

# MultiExperiment Viewer Quickstart Guide

## Table of Contents:

- I. Preface - 2
- II. Installing MeV - 2
- III. Opening a Data Set - 2
- IV. Filtering - 6
- V. Clustering
  - a. HCL - 8
  - b. K-means - 11
- VI. Modules - T-test – 13
- VII. Additional Resources

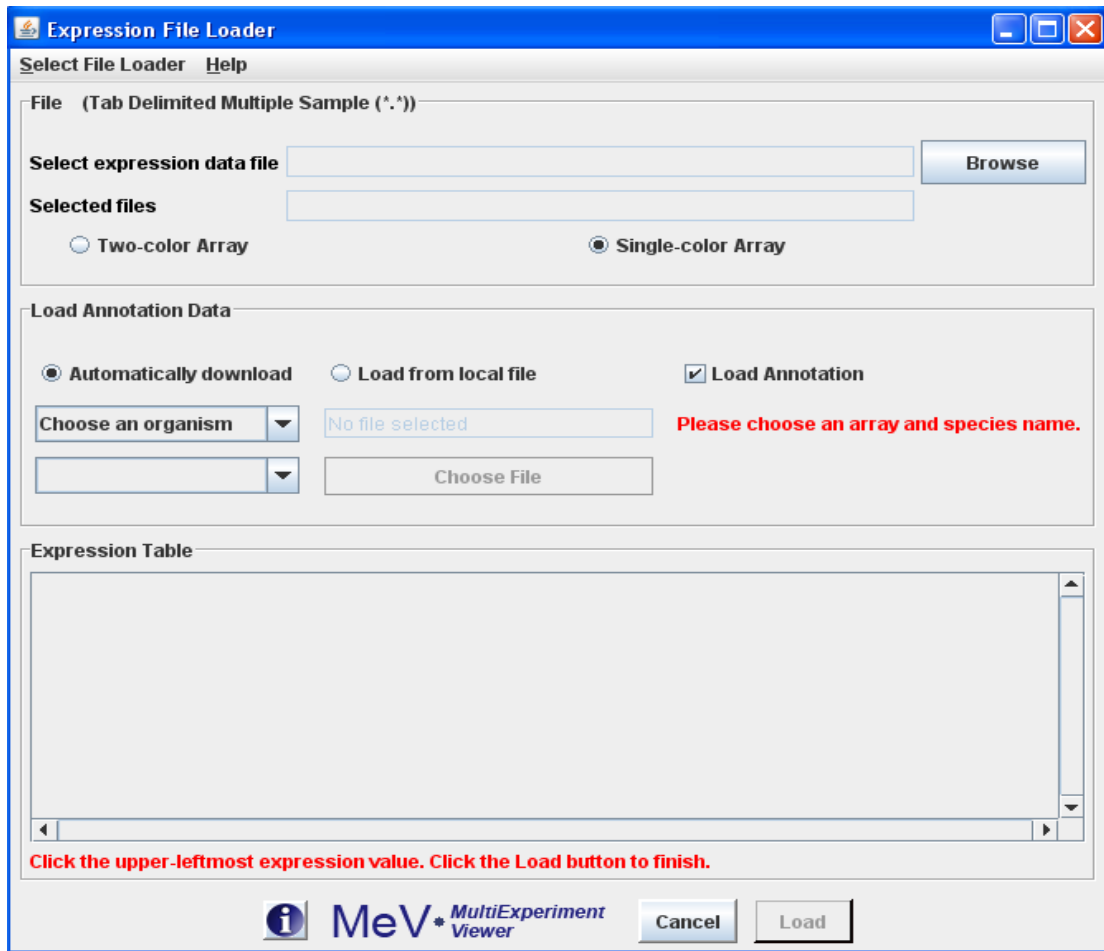
**Preface:** This guide is meant to be a quick reference for those who are just beginning to install and use MeV. Any questions that are not answered in this guide are most likely covered in the complete manual.

## **I. Installing MeV**

1. First make sure that Java is properly installed on your computer (not to be confused with Javascript, which is not the same). Java v1.6 or higher for a Windows PC/Linux and v1.5 or higher for Mac OSX needs to be installed in order for MeV to work. Go to <http://java.com/> to get the latest version. Certain MeV modules also require Java 3D, which can be found at <http://java3d.java.net/binary-builds.html>.
2. Go to [www.tm4.org](http://www.tm4.org). To begin download, go to the right hand side of the screen and click the “v4.8” link. The browser will be directed to a SourceForge download page specific to the current operating system.
3. A screen should pop up that asks what you want to do with the files. Whether using a PC, Mac, or Linux, open the files and download them.
4. Once downloaded, open the folder and unzip the file.
5. The unzipped folder can be copied to any convenient location on the hard drive.
6. Open the MeV\_4\_8 folder.
  - a. Windows users: double click the file called TMEV or tmev.bat to run the program.
  - b. Linux or Unix users: navigate to the MeV\_4\_8 folder in a terminal, and make tmev.sh executable by typing “chmod u+x tmev.sh” then “./tmev.sh”. MeV can hereafter be run by double-clicking the tmev.sh file.
  - c. Macintosh users: double-click the MeV application file.

## **Loading a Data Set**

1. In the Multiple Array Viewer, go to *File* → *Load Data*. Also, In the Multiple Experiment Viewer on the main menu bar, go to *File* → *New Multiple Array Viewer* and have more than one experiment window open at a time if needed.
2. When the Expression File Loader comes up, click *Browse* at the upper right side of the screen.



3. To choose one of the data sets that were downloaded with the program, choose *Desktop* from the drop down menu at the top. Open the MeV\_4\_8 folder, then open the “data” folder. If you have your own data set, find that file on your computer instead.
4. Choose a TDMS (Tab Delimited, Multiple Sample) file that ends with \*.txt and click *Open*. The one provided with the program is TDMS\_format\_sample.txt (The full MeV manual has a complete list and description of different file formats. The link to the manual can be found at <http://tm4.org/mev.html> under the heading Appendix: File Format Descriptions.)

### **Making Annotations**

Annotation gives you much more specific information about the gene, which could be more useful down the line.

- a. To use the Automatic Annotation loader, an Affymetrix dataset must be used.

**Expression File Loader**

Select File Loader Help

File (Tab Delimited Multiple Sample (\*.\*))

Select expression data file:

Selected file:

Two-color Array  Single-color Array

Load Annotation Data

Automatically download  Load from local file  Load Annotation

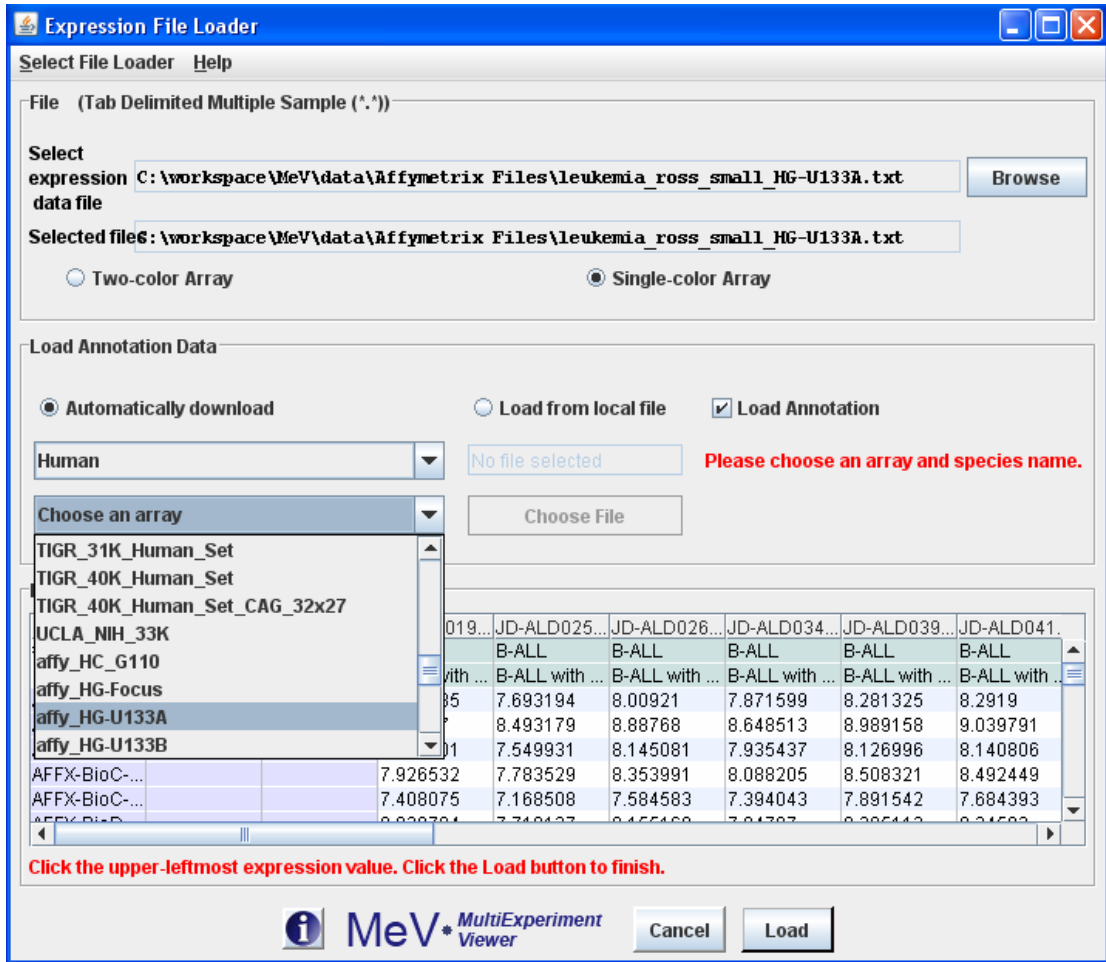
Choose an organism:   **Please choose an array and species name.**

Expression Table

| Probe_ID      | gene name | gene symbol  | JD-ALD019...   | JD-ALD025...   | JD-ALD026...   | JD-ALD034...   | JD-ALD039...   | JD-  |
|---------------|-----------|--------------|----------------|----------------|----------------|----------------|----------------|------|
| #cancer_s...  | --        | #cancer_s... | B-ALL          | B-ALL          | B-ALL          | B-ALL          | B-ALL          | B-A  |
| HG-U133A:     | gene name | #disease_... | B-ALL with ... | B-ALL with ... | B-ALL with ... | B-ALL with ... | B-ALL with ... | B-A  |
| AFFX-BioB-... |           |              | 7.557435       | 7.693194       | 8.00921        | 7.871599       | 8.281325       | 8.29 |
| AFFX-BioB-... |           |              | 8.22667        | 8.493179       | 8.88768        | 8.648513       | 8.989158       | 9.00 |
| AFFX-BioB-... |           |              | 7.572201       | 7.549931       | 8.145081       | 7.935437       | 8.126996       | 8.14 |
| AFFX-BioC-... |           |              | 7.926532       | 7.783529       | 8.353991       | 8.088205       | 8.508321       | 8.49 |
| AFFX-BioC-... |           |              | 7.408075       | 7.168508       | 7.584583       | 7.394043       | 7.891542       | 7.68 |
| AFFX-BioD...  |           |              | 8.028794       | 7.719127       | 8.155169       | 7.84797        | 8.285112       | 8.34 |

**Click the upper-leftmost expression value. Click the Load button to finish.**

- b. Note that in the sample data, the Affymetrix files have their own folder within the data folder. These have already been normalized. Again, for descriptions on other types of files please see the MeV Manual.
- c. To use the automatic annotation downloader, an internet connection is needed. Next, click the *Automatically download* radio button and check the *Load Annotation* check box. In the same panel there will be two drop down menus. From the top one, choose the species you are working with, which in the example case is human. Then in the bottom drop down menu choose the correct annotation file, which is the type of chip used in the experiment. This file will begin with “affy” and is found a little further down in the menu.



- d. To use an annotation file from the data folder or from your computer instead, click the *Load from local file* radio button within the Load Annotation Data box and click *Choose File*. Then click *Open*.
5. Once that is completed, whether an annotation was used or not, click the upper-leftmost expression value. That is, under the Expression Table, click the first box that contains a quantitative value. These include columns that contain numerical values such as sample data, but not columns that are labels, names, or other such unquantifiable values. (Note: There may be columns that have 0 or 1 in them to identify data, but are not expression values.)

| Expression Table |           |              |              |                |                |                |                |                |                |
|------------------|-----------|--------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Probe_ID         | gene name | gene symbol  | JD-ALD019... | JD-ALD025...   | JD-ALD026...   | JD-ALD034...   | JD-ALD039...   | JD-ALD041...   | JD-ALD051...   |
| #cancer_s...     | --        | #cancer_s... | B-ALL        | B-ALL          | B-ALL          | B-ALL          | B-ALL          | B-ALL          | B-ALL          |
| HG-U133A:        | gene name | #disease...  | B-ALL with   | B-ALL with ... | B-ALL with ... | B-ALL with ... | B-ALL with ... | B-ALL with ... | B-ALL with ... |
| AFFX-BioB...     |           |              | 7.557435     | 7.693194       | 8.00921        | 7.871599       | 8.281325       | 8.2919         | 7.119348       |
| AFFX-BioB...     |           |              | 8.22667      | 8.493179       | 8.88768        | 8.648513       | 8.989158       | 9.039791       | 7.911919       |
| AFFX-BioB...     |           |              | 7.572201     | 7.549931       | 8.145081       | 7.935437       | 8.126996       | 8.140806       | 6.996941       |
| AFFX-BioC...     |           |              | 7.926532     | 7.783529       | 8.353991       | 8.088205       | 8.508321       | 8.492449       | 8.618484       |
| AFFX-BioC...     |           |              | 7.408075     | 7.168508       | 7.584583       | 7.394043       | 7.891542       | 7.684393       | 7.930426       |
| AFFX-BioD...     |           |              | 8.028794     | 7.719127       | 8.155169       | 7.84797        | 8.285112       | 8.34583        | 8.54395        |
| AFFX-BioD...     |           |              | 10.207125    | 10.346668      | 10.580919      | 10.204725      | 10.526828      | 10.691249      | 10.917079      |
| AFFX-CreX...     |           |              | 11.218895    | 11.406426      | 11.350053      | 11.232634      | 11.351477      | 11.593959      | 11.556497      |
| AFFX-CreX...     |           |              | 11.503832    | 11.613685      | 11.844216      | 11.697612      | 11.809588      | 11.883938      | 11.956537      |
| AFFX-DapX...     |           |              | 2.214032     | 2.198321       | 2.272796       | 2.242986       | 2.349864       | 2.308443       | 2.355578       |
| AFFX-DapX...     |           |              | 2.294538     | 2.17948        | 2.139439       | 2.453214       | 2.133412       | 2.106146       | 2.213904       |

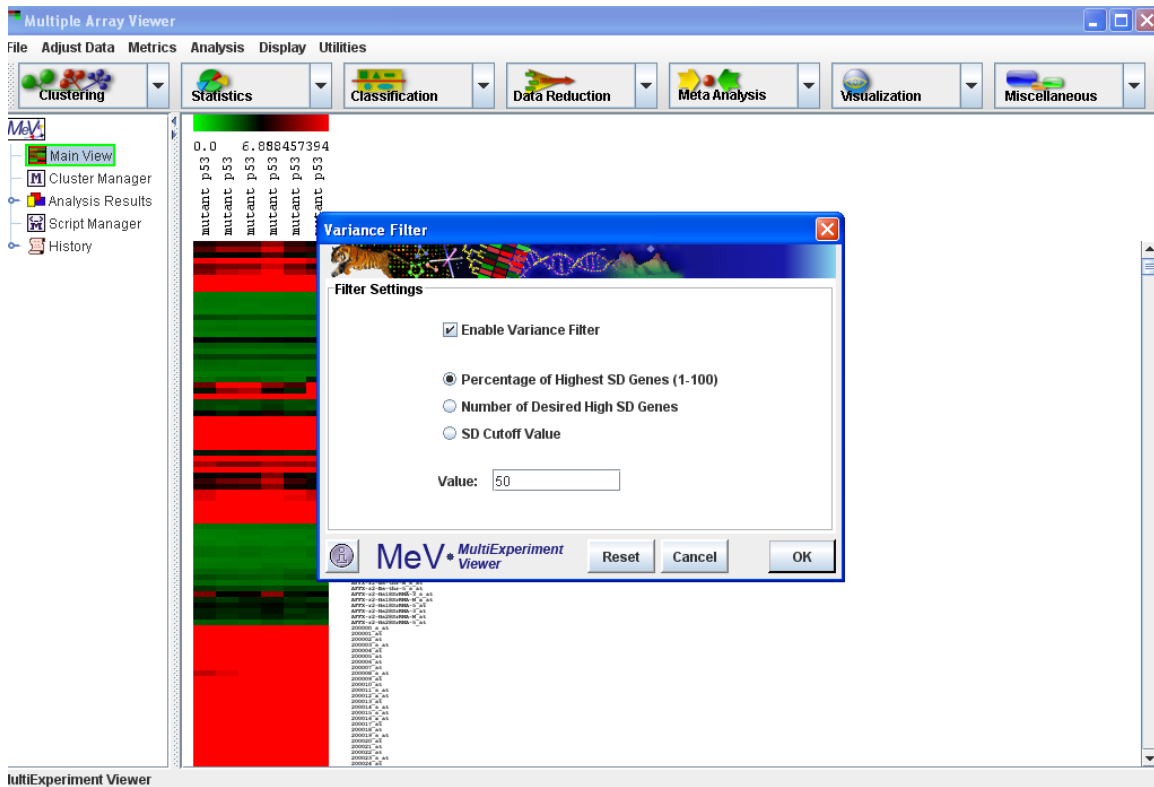
Click the upper-leftmost expression value. Click the Load button to finish.

6. Click *Load* at the bottom of the screen.
7. A diagram should come up that looks like lots of tiny green and red boxes. This is the **Heatmap View**. Each row represents a specific gene, whereas each column represents each sample, or experiment. The lightest green boxes are the most underexpressed genes, the brightest red being the most overexpressed genes. MeV expects that each sample loaded will have the same number of elements, in the same order, and that each gene (or spot) is aligned with that element in every other sample loaded. For example, using that rule, all input files will have data for gene x in row y. Clicking on a rectangle displays a dialog with detailed information about that spot.
8. The MeV toolbar on the left hand side should have 5 options. This is called the **Result Tree**. Any heading that has a node to the left of it means that when double clicked the information within that heading will appear below it. Click *Main View* to return to the original Heatmap View at any time.
9. To find the basic information for the data set, double click *Analysis Results* → *Data Source Selection Information*. This will tell you how many genes are in the data set and how many trials of the experiment were performed with each gene.
10. Under *History*, the History Log shows which tests were done on the data set and when.

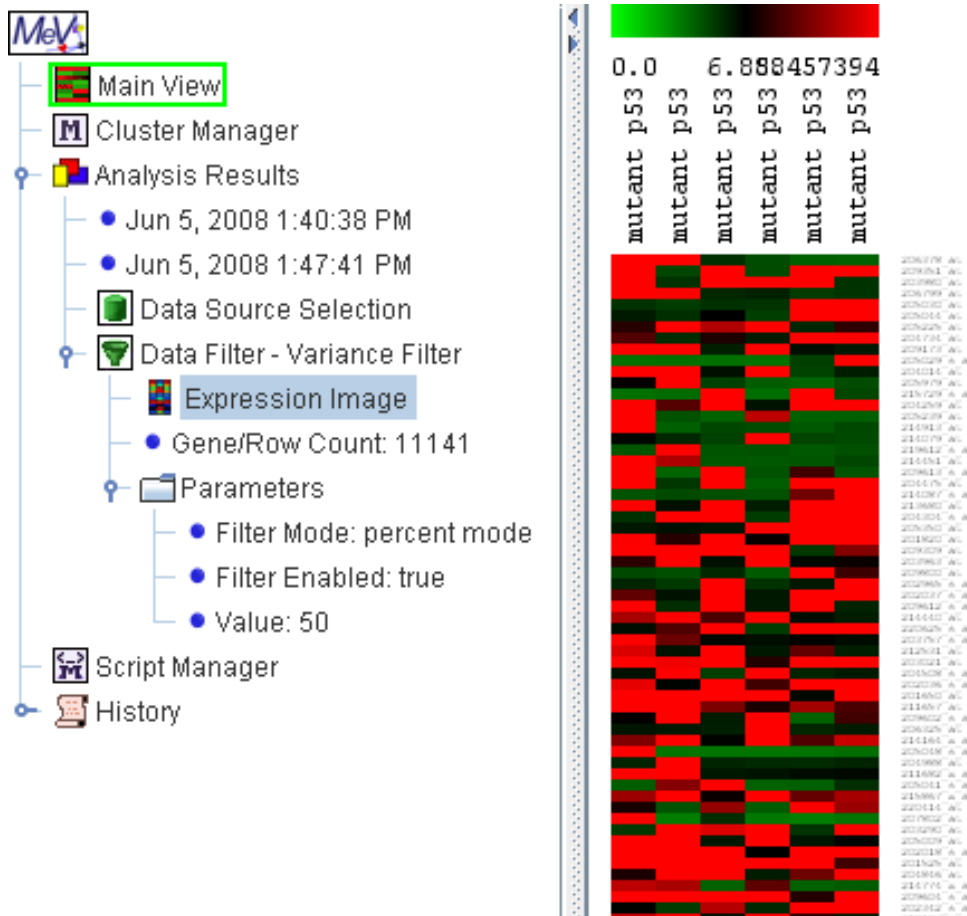
## Filtering

Many data sets contain tens to hundreds of thousands of genes, so filtering out the ones that contain little biological information makes doing analyses much easier. If the data set is not large, then filtering may not be needed.

1. Go to the main menu on top and click *Adjust Data* → *Data Filters* → *Variance Filters*. Put in the percentage that you wish to filter out. We will use 50% here as a default, but a lower percentage to filter out even more genes can be chosen.



2. The filtered data can be seen by using the navigation tree on the left. It is under *Analysis Results* → *Data Filter* → *Expression Image*. To see the type of filter that was used and the percentage filtered out look under parameters which is further down.



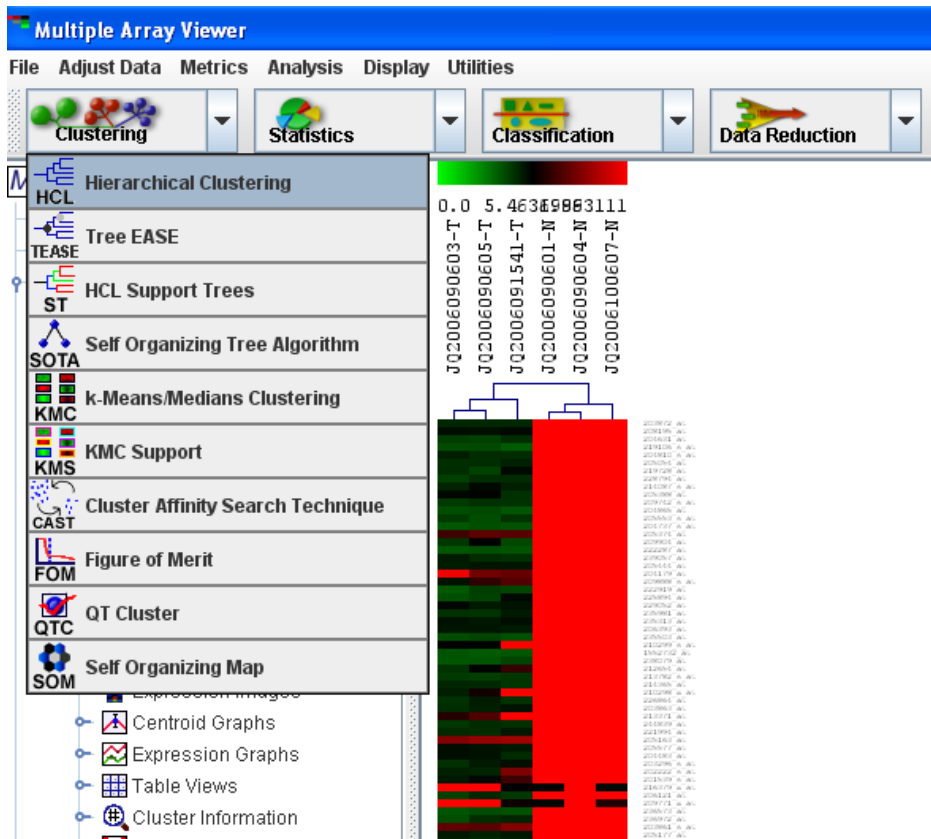
From now on, all tests done on the sample will be done on the filtered data.

The algorithm toolbar on top, where the icons say *Clustering*, *Statistics*,... etc, is where all the main modules are found. The same modules can also be found under *Analysis* in the main toolbar above it.

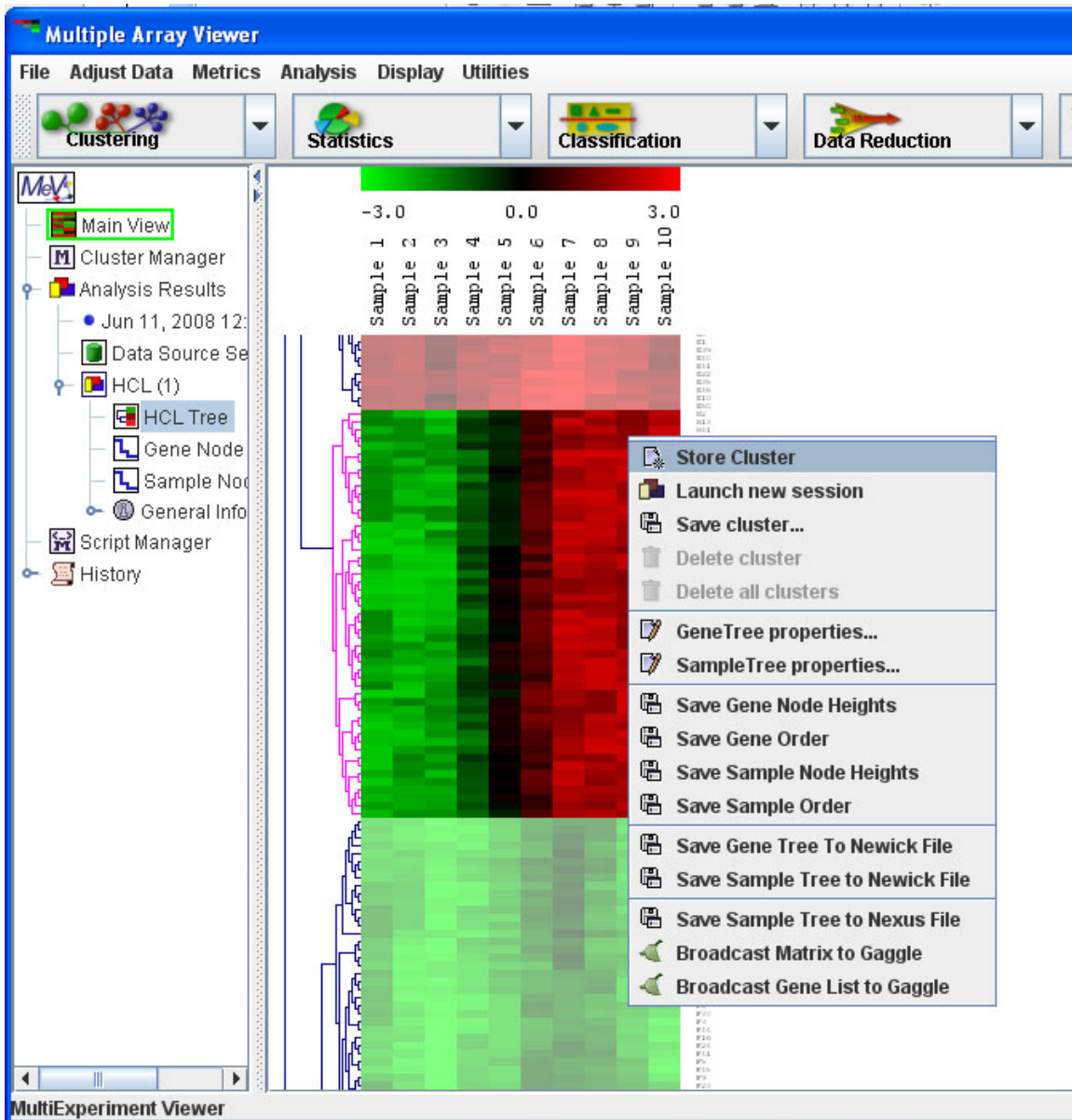
## Clustering

1. Clustering basically groups similar samples or genes together. This way it makes it much easier to test the data, since you can see which groups are more alike.
  - a. **HCL:** A commonly used clustering algorithm is Hierarchical Clustering, or HCL. From the top left click the *Clustering* drop down menu. Choose the first option, *Hierarchical Clustering*. **NOTE:** The downside to this test is that it can take a long time to perform if there are a lot of genes, and if there are too many an error message may appear. 10,000 genes is a standard cutoff point, but even with that many genes it may take up to 20 minutes. If only similarities among samples and not genes are being looked at, simply uncheck the box at the top that says Gene Tree. If there are a lot of genes and they need to be grouped into similar groups, the k-means cluster in the next section may be more useful

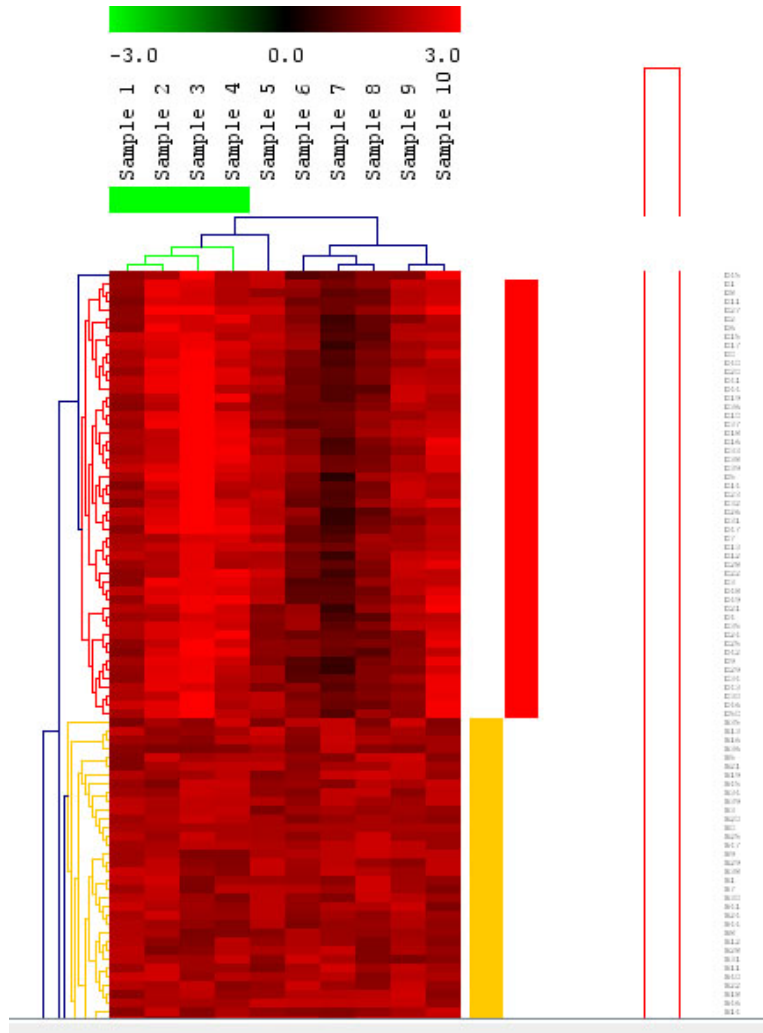




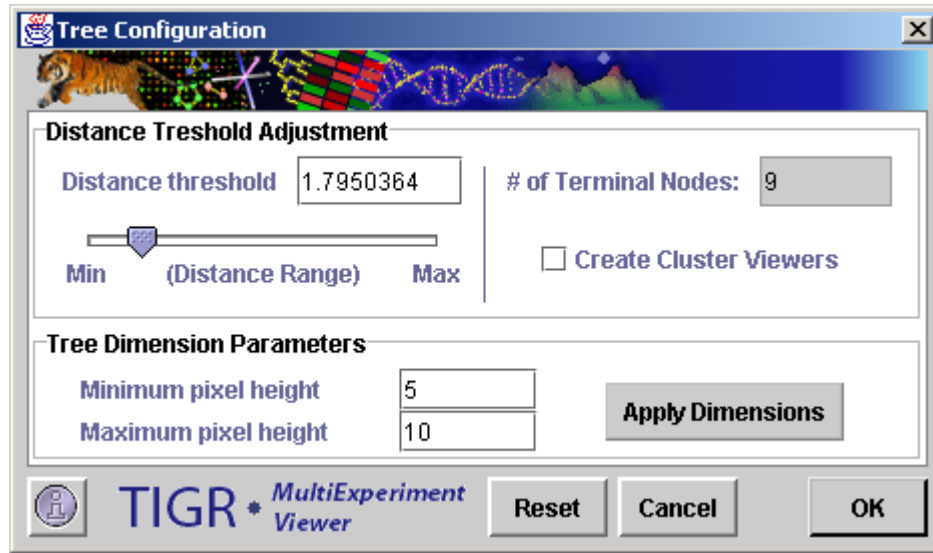
- b. Click *OK* at the bottom right.
- c. To view the clusters, find and double click *HCL* on the navigation tree on the left.
- d. To store the clusters, click on the groups that you want together. That is, click the line that connects a certain group of genes. That group will then be highlighted. Right click on the line and choose *Store Cluster*.



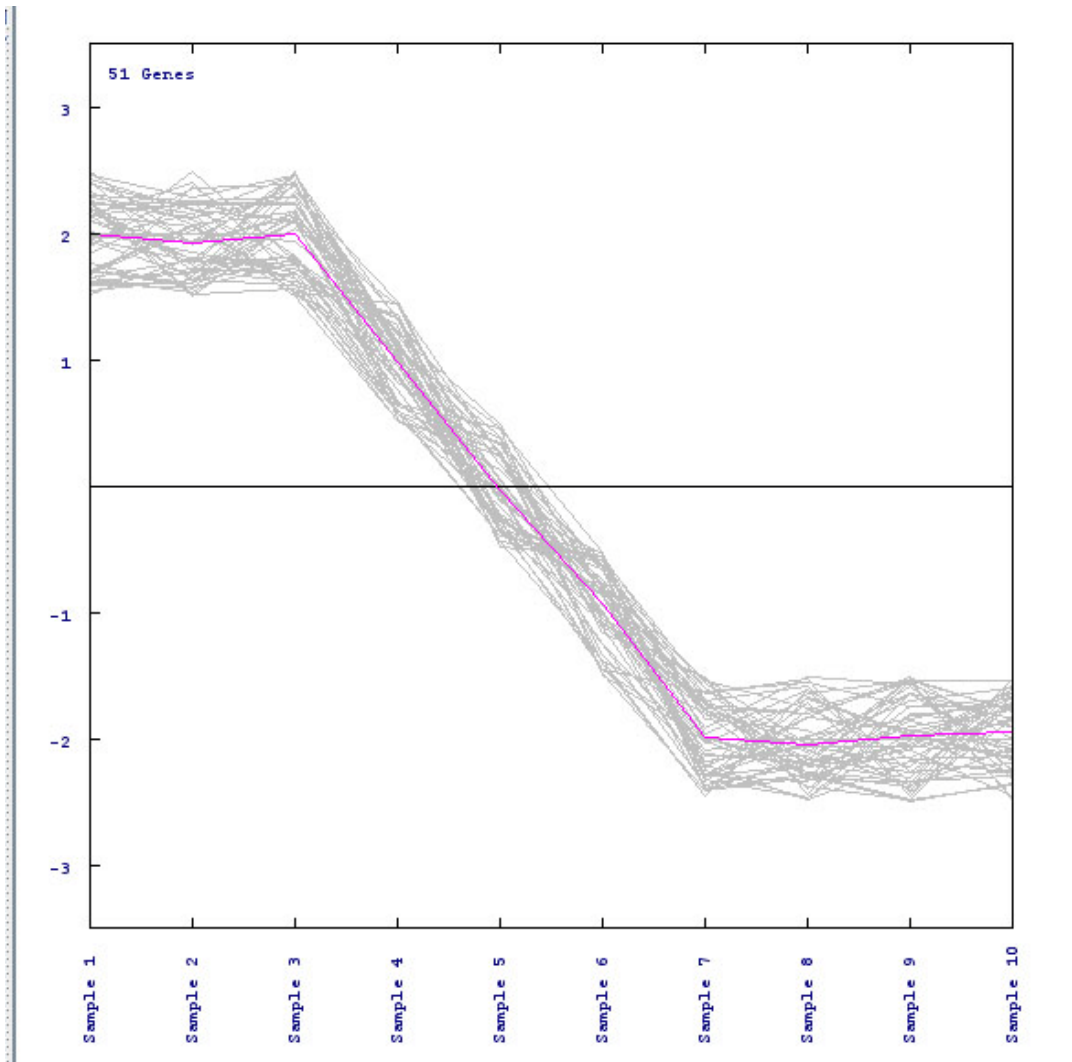
- e. The program should automatically choose a new color for each cluster so that they can easily be viewed and distinguished.



- f. To get a more detailed result, right click in the gene tree and select *Gene Tree Properties*.

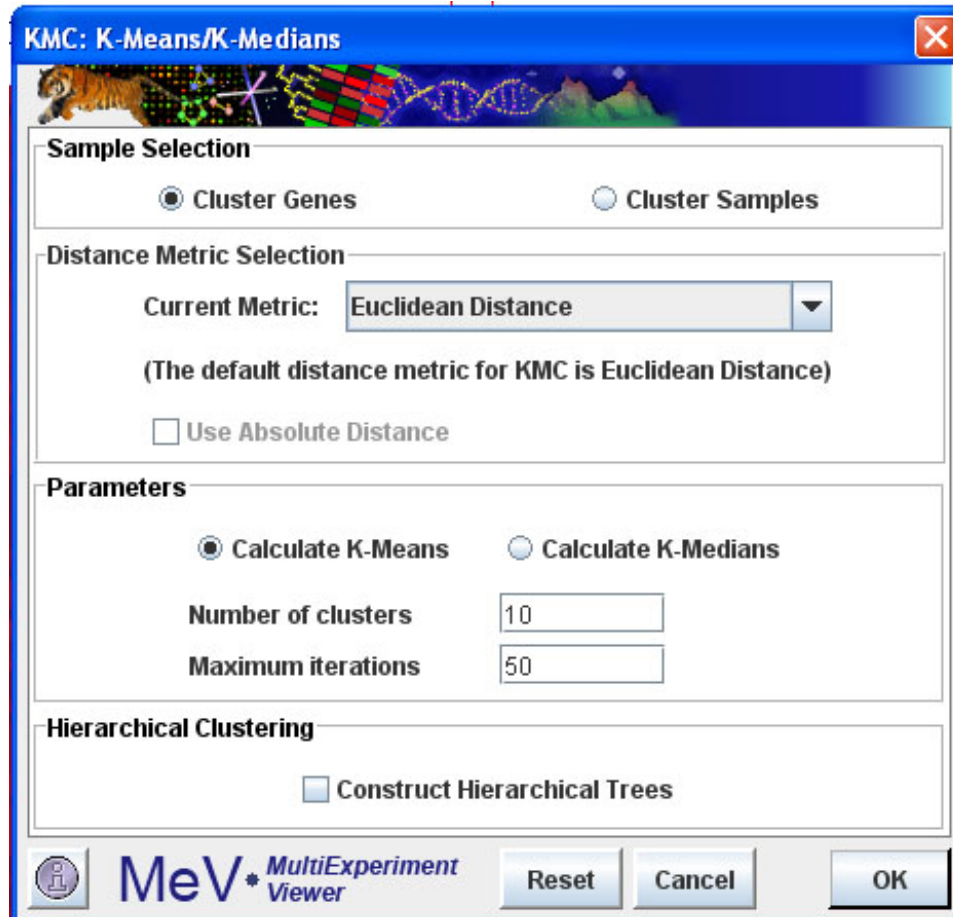


- g. Under *Distance Threshold Adjustment*, the number of clusters needed can be adjusted. The lower the distance range, the more clusters there will be since a shorter distance between similar genes means more groups. A light gray triangular ‘wedges’ along the side will appear as this distance threshold changes. Either use the slide or type in a value for the distance threshold until you get the desired number of clusters.
- h. To get different graphs and charts for each cluster created click the box on the right *Create Cluster Viewers*. These can be very useful visualization tools as well being able to look closer at each cluster.
- i. Click *OK*.
- j. If the desired number of clusters were not the right number, going back is always an option. Also, if the navigation tree gets too cluttered, simply right click the item you wish to delete and click *Delete*.
- k. Each individual cluster in the navigation tree on the left can be viewed. Either view different graphs of an individual cluster, or choose *All Clusters* and see all of them on one screen.
- l. The Expression images are the red and green rectangles like that of the main view. The Centroid Graphs show the means and ranges of each sample. The Expression Graphs give a more detailed representation of the gene behavior.



Each individual line represents one gene, and the pink line in the middle shows the average expression of all the genes in each sample.

- m. To do an analysis on just one particular cluster (such as a t-test, which will be discussed in the Modules section), right click on the individual cluster and choose *Launch New Session*. This will open a new window with just the particular cluster you wish to analyze.
  - n. To see a list of all of the clusters you made double click Cluster Manager in the navigation tree.
2. **K-Means:** This cluster is useful if you have a large data set. It can also be found in the *Clustering* drop down menu. The benefits of this cluster are that it is much faster than the HCL and that the exact amount of clusters wanted can be chosen.
- a. Select *Clustering* → *k-Means/Medians Clustering* from the drop down menu.
  - b. Choose whether you want to cluster the genes or the samples.

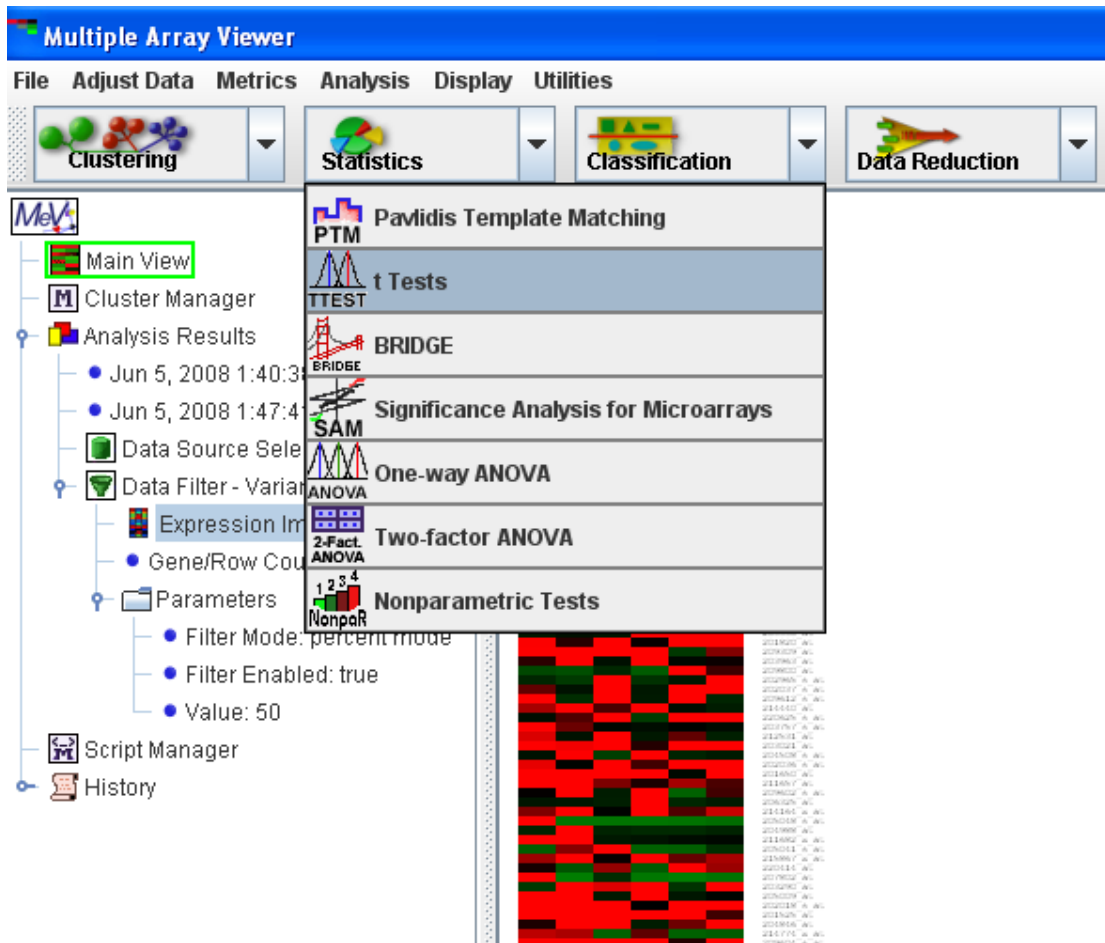


- c. Choose the amount of clusters to form. Here we will use 10, but it is always an option to go back and select fewer (or more) if those results do not give you the information you want.
  - d. The clusters in the navigation tree can be viewed on the left just as the HCL was viewed
3. For more information on clustering, there is a more detailed section in the Manual called Working with Clusters.

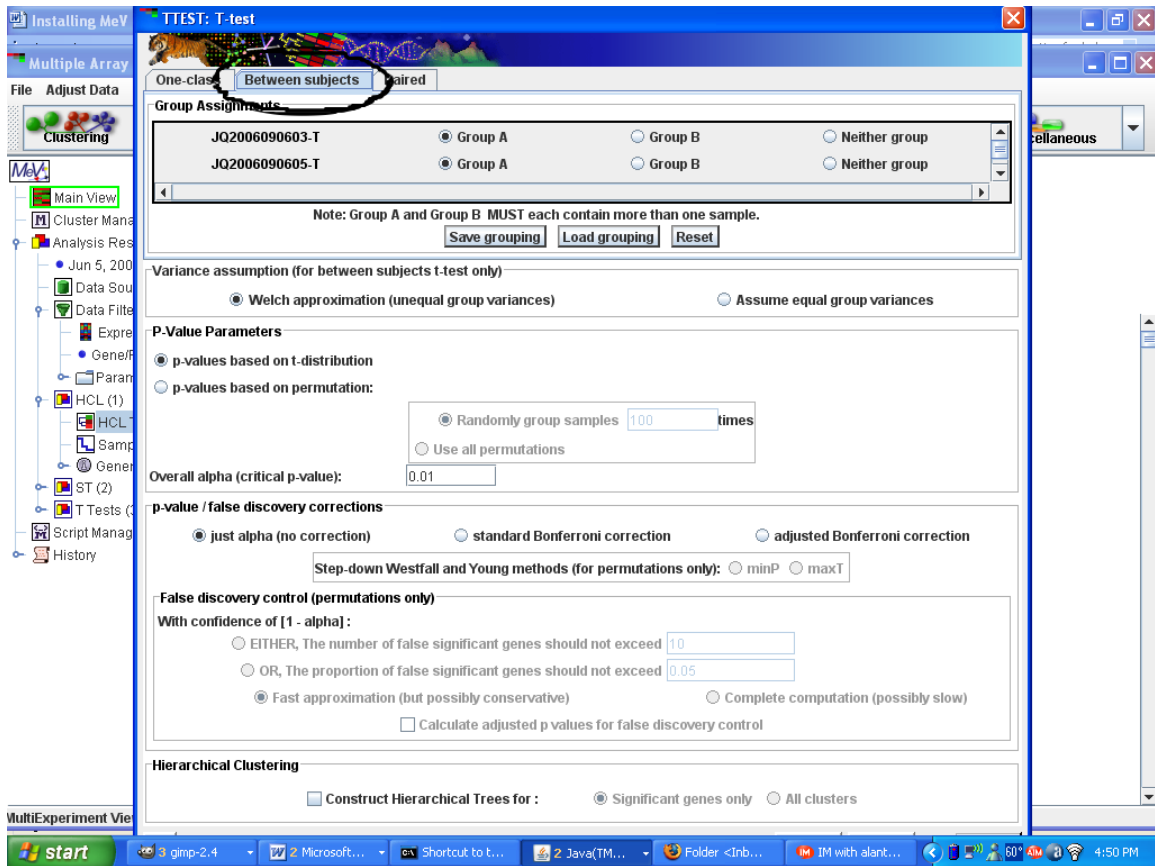
## Modules

Modules test for differences, patterns, and other such relationships among genes in the samples.

1. A **t-test** is the basic test to look at the differences among groups, if any, and how significant those differences are. In other words, it takes the means of two groups and determines if the difference in those means is significant.
2. Below the main drop down menu, click the icon that says *Statistics*, and then choose *T-tests*.



3. First, make sure you're choosing a 2 sample t-test, since the screen will automatically come up with a one sample. The difference is that you want to test for the differences among two groups, such as cancerous cells versus normal ones, to try and find the significant differences between the two groups. Do this by choosing the *Between Groups* tab at the top of the t-test screen.



4. Select which samples belong to which group. For example, all the columns that are labeled as cancerous, put them in Group A. All the normal samples, click the bubble next to Group B. For other groups that are not being tested against each other, they can be marked as Neither group.
5. Check *Assume equal variance* on the right. This means that the variance within all of the samples is the same, since the data has all been normalized.
6. Note that the T-test box is rather large and the *OK* button may not automatically appear at the bottom of the page. This is merely a technical problem. In order to fix that, move the window down by clicking the bar on the top that says *TTEST: T-test* and dragging it. Shrink it down using the arrows at the edge of the window. Once shrunken, move the window back up and the *OK* button should appear on the bottom right. Click it to run the test.
7. The result to the T-test is in the navigation tree along with the filtering and clustering results.

### Additional Resources

For more information on the different modules, consult the MeV Manual, found in the “documentation” folder, or go to <http://www.tm4.org>.