HierarchicalClusteringViewer Documentation

Module name:	HierarchicalClusteringViewer
Description:	Hierarchical clustering viewer that reads cdt, atr, and gtr files.
Author:	Joshua Gould (Broad Institute), gp-help@broad.mit.edu

Summary:
Reads in files in Eisen's tree view format and displays hierarchical trees. At least one of gtr filename and atr filename files is required. A cdt file is always required. For more information about these file formats see
http://www.broad.mit.edu/cancer/software/genepattern/tutorial/gp_fileformats.html.
The HierarchicalClustering module produces output that can be used as input to this module.

Expression values are displayed in a heat map format where the largest values are displayed as the reddest (hot), the smallest values are displayed as the bluest (cool), and intermediate values are a lighter color of either blue or red. This is a very convenient way to display three-dimensional data (samples, features, and expression values).

Display Options

There are a number of options controlling the appearance of the heat map that can be set by selecting the *Display Options* item in the *Edit* menu. You can choose between radio buttons for *Relative* and *Global* color schemes. The relative choice causes the colors for expression values to be scaled as a function of the number of standard deviations relative to the mean for each row. The global choice scales values using the minimum and maximum values in the entire dataset. To display the color legend, select the *Legend* item from the *View* menu.

You can choose whether to show or hide row names and descriptions and column names from the display by deselecting the corresponding check box. The row size and column size sliders control the size of each element in the heat map. When the *Maintain Square Aspect* checkbox is selected, the row and column sizes are kept in synch. The show grid checkbox controls whether a grid is drawn around each element. The *Show Profile* checkbox controls whether the profile is shown in the table for each feature. The *Feature Label Width* checkbox controls the size of feature labels, which you can load by selecting *Label Features* item from the *File* menu. Similarly, *Sample Label Height* checkbox controls the size of sample labels, which you can load by selecting *Label Samples* item from the *File* menu.

Use *Edit> Sample Dendrogram Branch Color* to set the color of a selected branch in the sample dendrogram. Similarly, to set the color for the feature tree, select *Edit > Feature Dendrogram Branch Color*. You can set the height of the sample and feature

dendrograms by changing the values of the *Sample Dendrogram Height* and *Feature Dendrogram Height* fields.

Saving Images
Users can save the heat map image to a file by selecting *File>Save Image*.

Finding Features
To find a feature in the heatmap, select *Edit>Find*. You can choose whether to match the case of the text you're searching for by selecting the *Match case* checkbox at the bottom of the find dialog. You can also choose whether to include the description column in your search by selecting the *Include Descriptions In Search* checkbox.

Expression Profiles
The expression profile for a feature plots expression value per sample.
To display an expression profile:
1. Select one or more features and optionally select one or more samples.
2. Click *View>Profile*. Alternatively, right-click and select *Profile* from the context menu or click the profile plot in the feature table.

Centroid Plot
The centroid plot shows the mean expression value for each sample. The error bars represent the standard deviation.
To display a centroid plot:
1. Select two or more features and optionally select one or more samples.
2. Click *View>Centroid Plot.* Alternatively, right-click and select *Centroid Plot* from the context menu.

Histogram
The histogram plot shows the distributions of expression values. The vertical axis represents number of occurrences. The horizontal axis represents the binned expression values.
To display a histogram:
3. Select one or more features and optionally select one or more samples.
4. Click *View>Histogram.* Alternatively, right-click and select *Histogram* from the context menu.

Saving Datasets
You can use the viewer to create a new dataset.
1. Select *File>Save Dataset*. A window appears.
2. Choose the features and samples to include in the dataset by selecting the features and sample names in the viewer.
3. Choose a location and name for the new dataset.
4. Click *Save* to save the new dataset.

Feature and Sample Labels
The viewer provides two labeling methods:

- Feature and sample labels use color to annotate features and samples in the heatmap

To use feature annotations:

1. Create a feature list file.
2. Select *File>Label Features* to open your feature list file. In the Feature column of the feature table, a color bar appears next to each feature in the feature list.
3. Select E*dit>Feature Labels* to edit the color or close the feature list:
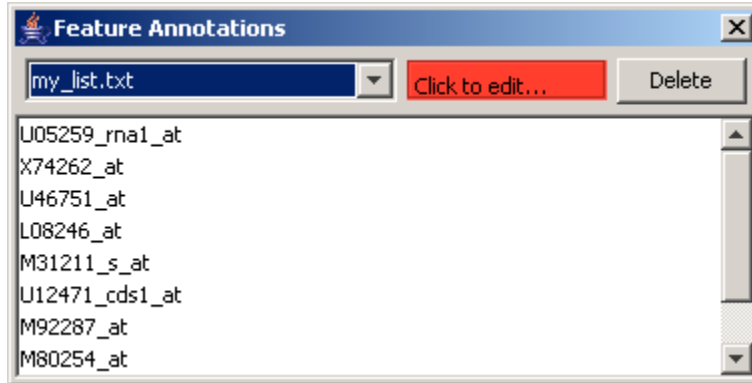
4. In the Feature Labels window, select your feature list from the drop-down list. The color assigned to that feature list appears in the box to the right.
   - To change the color, click the box and select a new color.
   - To close the feature list and remove the color bars from the table, click *Delete*.

To use sample annotations:

1. Create a cls file or sample info file.
2. Select *File>Label Samples* to open your cls or sample info file. A color bar appears below each sample name.
3. Edit and delete sample annotations as described above for feature labels

Loading Descriptions

GeneCruiser retrieves information about Affymetrix probe identifiers and adds the information to the feature table.

To use GeneCruiser annotations:

1. Select *File>Load Descriptions>GeneCruiser*.
2. Select the features that you want to retrieve annotations for in the table.
3. Choose which fields to retrieve from GeneCruiser in the GeneCruiser dialog.
4. The annotations appear in additional columns in the table.

Keyboard Shortcuts

You can use your keyboard to quickly accomplish many tasks. To find the shortcuts for common commands, look in the menus or select *Help>Keyboard Shortcuts* to see a list of available shortcuts

Parameters

cdt.filename	Clustered data table, contains the original data, but reordered

gtr.filename	Gene Tree file, records the order in which genes (rows) were joined (at least one of gtr filename and atr filename is required)
atr.filename	Array Tree file, records the order in which samples (columns) were joined (at least one of gtr filename and atr filename is required)