciences NCI

CIT Center for Information Technology

NCI/DCS µArray Center mAdb Gateway

Page Updated: Wednesday, 13-Sep-2000 13:42:44 EDT

Welcome to the mAdb (aka *Mad Bee*) Home page.
CIT/BIMAS is collaborating with NCI/DCS in the development of the BioInformatics to manage, access and analyze cDNA µArray data generated by the NCI/DCS µArray Center.

mAdb Amusement

- Gateway - Data Upload and Analysis Tools
  (Note: Must be a registered user - Login/Password required.)
- Forums - for discussion of microArray issues.
  (Note: Must be a registered user - Login/Password required.)
- Reference Information - Protocols, mAdb User Manual, Axon GenePix User Manual, RNA Amplification Protocol
- Download - Programs including Axon GenePix, Stanford's Cluster and TreeView, various

http://www.cit.nih.gov/

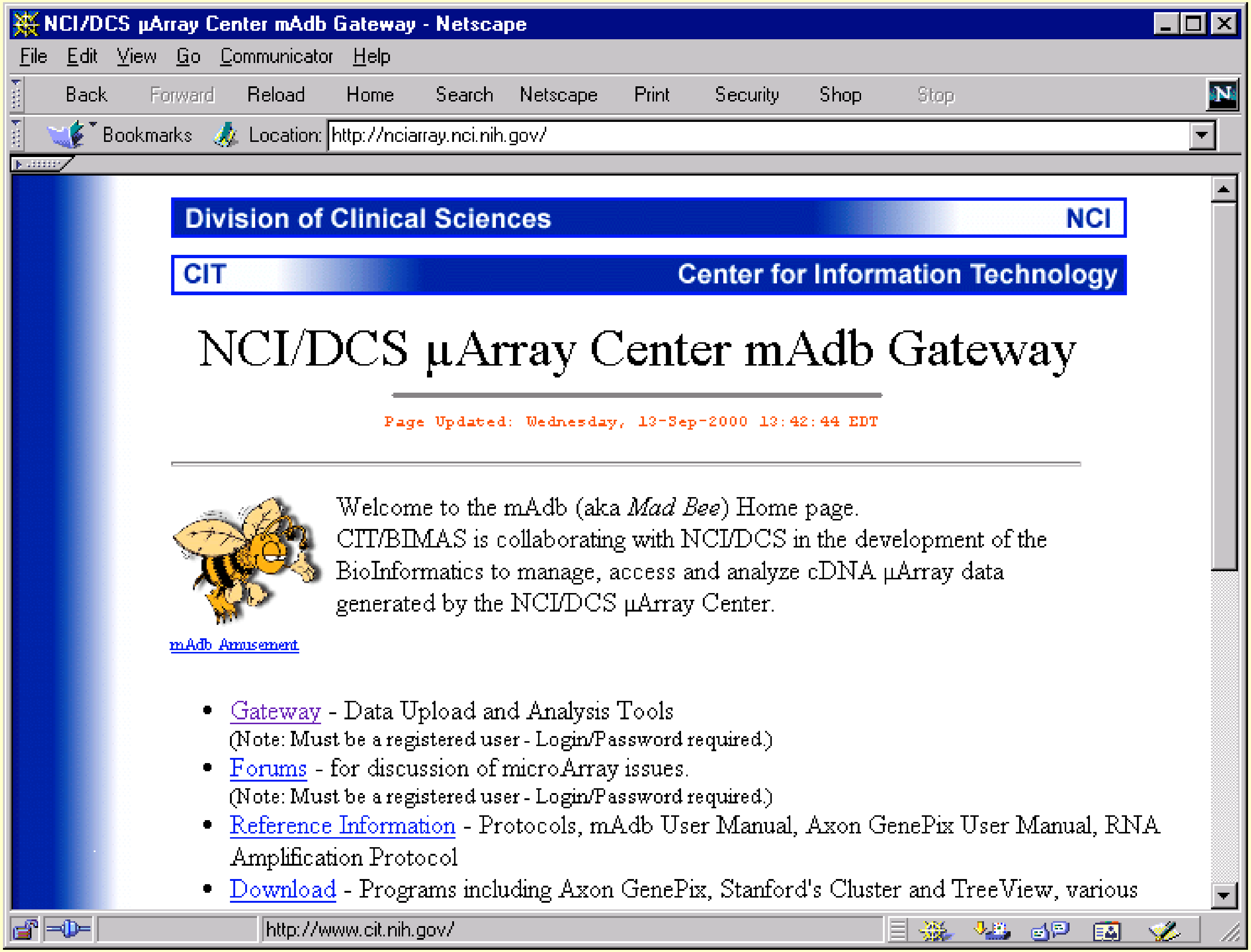
Outline

1. Log into mAdb
2. Select your data
3. Export it as a Zip file to your computer
4. Unpack the Zip file
5. Click on the MAE Start.mae

IV. Procedure to use MAExplorer on mAdb data

- 1. Install MAExplorer if not already installed (see previous Procedure 1).
- 2. Go to **http://nciarray.nci.nih.gov/** with your Web browser
- 3. Go to "**Gateway**"
- 4. Go to "**Tools**"
- 5. Select the set of projects to be exported from the scrollable list.
- 6. Select "**BETA formated array data retrieval tool**".
- 7. Select "**LECB/NCI MAExplorer**" for the "Retrieval format".
- 8. **Submit**. This will eventually replace the Web page with a new page containing a numbered (number related to date and time of day) file ending in **.zip**. The file will be purged after a while, so it should not be treated as a permanent link.
- 9. Click on the **.zip** file and save it locally to your disk.
- 10. Unpack the **.zip** file to a new directory, for example "*myData*"
- 11. On Windows systems, double click on Start.mae in the *myData*\MAE\ directory. This will start up MAExplorer.

IV.1 NCI/CIT mAdb Web server home page http://nciarray.nci.nih.gov/



Division of Clinical Sciences **NCI**

CIT **Center for Information Technology**

NCI/DCS µArray Center mAdb Gateway

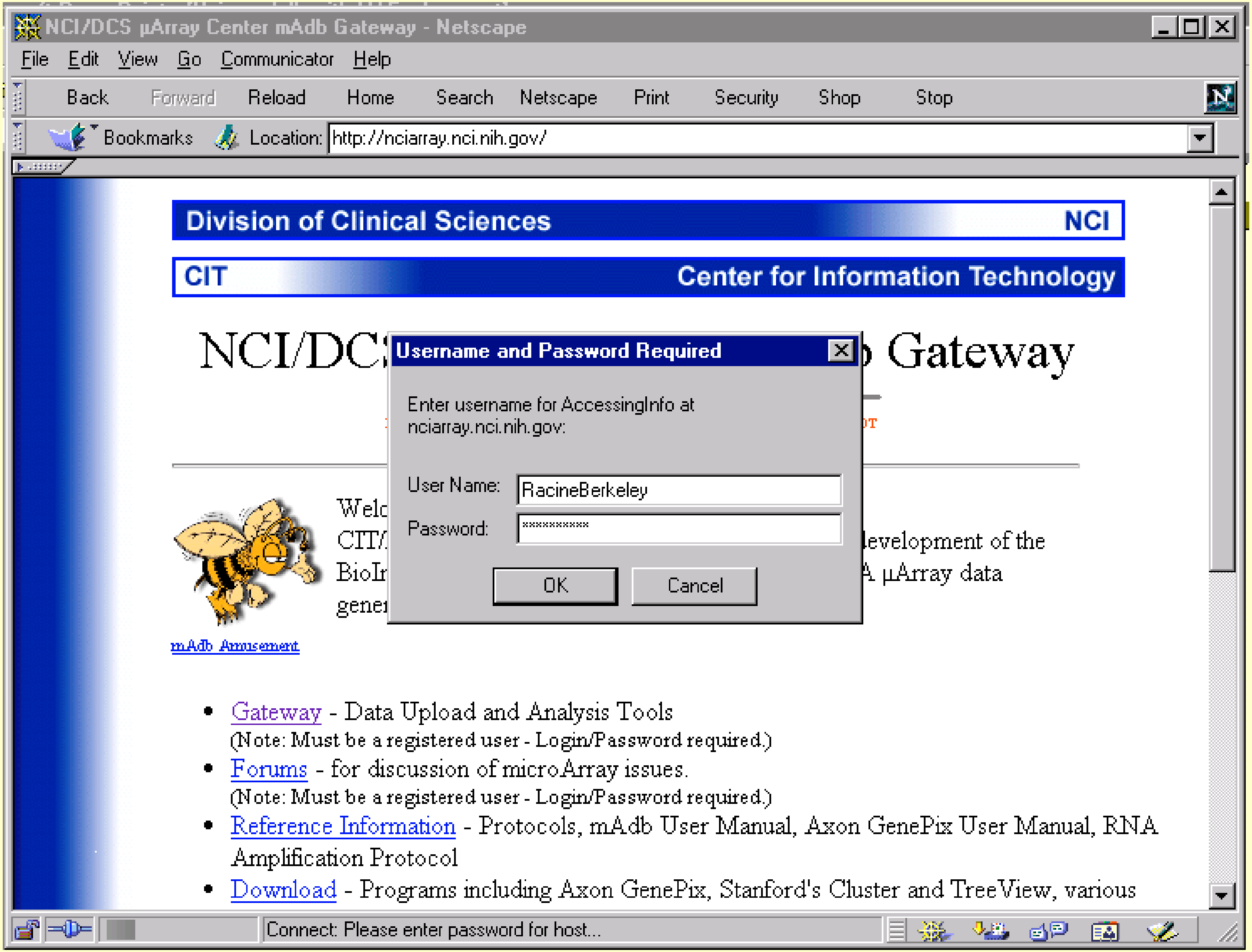
Page Updated: Wednesday, 13-Sep-2000 13:42:44 EDT

Welcome to the mAdb (aka *Mad Bee*) Home page.
CIT/BIMAS is collaborating with NCI/DCS in the development of the BioInformatics to manage, access and analyze cDNA µArray data generated by the NCI/DCS µArray Center.

mAdb Amusement

- Gateway - Data Upload and Analysis Tools
  (Note: Must be a registered user - Login/Password required.)
- Forums - for discussion of microArray issues.
  (Note: Must be a registered user - Login/Password required.)
- Reference Information - Protocols, mAdb User Manual, Axon GenePix User Manual, RNA Amplification Protocol
- Download - Programs including Axon GenePix, Stanford's Cluster and TreeView, various

IV.2 Press “Gateway” & Log on to mAdb server



http://nciarray.nci.nih.gov/

NCI/DC… Gateway

Username and Password Required

Enter username for AccessingInfo at nciarray.nci.nih.gov:

User Name: RacineBerkeley

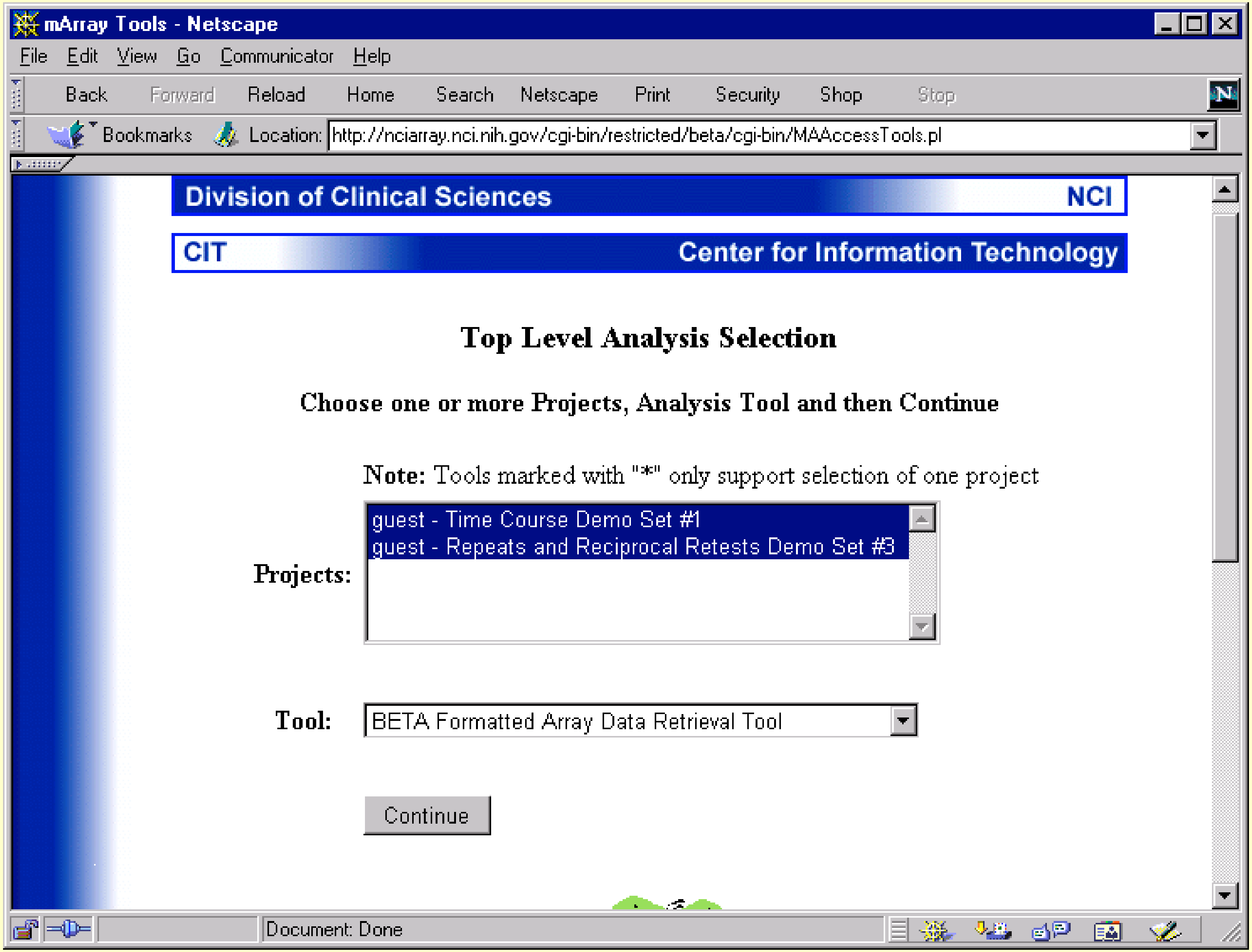
Password: ***********

Welc…
CIT/… …evelopment of the
BioIr… …A µArray data
gener…

- Gateway - Data Upload and Analysis Tools
  (Note: Must be a registered user - Login/Password required.)
- Forums - for discussion of microArray issues.
  (Note: Must be a registered user - Login/Password required.)
- Reference Information - Protocols, mAdb User Manual, Axon GenePix User Manual, RNA Amplification Protocol
- Download - Programs including Axon GenePix, Stanford's Cluster and TreeView, various

Connect: Please enter password for host...

IV.3 Select: a) Projects, b) “Formated Array data Retrieval Tool”, c) then press “Continue”



Top Level Analysis Selection

Choose one or more Projects, Analysis Tool and then Continue

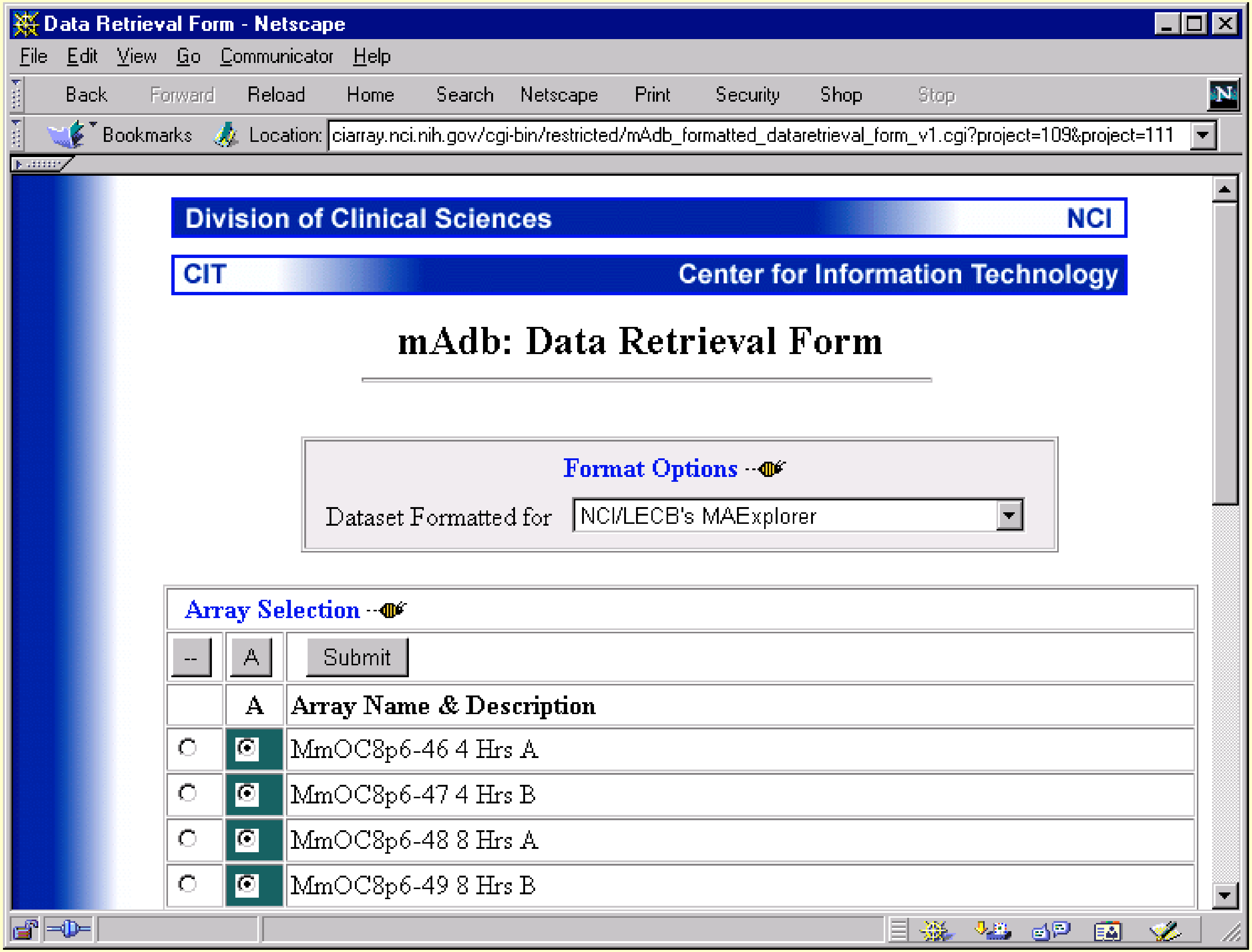
Note: Tools marked with "*" only support selection of one project

Projects:
- guest - Time Course Demo Set #1
- guest - Repeats and Reciprocal Retests Demo Set #3

Tool: BETA Formatted Array Data Retrieval Tool

Continue

IV.4 Set a) Format option to “MAExplorer”, b) select arrays to be analyzed, c) press “Submit”



mAdb: Data Retrieval Form

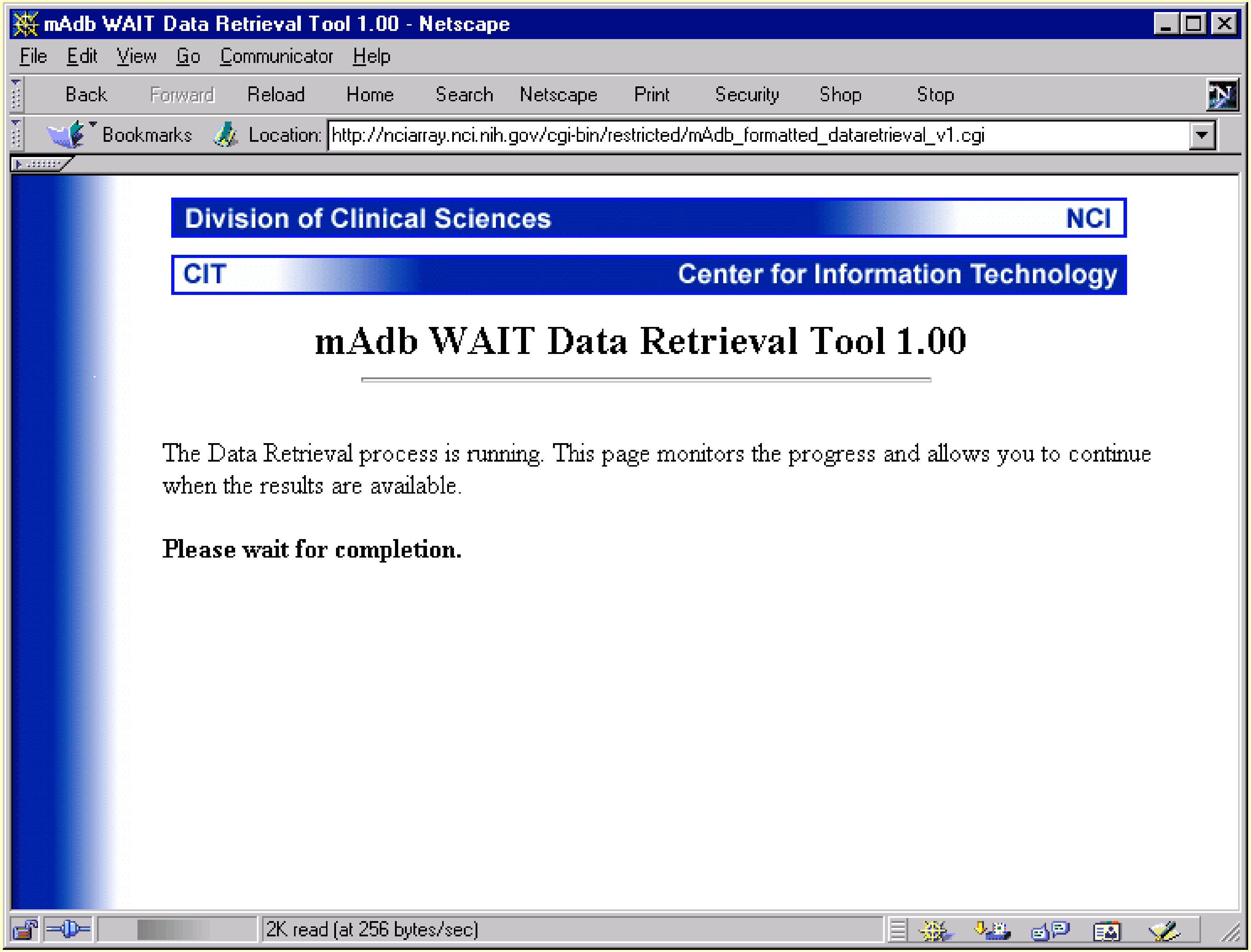
Format Options

Dataset Formatted for: NCI/LECB's MAExplorer

Array Selection

	A	Array Name & Description
○	●	MmOC8p6-46 4 Hrs A
○	●	MmOC8p6-47 4 Hrs B
○	●	MmOC8p6-48 8 Hrs A
○	●	MmOC8p6-49 8 Hrs B

IV.5 It will contact the mAdb server to get data



http://nciarray.nci.nih.gov/cgi-bin/restricted/mAdb_formatted_dataretrieval_v1.cgi

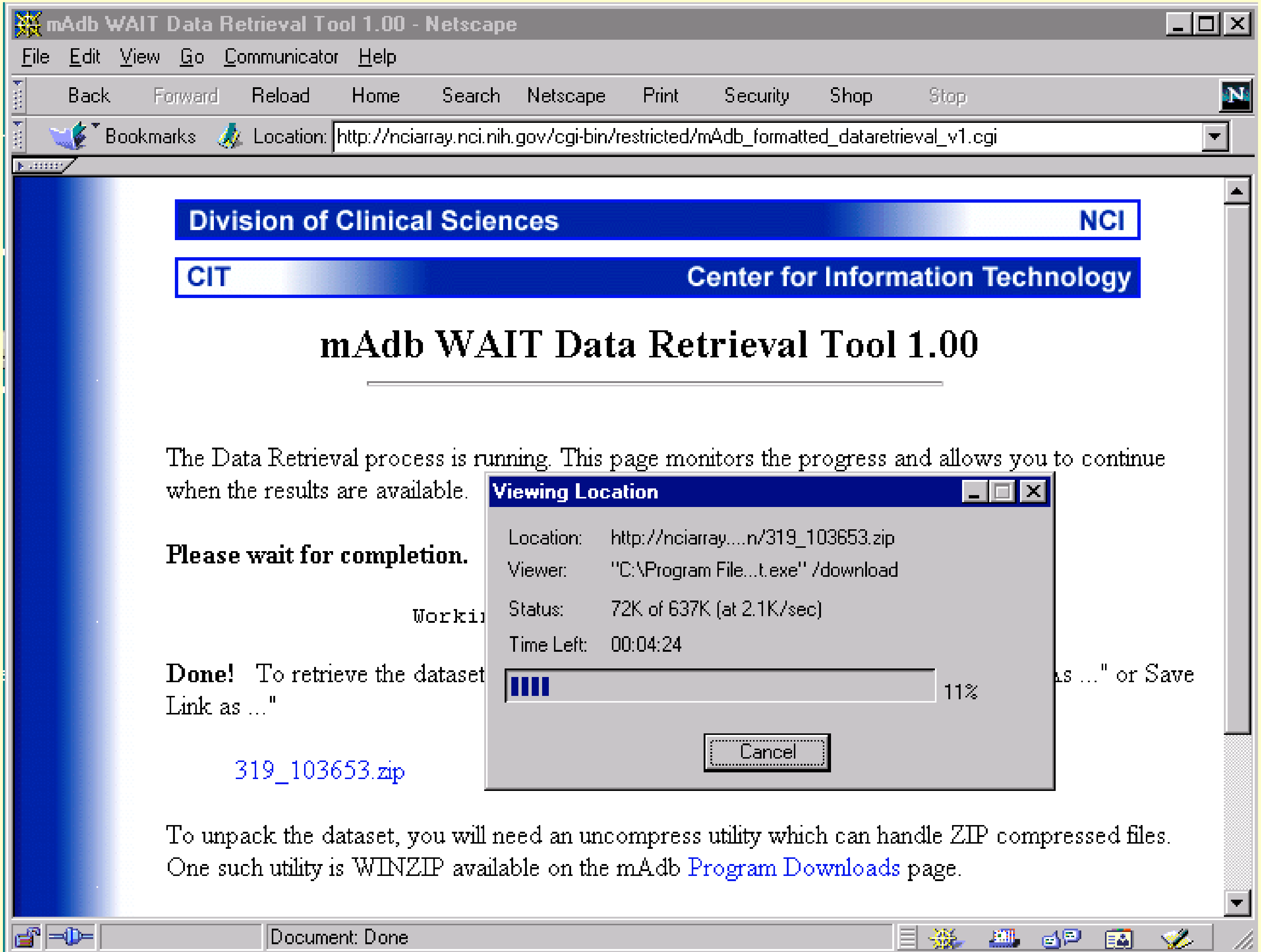
mAdb WAIT Data Retrieval Tool 1.00

The Data Retrieval process is running. This page monitors the progress and allows you to continue when the results are available.

Please wait for completion.

2K read (at 256 bytes/sec)

IV.6 Click on Zip file (e.g. 319-103653.zip) result to download to your computer.



mAdb WAIT Data Retrieval Tool 1.00

The Data Retrieval process is running. This page monitors the progress and allows you to continue when the results are available.

Please wait for completion.

Worki

Done! To retrieve the dataset …s ..." or Save Link as ..."

319_103653.zip

To unpack the dataset, you will need an uncompress utility which can handle ZIP compressed files. One such utility is WINZIP available on the mAdb Program Downloads page.

Viewing Location

Location: http://nciarray....n/319_103653.zip

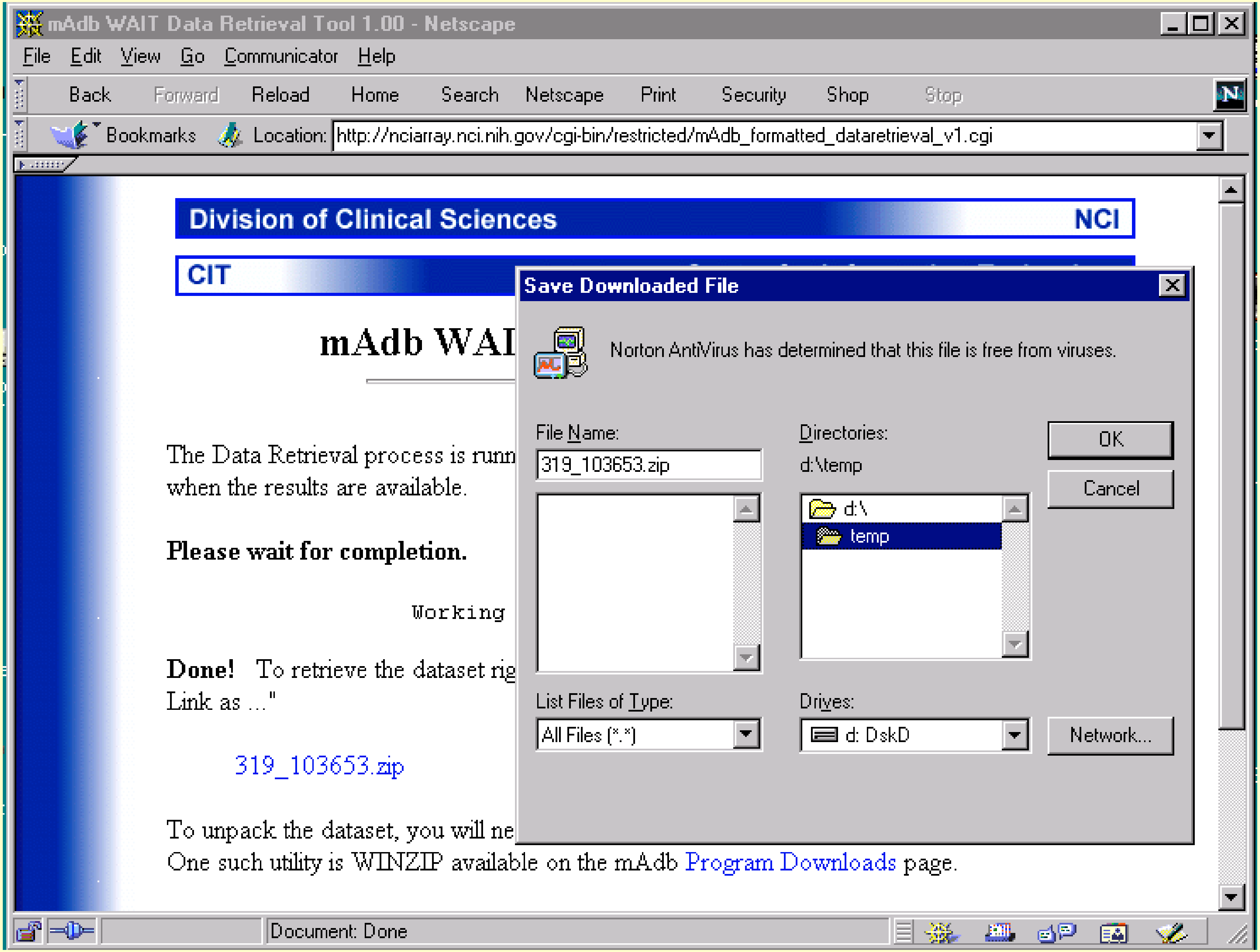
Viewer: "C:\Program File...t.exe" /download

Status: 72K of 637K (at 2.1K/sec)

Time Left: 00:04:24

11%

IV.7 Save the Zip data file on your local disk



Location: http://nciarray.nci.nih.gov/cgi-bin/restricted/mAdb_formatted_dataretrieval_v1.cgi

mAdb WAI

The Data Retrieval process is runn
when the results are available.

Please wait for completion.

Working

Done! To retrieve the dataset rig
Link as ..."

319_103653.zip

To unpack the dataset, you will ne
One such utility is WINZIP available on the mAdb Program Downloads page.

Norton AntiVirus has determined that this file is free from viruses.

File Name: 319_103653.zip

Directories: d:\temp

d:\
temp

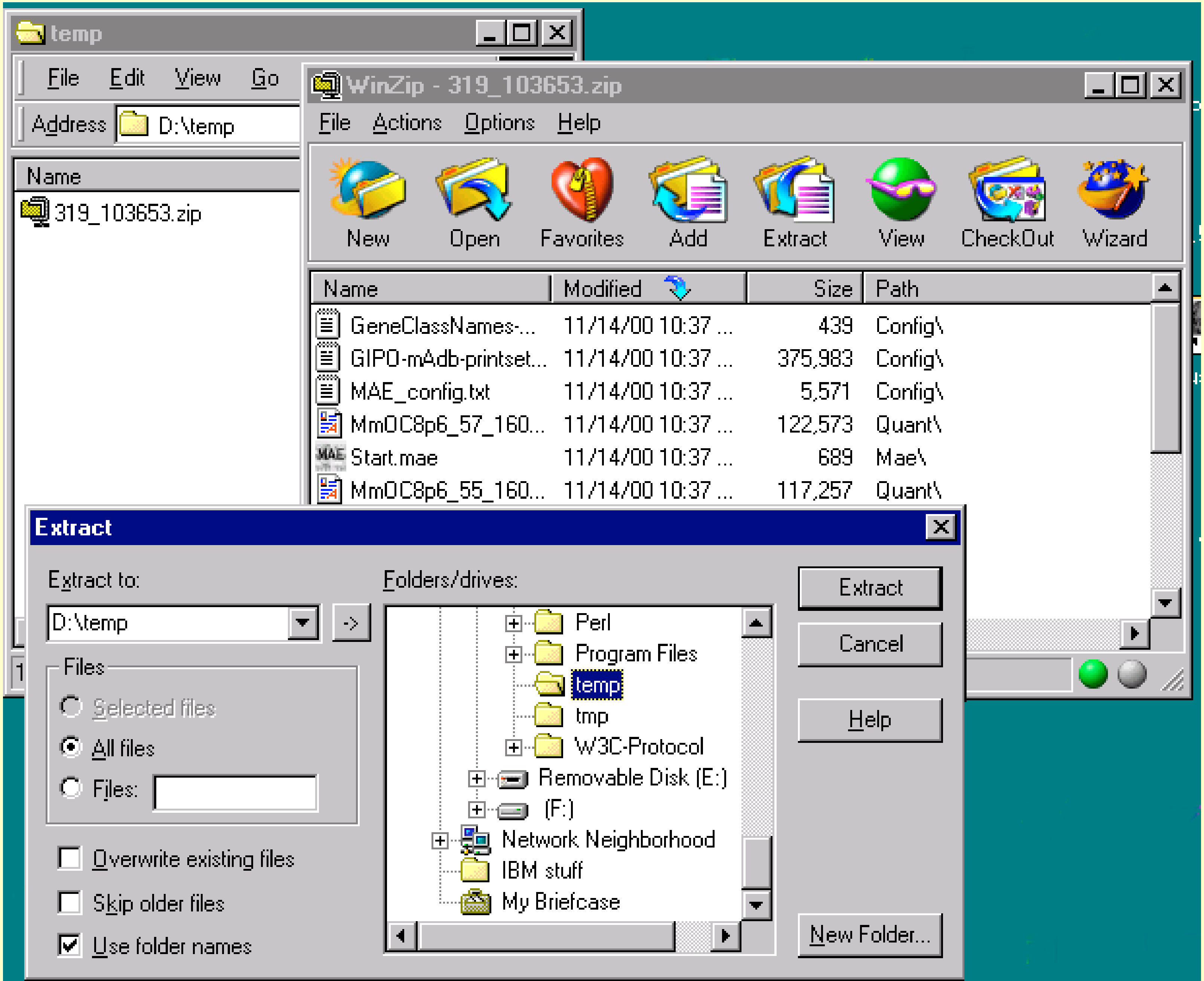
List Files of Type: All Files (*.*)

Drives: d: DskD

Document: Done

IV.8 Unzipping the Zip data file

- (WinZip is available from the mAdb download Web site)



Address: D:\temp

Name
319_103653.zip

Name	Modified	Size	Path
GeneClassNames-...	11/14/00 10:37 ...	439	Config\
GIPO-mAdb-printset...	11/14/00 10:37 ...	375,983	Config\
MAE_config.txt	11/14/00 10:37 ...	5,571	Config\
MmOC8p6_57_160...	11/14/00 10:37 ...	122,573	Quant\
Start.mae	11/14/00 10:37 ...	689	Mae\
MmOC8p6_55_160...	11/14/00 10:37 ...	117,257	Quant\

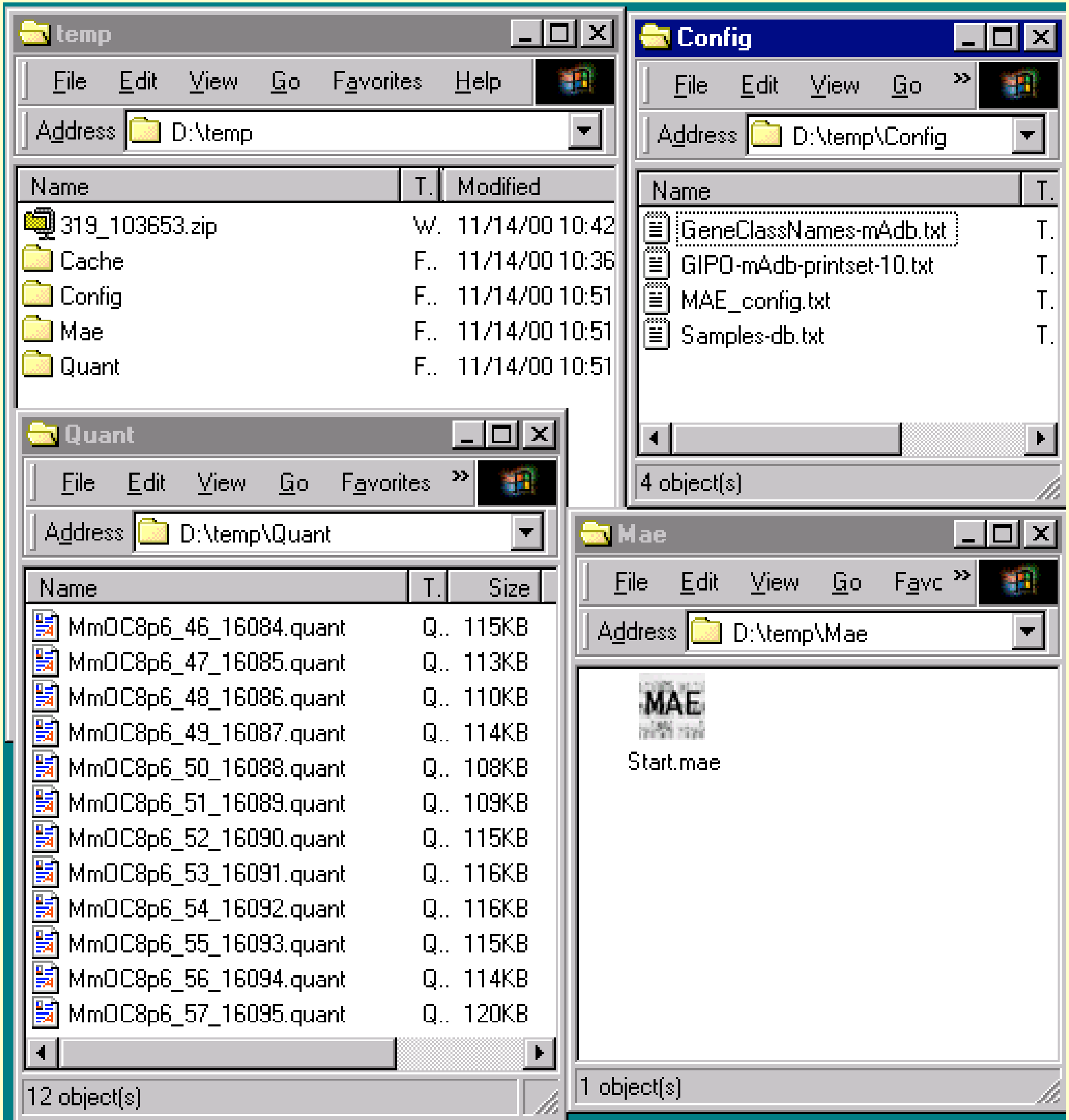
Extract to: D:\temp

Files
○ Selected files
● All files
○ Files:

☐ Overwrite existing files
☐ Skip older files
☑ Use folder names

Folders/drives:
- Perl
- Program Files
- temp
- tmp
- W3C-Protocol
- Removable Disk (E:)
- (F:)
- Network Neighborhood
- IBM stuff
- My Briefcase

IV.9 Inspecting the unzipped data files



Address D:\temp

Name	T.	Modified
319_103653.zip	W.	11/14/00 10:42
Cache	F..	11/14/00 10:36
Config	F..	11/14/00 10:51
Mae	F..	11/14/00 10:51
Quant	F..	11/14/00 10:51

Address D:\temp\Config

Name	T.
GeneClassNames-mAdb.txt	T.
GIPO-mAdb-printset-10.txt	T.
MAE_config.txt	T.
Samples-db.txt	T.

4 object(s)

Address D:\temp\Quant

Name	T.	Size
MmOC8p6_46_16084.quant	Q..	115KB
MmOC8p6_47_16085.quant	Q..	113KB
MmOC8p6_48_16086.quant	Q..	110KB
MmOC8p6_49_16087.quant	Q..	114KB
MmOC8p6_50_16088.quant	Q..	108KB
MmOC8p6_51_16089.quant	Q..	109KB
MmOC8p6_52_16090.quant	Q..	115KB
MmOC8p6_53_16091.quant	Q..	116KB
MmOC8p6_54_16092.quant	Q..	116KB
MmOC8p6_55_16093.quant	Q..	115KB
MmOC8p6_56_16094.quant	Q..	114KB
MmOC8p6_57_16095.quant	Q..	120KB

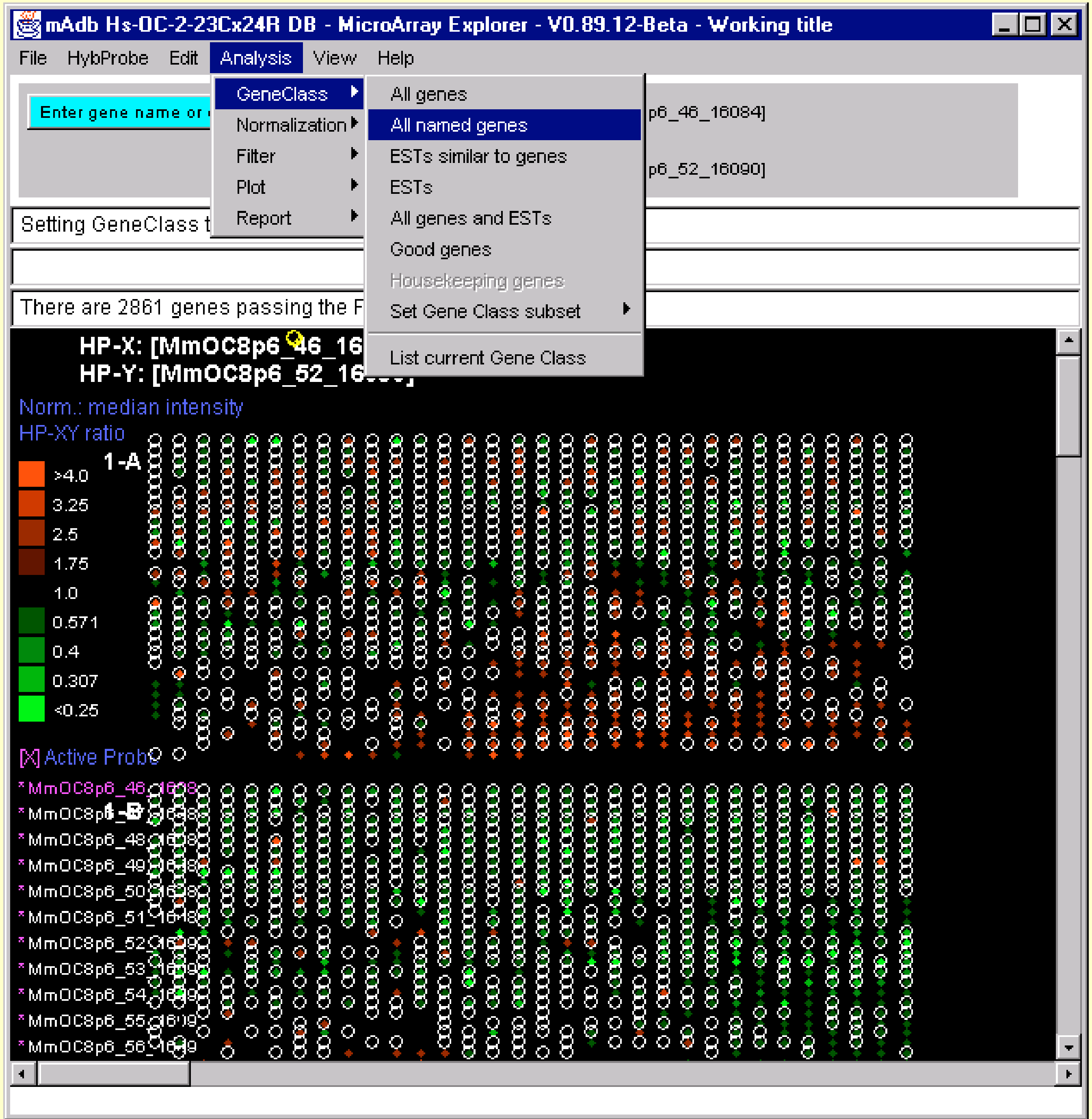
12 object(s)

Address D:\temp\Mae

Start.mae

1 object(s)

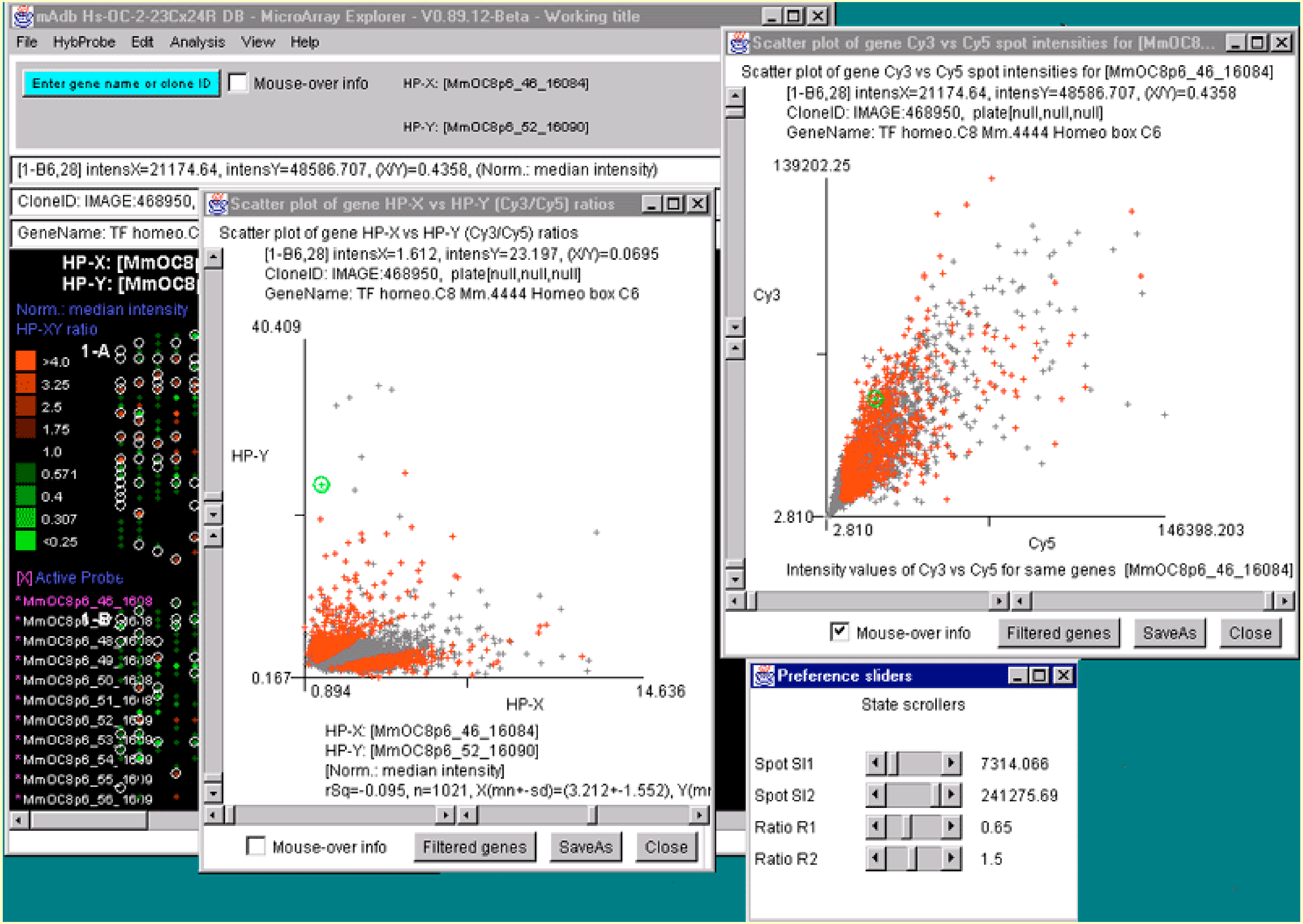
IV.10 Click " MAE Start.mae" to start MAExplorer



Setting GeneClass t

There are 2861 genes passing the F

IV.11 Explore data using data filters, plots, etc.



mAdb Hs-OC-2-23Cx24R DB - MicroArray Explorer - V0.89.12-Beta - Working title

HP-X: [MmOC8p6_46_16084]

HP-Y: [MmOC8p6_52_16090]

[1-B6,28] intensX=21174.64, intensY=48586.707, (X/Y)=0.4358, (Norm.: median intensity)

CloneID: IMAGE:468950,

GeneName: TF homeo.C

Norm.: median intensity

HP-XY ratio

[X] Active Probe

Scatter plot of gene HP-X vs HP-Y (Cy3/Cy5) ratios

Scatter plot of gene HP-X vs HP-Y (Cy3/Cy5) ratios
[1-B6,28] intensX=1.612, intensY=23.197, (X/Y)=0.0695
CloneID: IMAGE:468950, plate[null,null,null]
GeneName: TF homeo.C8 Mm.4444 Homeo box C6

HP-X: [MmOC8p6_46_16084]
HP-Y: [MmOC8p6_52_16090]
[Norm.: median intensity]
rSq=-0.095, n=1021, X(mn+-sd)=(3.212+-1.552), Y(m

Scatter plot of gene Cy3 vs Cy5 spot intensities for [MmOC8...

Scatter plot of gene Cy3 vs Cy5 spot intensities for [MmOC8p6_46_16084]
[1-B6,28] intensX=21174.64, intensY=48586.707, (X/Y)=0.4358
CloneID: IMAGE:468950, plate[null,null,null]
GeneName: TF homeo.C8 Mm.4444 Homeo box C6

Intensity values of Cy3 vs Cy5 for same genes [MmOC8p6_46_16084]

Preference sliders

State scrollers

Spot SI1 7314.066
Spot SI2 241275.69
Ratio R1 0.65
Ratio R2 1.5

Summary of Downloading a mAdb data set

- This procedure downloads one or more projects into a directory on your local computer.

- At this point, data mining may proceed using MAExplorer independent of the Internet connection to mAdb.

- If you want to add additional hybridized samples, you should download all of the samples again (this will be resolved in the future). Currently, you can’t easily merge data from several downloaded data sets.