

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

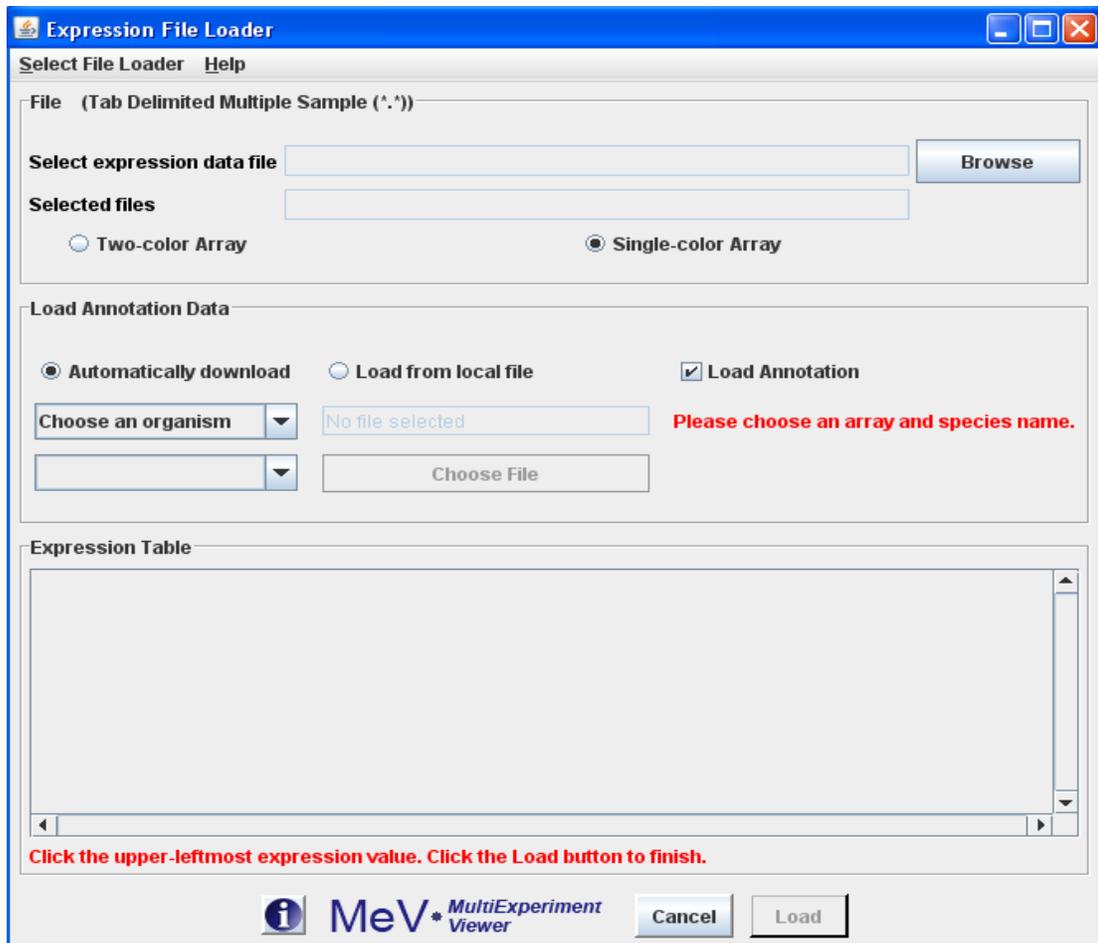
Preface: This guide is meant to be a quick reference for those who are just beginning to install and use MeV. Any questions that are not answered in this guide are most likely covered in the complete manual.

I. Installing MeV

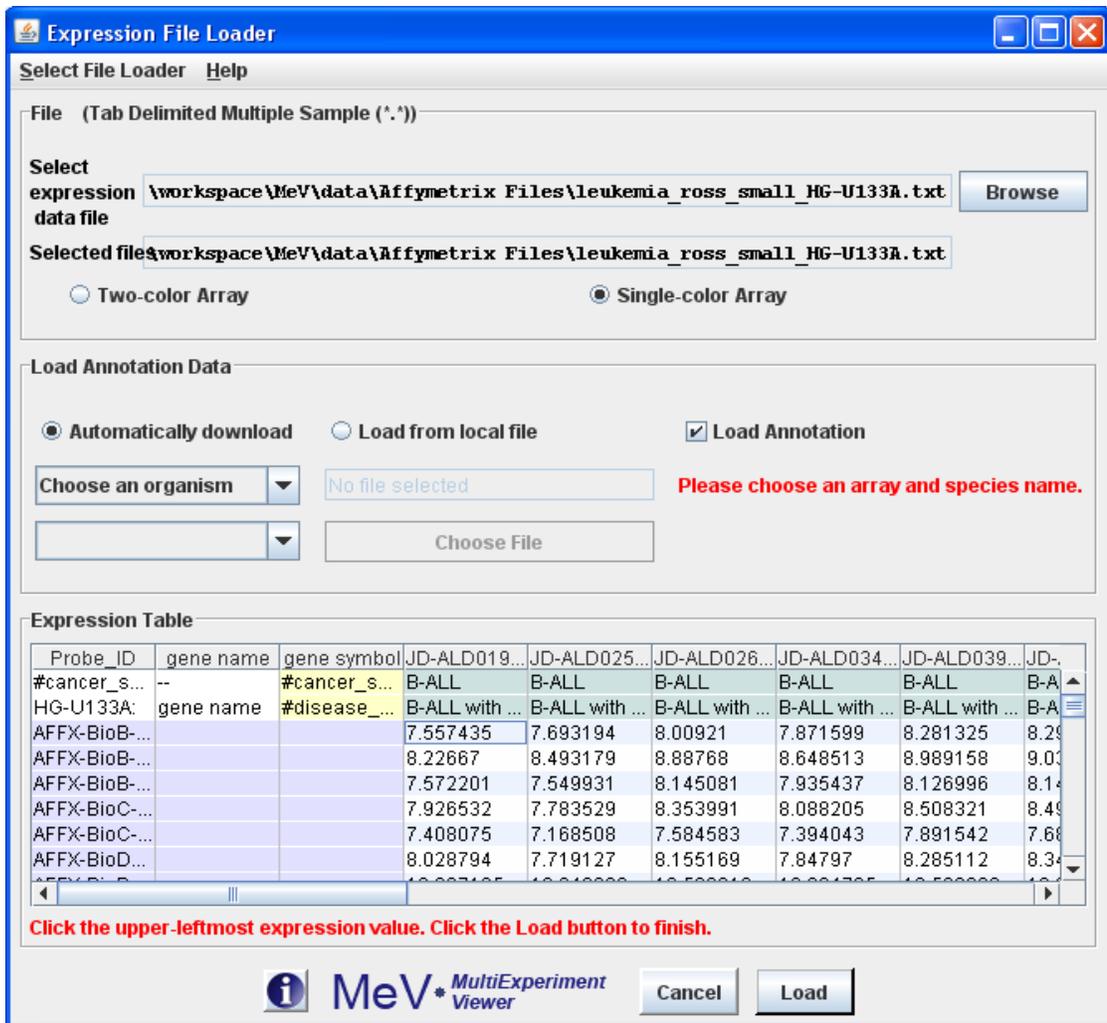
1. First make sure that Java is properly installed on your computer (not to be confused with Javascript, which is not the same). You will need to have Java v1.6 or higher for a Windows PC/Linux and v1.5 or higher for Mac OSX. Go to <http://java.com/> to get the latest version. Certain MeV modules also require Java 3D, which can be found here: <http://java.sun.com/products/java-media/3D/download.html>
2. Go to www.tm4.org. To begin download, go to the right hand side of the screen and click the “v4.6” link for the operating system you use. Note that there are different links for PC, Mac and Linux.
3. A screen should pop up that asks what you want to do with the files. Whether using a PC, Mac, or Linux, you want open the files and download them to your computer. Then save the files to your computer if it does not automatically do so.
4. Once downloaded, open the folder and unzip all files.
5. To best access the folder again, copy the MeV_4_6 folder onto a location on you're your hard drive.
6. Open the MeV_4_6 folder. If using Windows, double click the file called tmev.bat to run the program. For Linux or Unix users, navigate to the MeV_4_6 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. For Macintosh users, open the TMEV_MacOSX_4_6 file. A black MSDOS prompt screen should open, and MeV should open shortly after.

Opening a Data Set

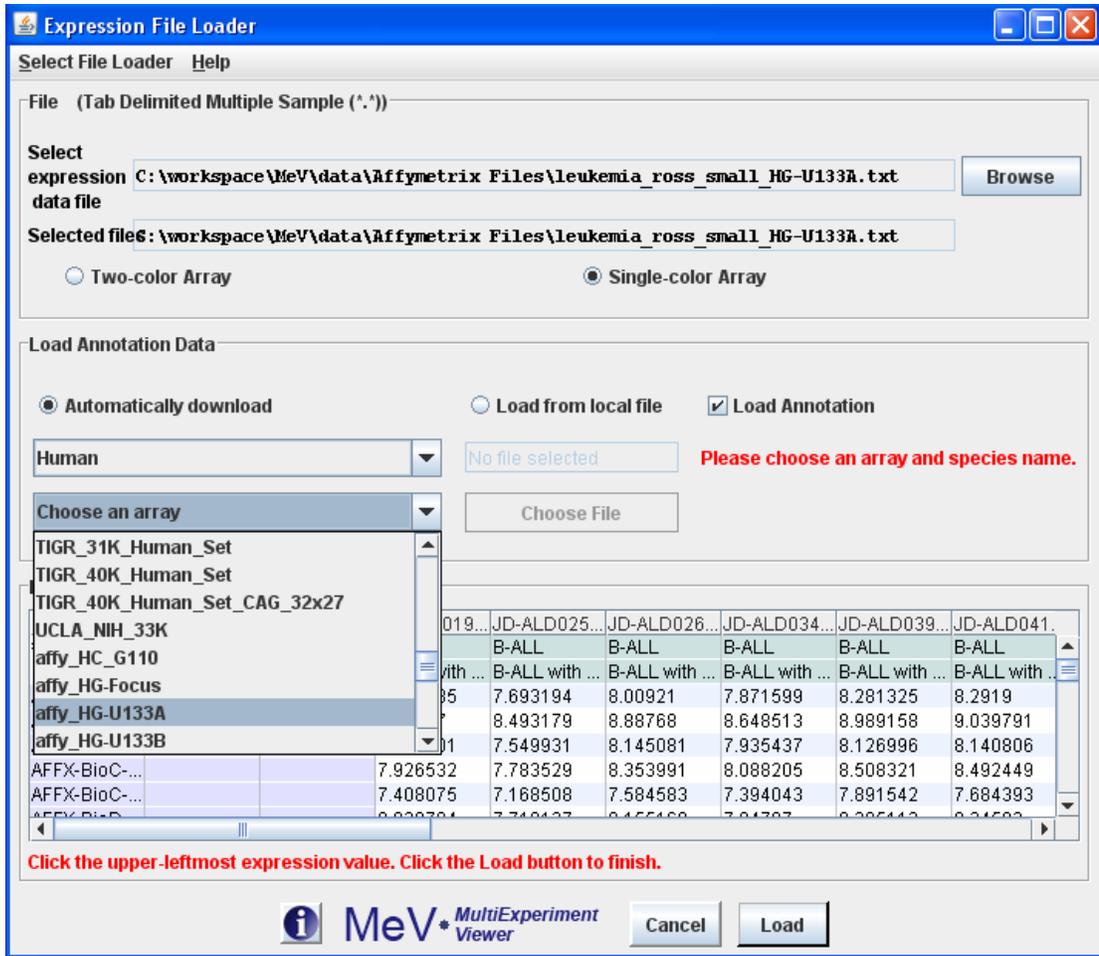
1. In the Multiple Array Viewer, go to *File* → *Load Data*. (There is also a Single Array Viewer that can only open one set of data at a time, but only the Multiple allows you to view many samples together. The real power of MeV is in the program's analysis modules, found only in the Multiple Array Viewer. That is where the clustering and visualization of data can take place. Therefore, the remainder of this guide will focus on the Multiple Array Viewer. Also, In the Multiple Experiment Viewer, you can go to *File* → *New Multiple Array Viewer* and have more than one experiment window open at a time.)
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_6 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 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.)
5. When do I use Annotation? 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, you must be using an Affymetrix dataset.



- 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. If you want to use the automatic annotation downloader, you must be connected to the Internet. 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. If you want 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*.
6. 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.)

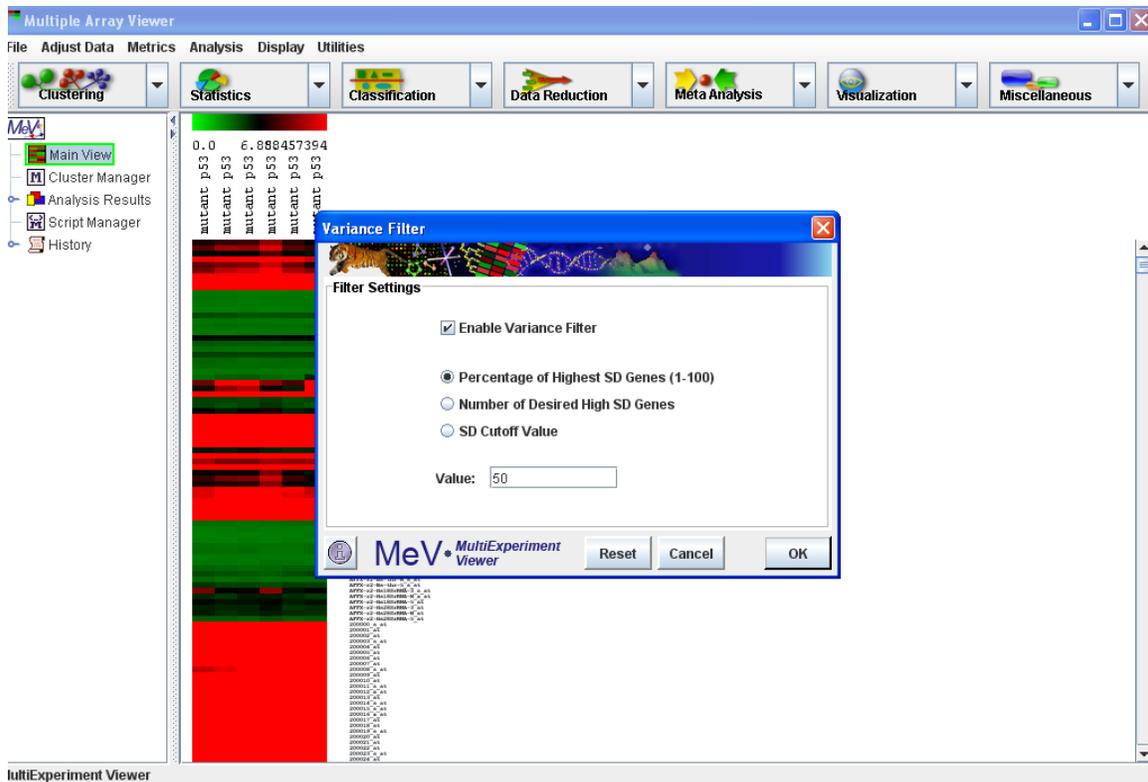
Probe_ID	gene name	gene symbol	JD-ALD019...	JD-ALD025...	JD-ALD026...	JD-ALD034...	JD-ALD039...	JD-ALD041...	JD-ALD051...
#cancer_s...	--	#cancer_s...	B-ALL	B-ALL	B-ALL	B-ALL	B-ALL	B-ALL	B-ALL
HG-U133A:	gene name	#disease...	B-ALL with	B-ALL with ...					
AFFX-BioB...			7.557435	7.693194	8.00921	7.871599	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.

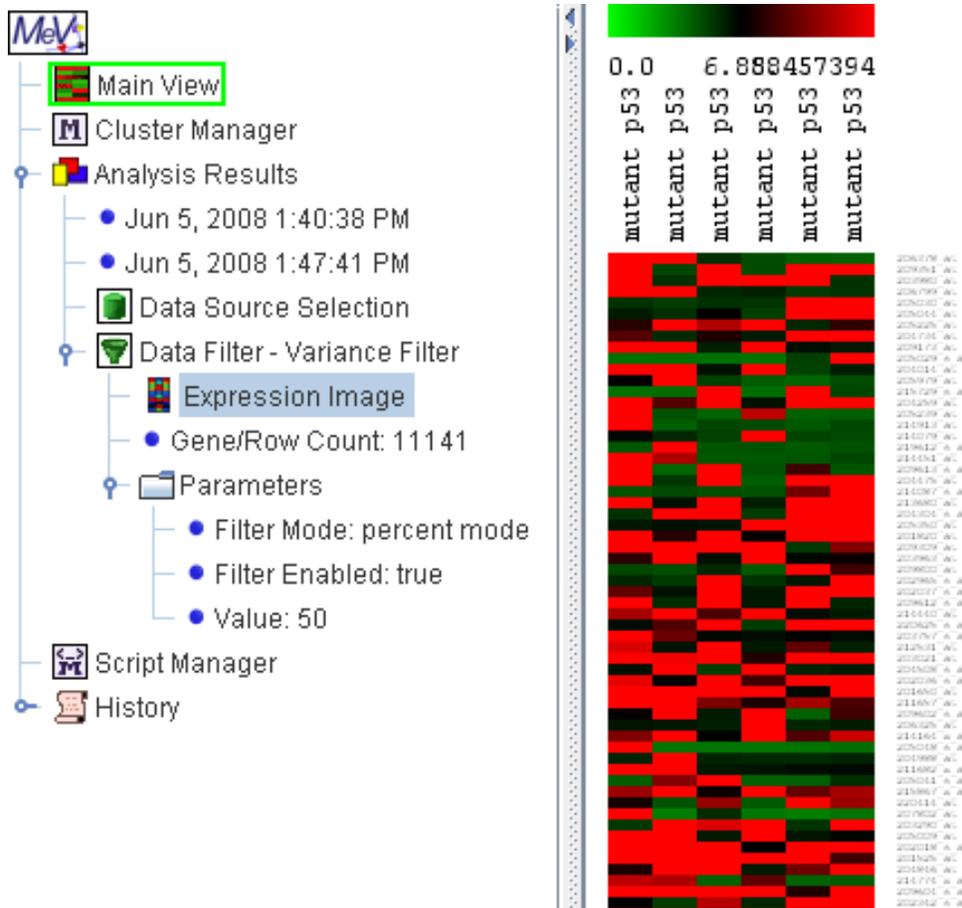
- Click *Load* at the bottom of the screen.
- A diagram should come up that looks like lots of tiny green and red boxes. This is the **Heatmap View**. Each row represents a specific gene, whereas each column represents each sample, or experiment. The lightest green boxes are the most underexpressed genes, 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.
- The MeV toolbar on the left hand side should have 5 options. This is called the **Navigation Tree**. Any heading that has a node to the left of it means that if you double click it, the information within that heading will appear below it. Click *Main View* to return to the original Heatmap View at any time.
- 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.
- Under *History*, the History Log will show you which tests were done on the data set and when.

Filtering

- Why filter? 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 you may not need to filter it.
- 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 you can choose a lower percentage to filter out even more genes.



3. The filtered data can be seen by using the navigation tree on the left. It is under *Analysis Results* → *Data Filter* → *Expression Image*. If you go further and look under parameters, it will tell you the type of filter that was used and the percentage filtered out.

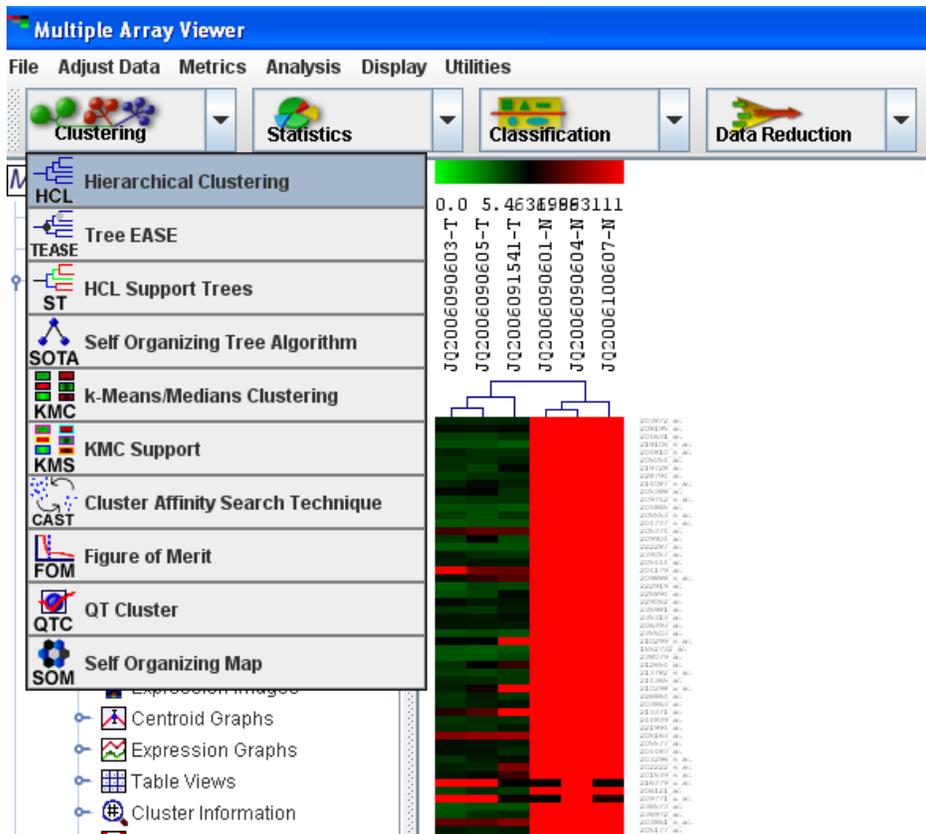


From now on, all tests done on the sample will be done on the filtered data.

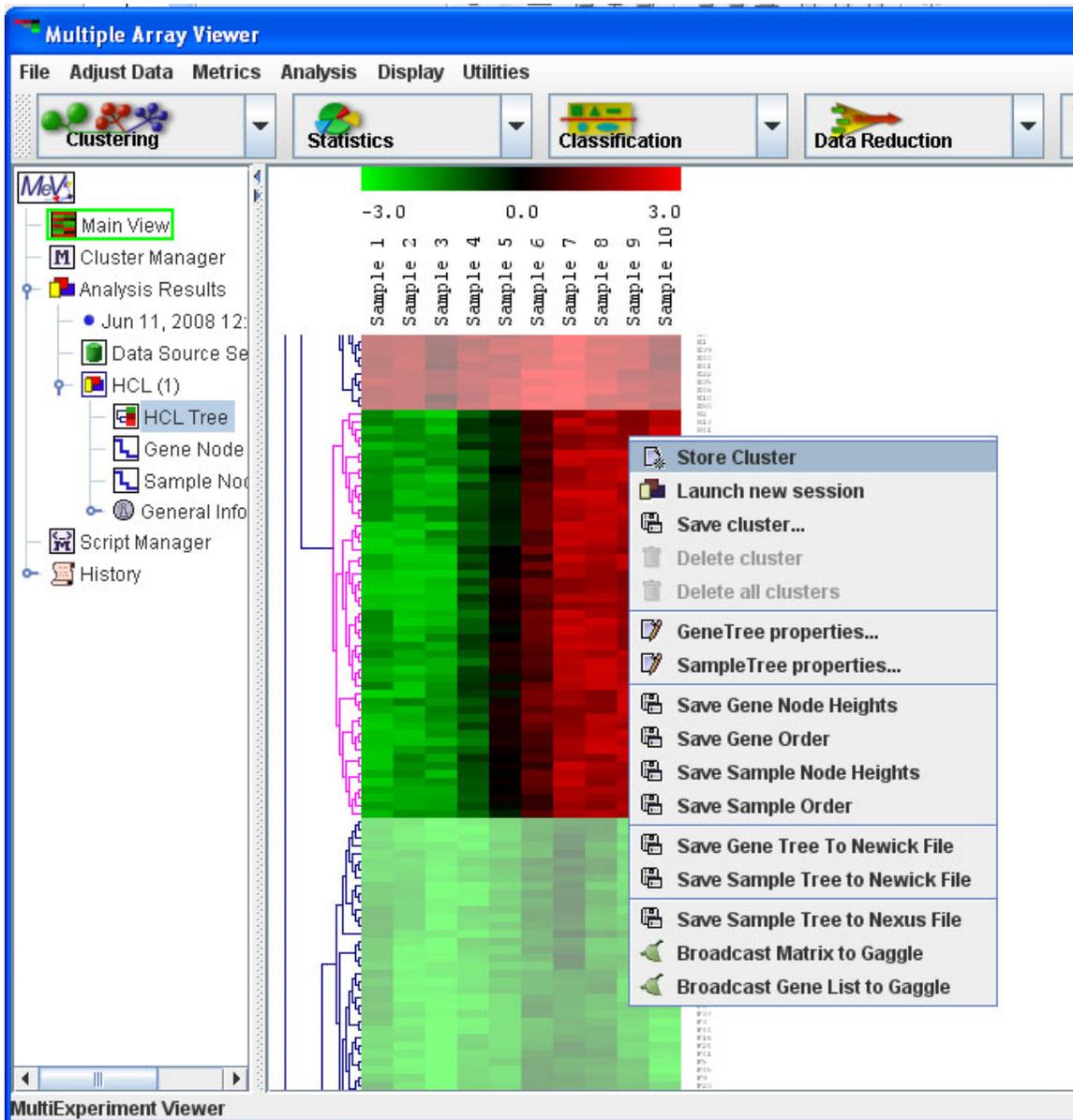
The algorithm toolbar on top, where the icons say *Clustering*, *Statistics*,... etc, is where all the main modules are found. The same modules can also be found under *Analysis* in the main toolbar above it.

Clustering

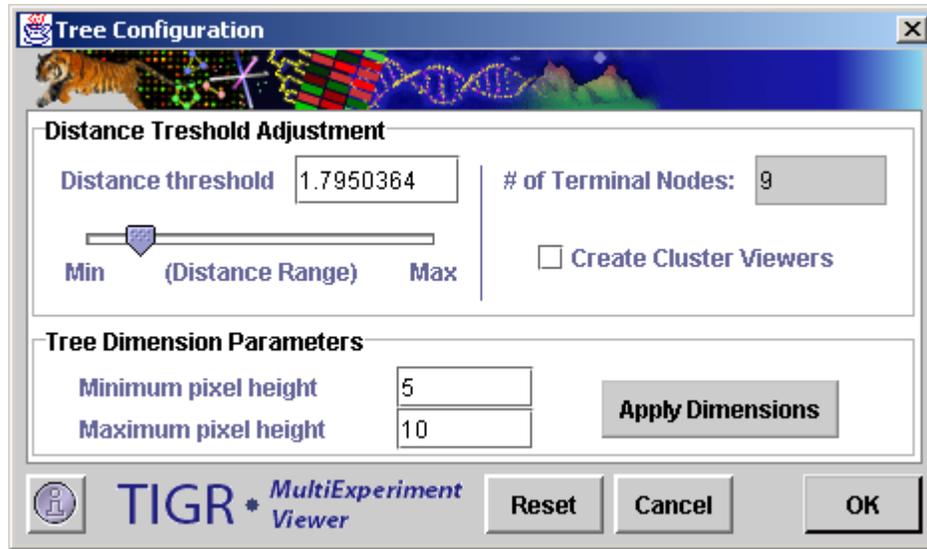
1. Clustering basically groups similar samples or genes together. This way it makes it much easier to test the data, since you can see which groups are more alike.
 - a. **HCL:** A commonly used clustering algorithm is Hierarchical Clustering, or HCL. From the top left click the *Clustering* drop down 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 you just want to look at the similarities among samples and not genes, simply uncheck the box at the top that says Gene Tree. If you do have a lot of genes and want to group them into similar groups, the k-means cluster in the next section may be more useful



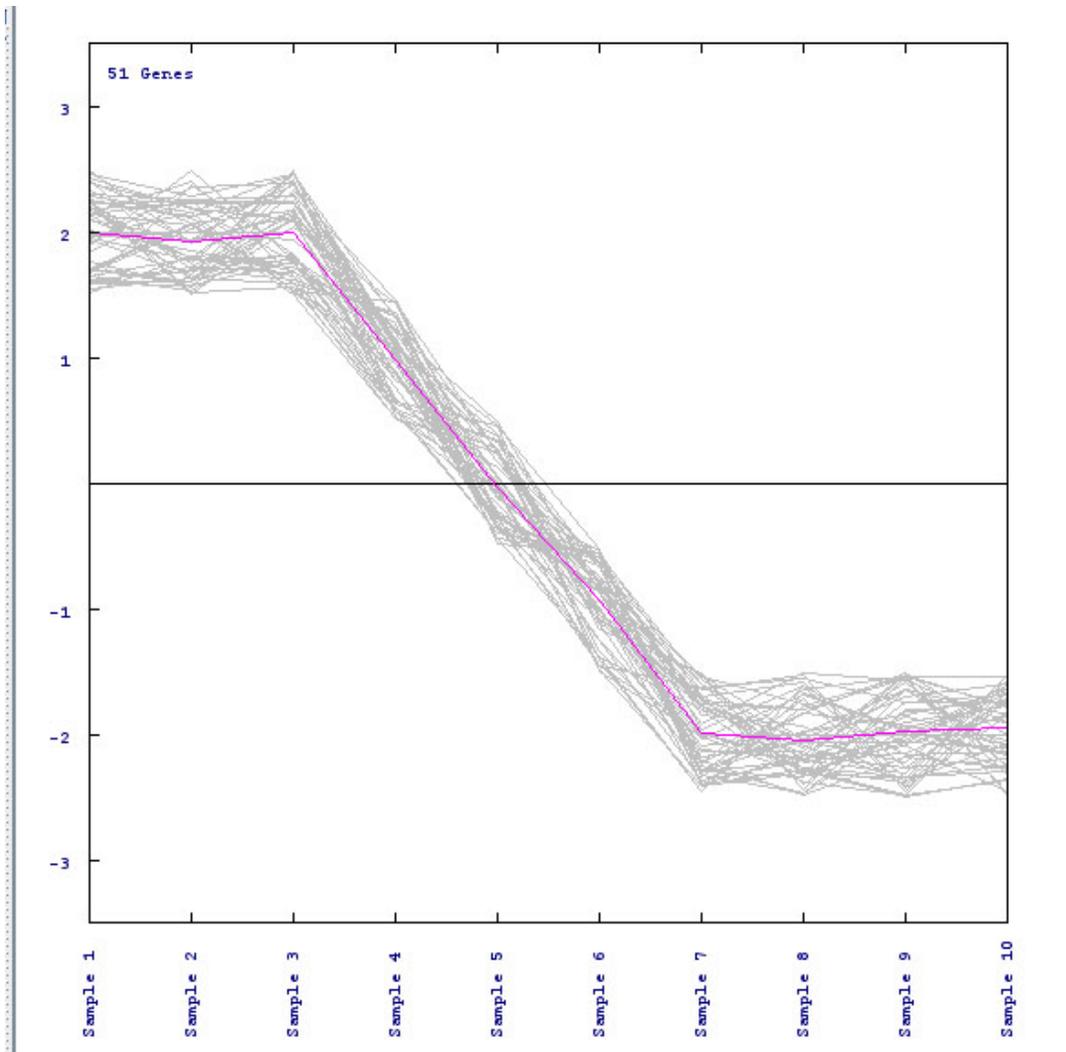
- b. Click *OK* at the bottom right.
- c. To view the clusters, find and double click *HCL* on the navigation tree on the left.
- d. You can store the clusters by clicking on the groups that you want together. That is, click the line that connects a certain group of genes. That group will then be highlighted.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



- e. The program should automatically choose a new color for each cluster so that you can easily view and distinguish them.

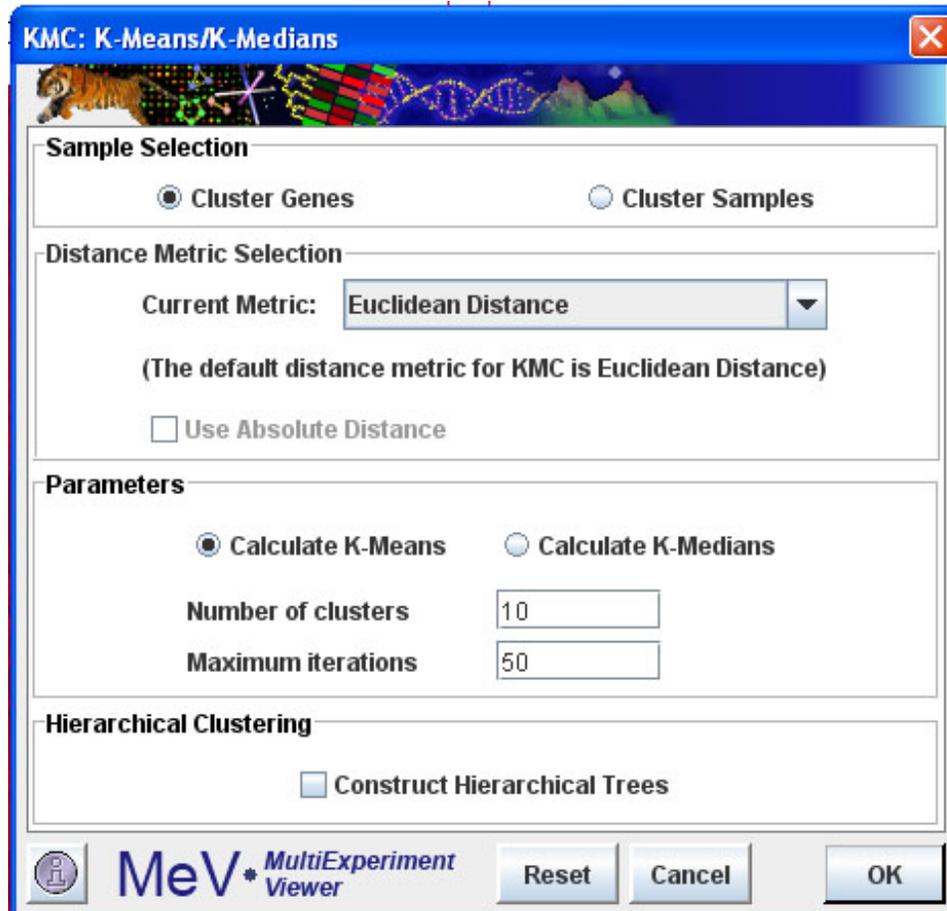


- g. Under *Distance Threshold Adjustment*, you can adjust how many clusters will be created. The lower the distance range, the more clusters there will be since a shorter distance between similar genes means more groups. You will see a light gray triangular ‘wedges’ along the side as this distance threshold changes. Either use the slide or type in a value for the distance threshold until you get the desired number of clusters.
- h. If you click the box on the right *Create Cluster Viewers*, you will get different graphs and charts for each cluster created. These can be very useful visualization tools as well being able to look closer at each cluster.
- i. Click *OK*.
- j. If you did not get the desired number of clusters, you can always go back. Also, if you navigation tree gets too cluttered, simply right click the item your wish to delete and click *Delete*.
- k. You can view each individual cluster in the navigation tree on the left. You can 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. If you wish to do an analysis on just one particular cluster (such as a t-test, which will be discussed in the Modules section), you can 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. If you double click Cluster Manager in the navigation tree, it will show you the list of all the clusters you made.
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 you can choose exactly how many clusters you want.
- a. Select *Clustering* → *k-Means/Medians Clustering* from the drop down menu.
 - b. Choose whether you want to cluster the genes or the samples.

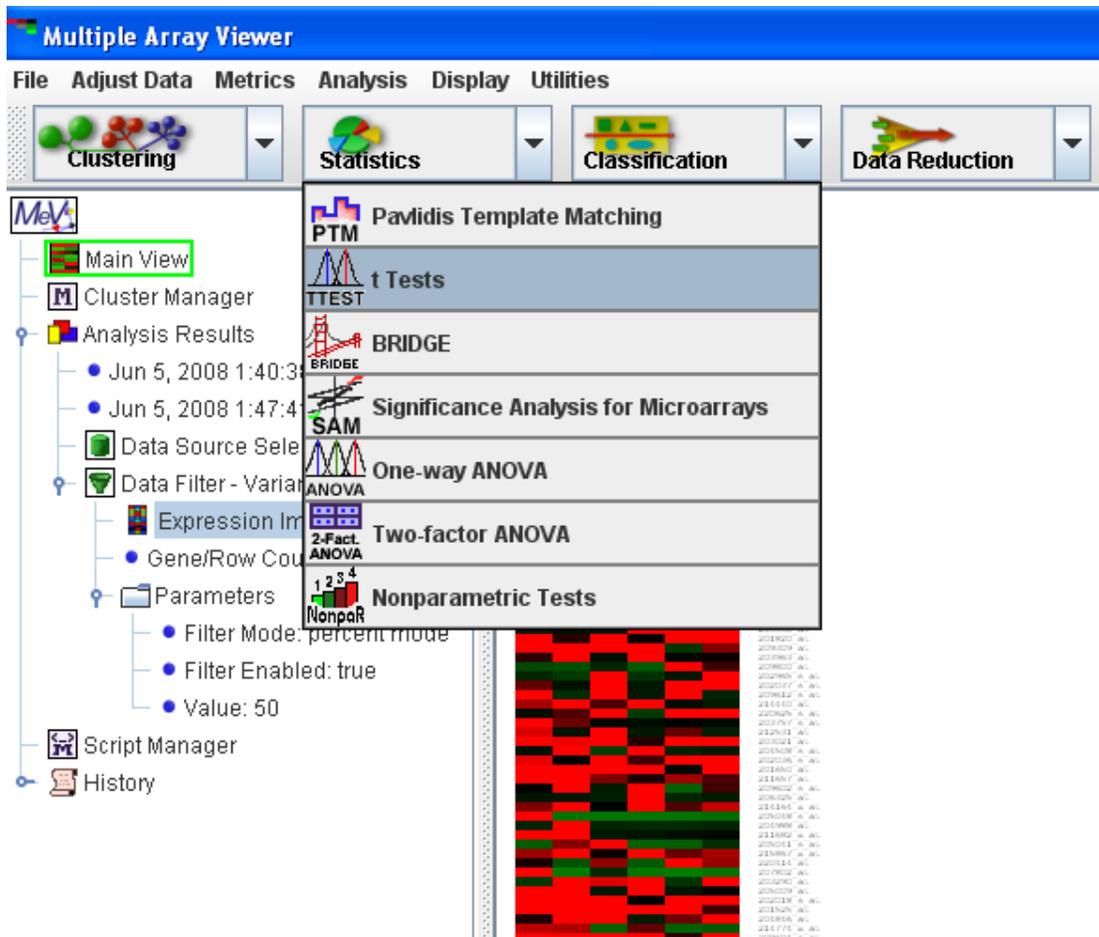


- c. Choose how many clusters you wish to form. Here we will use 10, but you can always go back and select fewer (or more) if those results do not give you the information you want.
 - d. You can view the clusters in the navigation tree on the left just as you could for the HCL.
3. For more information on clustering, there is a more detailed section in the Manual called Working with Clusters.

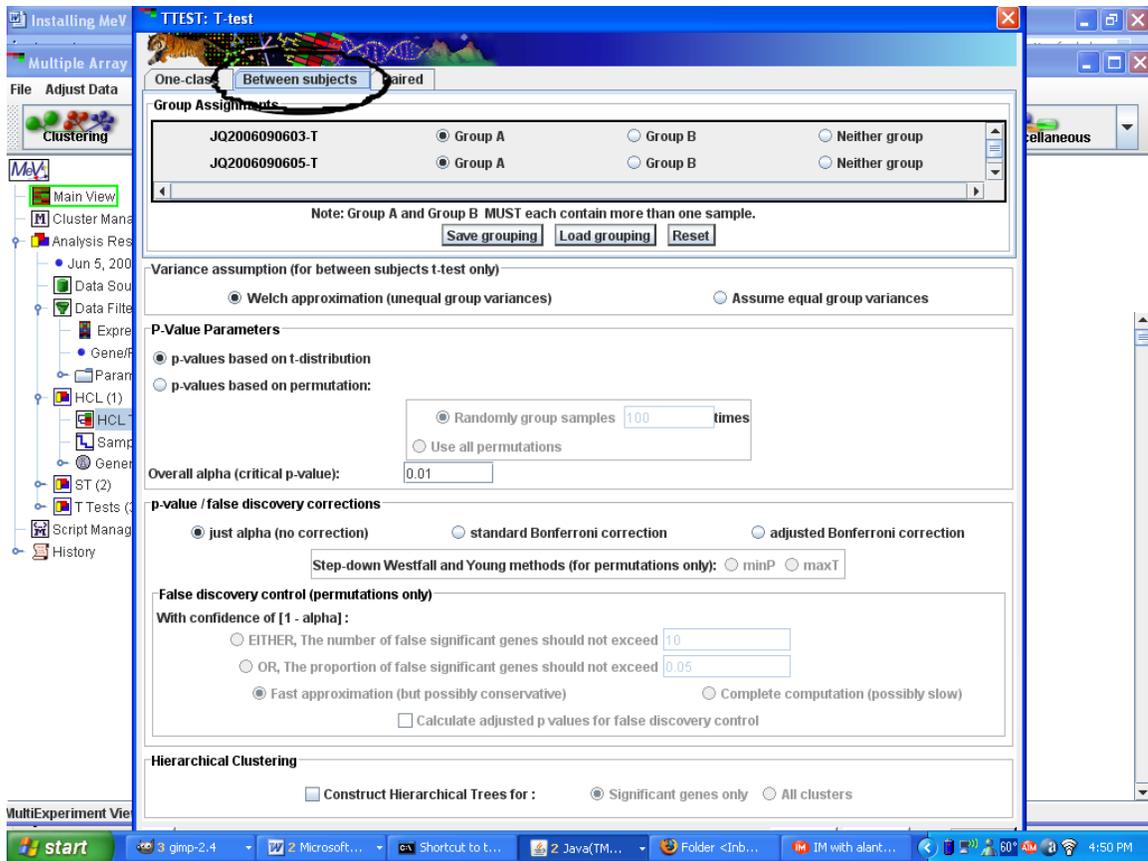
Modules

What do these do? They 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. You first want to 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. If you have other groups that you don't want to test against each other, you can mark them as Neither group.
5. You also want to 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>.