

## MeV+R Quick Tutorial

### 1. Installing R and Rserve

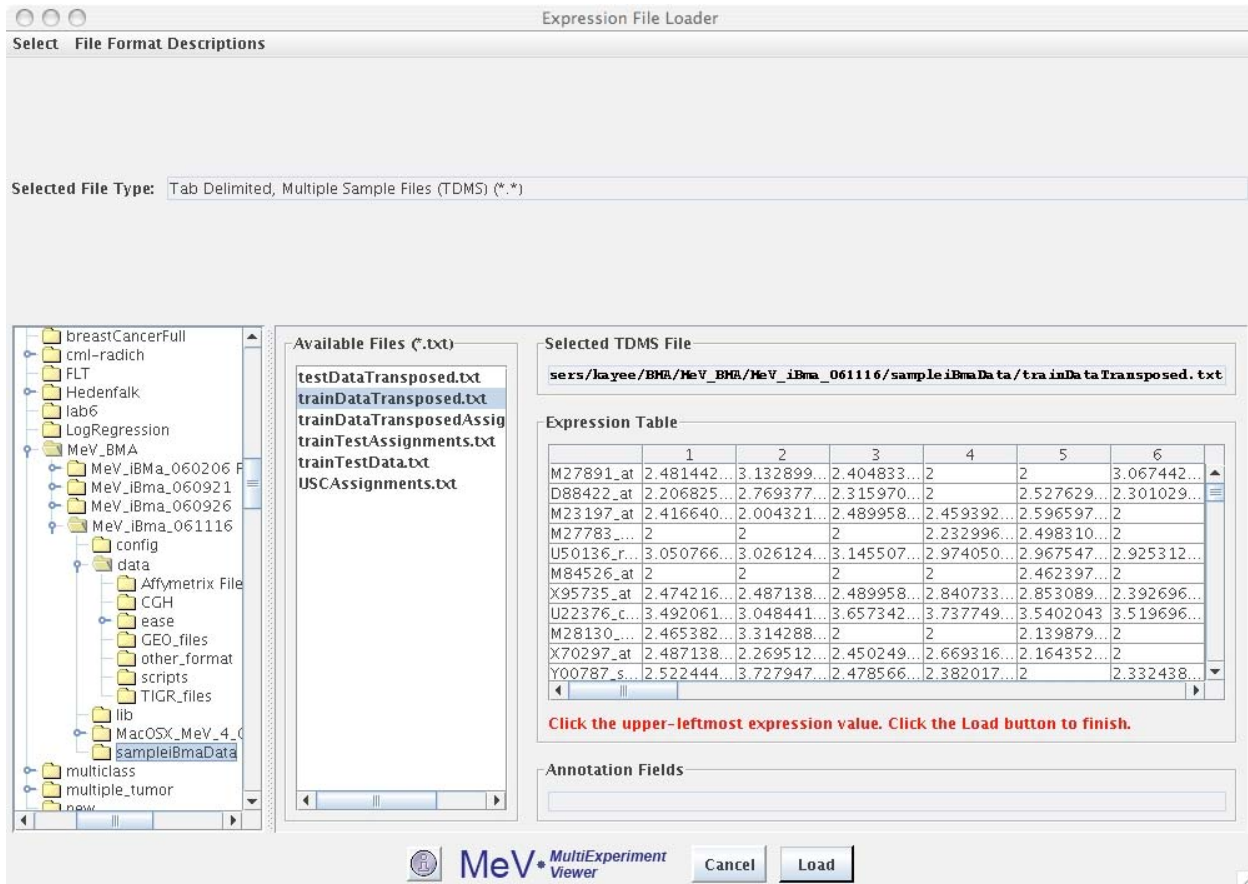
Please see section 19 in the MeV User Manual for details. The manual refers to R 2.4. The latest version of R can be installed in a similar manner. Essentially, the user needs to install R, Rserve and the bioconductor packages (Rama, Bridge and IterativeBMA) in R.

### 2. Load Data

MeV can interpret files of several types, including the MultiExperiment Viewer format (.mev), the TIGR ArrayViewer format (.tav), the TDMS file format (Tab Delimited, Multiple Sample format), the Affymetrix file format, and GenePix file format (.gpr). Please see section 4 in the MeV User Manual for details.

Our test data sets are in the tab-delimited format. Here is a brief summary of loading our test data (from Section 4.3 in the MeV User manual):

Select the Tab Delimited, Multiple Sample Files (TDMS) (\*.\*) option from the drop-down menu to load text files. Navigate to the folder containing the file and select the desired file. The file will be displayed in a tabular format. Click the cell in the table which contains the upper-leftmost **expression value** in the file. Click *Load*.



### 3. Start Rserve

**Running under OS X:** Open a terminal instance. Type *R CMD Rserve*

**Running under Windows:** Double click *Rserve.exe*

### 4. Run Rama

RAMA uses a Bayesian hierarchical model for the robust estimation of cDNA microarray intensities with replicates. This is highly relevant for replicated microarray experiments because even one outlying replicate (such as due to scratches or dust) can have a disastrous effect on the estimated signal intensity. Our model borrows strength from all the genes to decide if a measurement is an outlier, and hence it is superior in detecting outliers based on a small number of replicate measurements compared to other classical robust estimators. The case studies in the full manuscript and our previous work [1] showed that RAMA produces superior results compared to other methods of estimating gene intensities from replicated microarray data.

For the user interface, please see section 5.2 in the MeV User Manual.

### 5. Run Bridge

BRIDGE fits a robust Bayesian hierarchical model to test for differentially expressed genes on microarray data. It can be used with both two-color microarrays and single-channel Affymetrix chips. BRIDGE is powerful even with a small number of samples (either biological or technical replicates) under each experimental condition. Our case studies in the full manuscript and our previous work [2] showed that BRIDGE produces better results than other MeV tools (one-sample t-test and SAM).

For the user interface, please see section 11.28 in the MeV User Manual.

#### 6. Run iterativeBMA

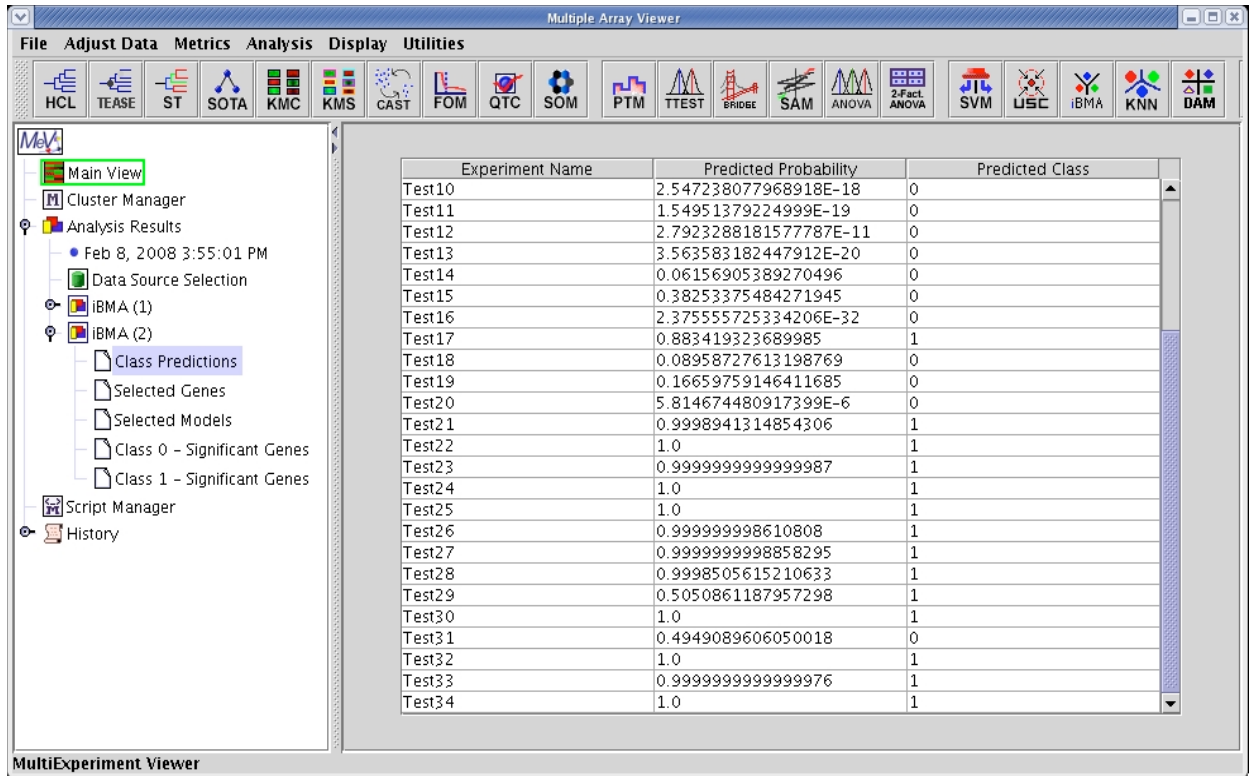
The iterativeBMA algorithm is a multivariate technique for gene selection and classification of microarray data. Bayesian Model Averaging (BMA) takes model uncertainty into consideration by averaging over the posterior distributions of predicted probabilities based on multiple models, weighted by their posterior model probabilities. In the training data, the classes (or labels) of the samples are used to select a small number of genes to predict the classes (or labels) of the samples in the test set.

After loading the data, the user is asked to label the two classes. The default labels for the two classes are 0 and 1 respectively. In the same dialog box, the user is asked to establish a Rserve connection. The user is also given the option of specifying advanced parameters for the analysis. The advanced parameters are documented in the iterativeBMA vignette.

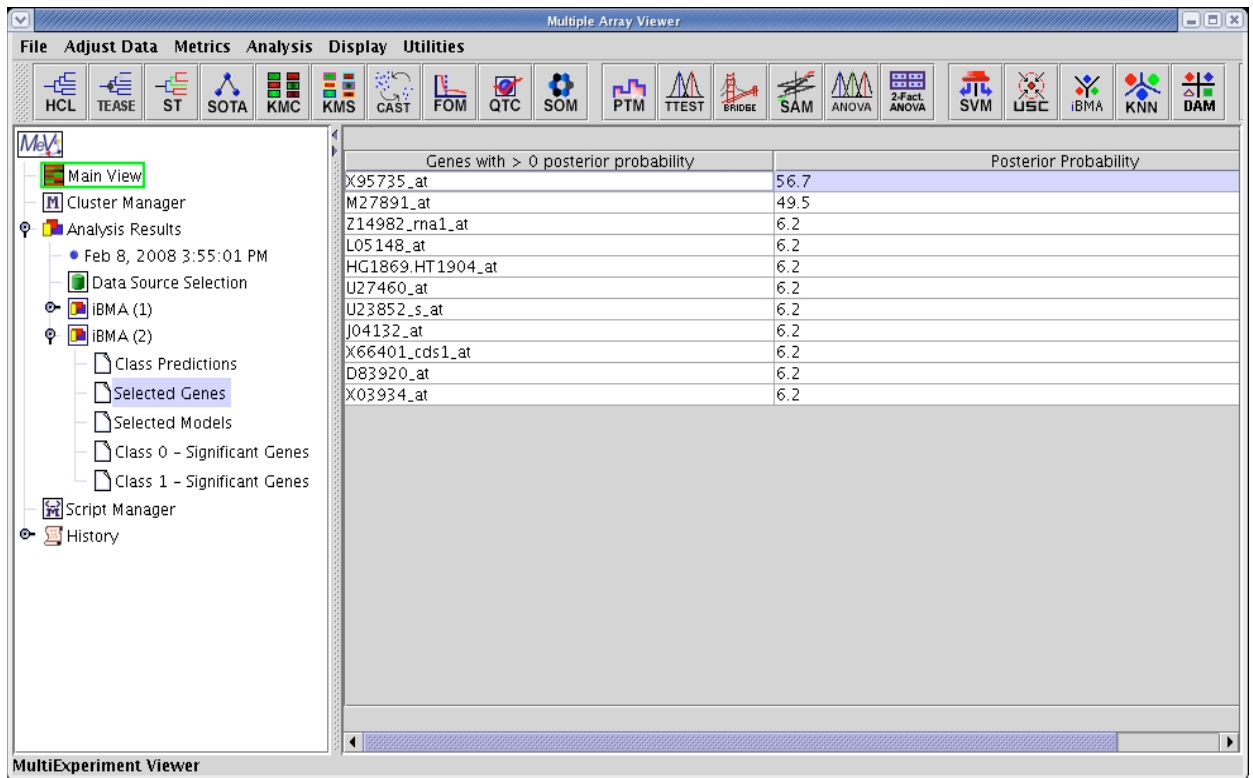
The next dialog box asks the user to assign labels to each of the samples in the data, either by using a pull-down menu or loading an assignment file. At this point, if Rserve is not already running, the user is reminded to start the connection. Then, the data and the parameters are sent to R, and a progress bar is shown warning the user that the computation could take a long time.

After the iterativeBMA Bioconductor package finishes running, the following analysis results are displayed. The following screen shots were obtained using the leukemia data in the case studies documented in the full manuscript.

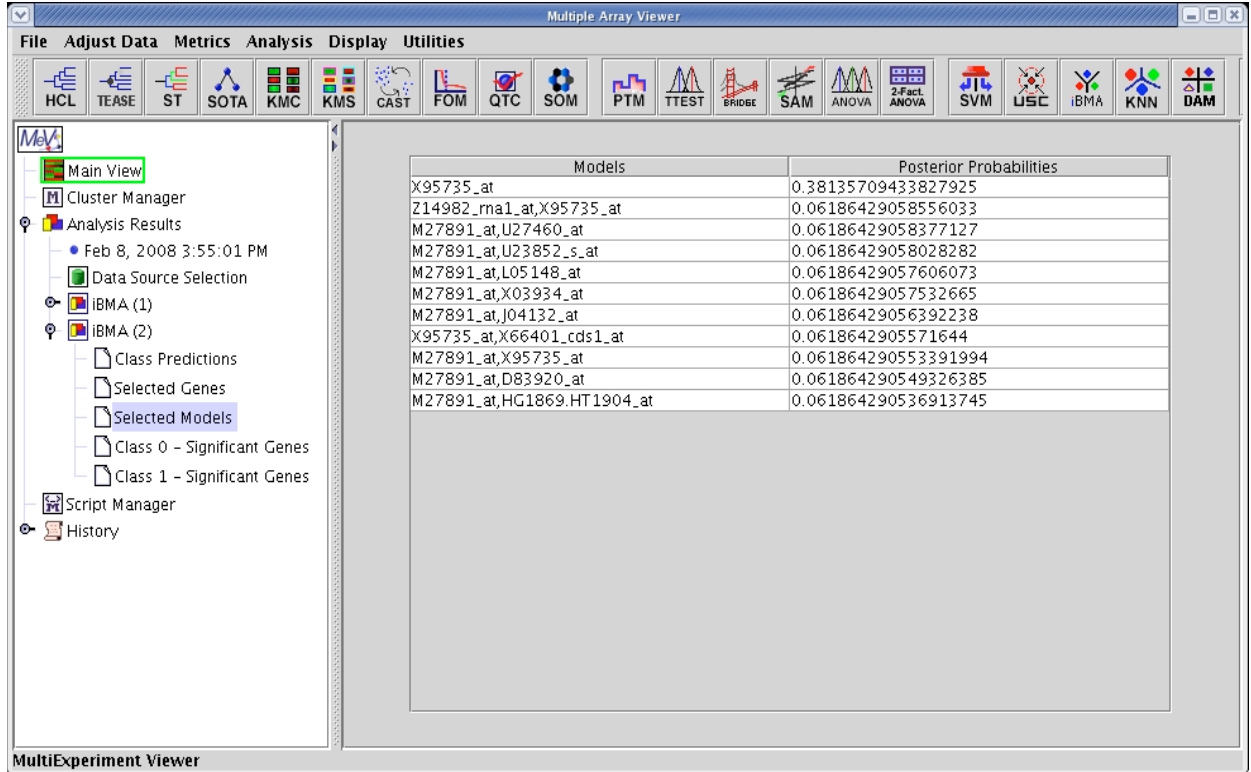
a. the predicted probability and class for each test sample



b. the posterior probabilities of the selected genes sorted in descending order,



c. the posterior probability of the selected model sorted in descending order,



d. the heatmaps of the selected genes in both classes (see Figure 4 in the manuscript).

## REFERENCES

1. Gottardo R, Raftery AE, Yeung KY, Bumgarner RE: **Robust estimation of cDNA microarray intensities with replicates.** *Journal of the American Statistical Association* 2006, **101**(473):30-40.
2. Gottardo R, Raftery AE, Yeung KY, Bumgarner RE: **Bayesian Robust Inference for Differential Gene Expression in cDNA Microarrays with Multiple Samples.** *Biometrics* 2006, **62**:10-18.