HierarchicalClusteringViewer Documentation

Description: Views results of hierarchical clustering.
Author: (Broad Institute), gp-help@broadinstitute.org

Summary:
Displays a heat map with hierarchical trees. At least one of the gtr file or atr file parameters is required. A cdt file is always required. A complete description of these formats is available on the GenePattern website. The HierarchicalClustering module produces output that can be used as input to this module.

Values are dislayed 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 values).

Display Options
There are a number of options controlling the appearance of the heat map that can be set by selecting *View>View Options*:

- You can choose between radio buttons for *Relative* and *Global* color schemes. When the relative choice is selected, the heat map converts values to colors using the mean, minimum, and maximum values in each row. When the global choice is selected the heat map converts values to colors using the mean, minimum, and maximum values in the entire dataset. To display the color legend, select *View>Color Scheme Legend*.
- You can choose between the radio buttons *Use Color Gradient* and *Use Discrete Colors*. The heat map uses linear color gradient to convert values to colors when the *Use Color* Gradient button is selected, while a discrete color map is used if the *Use Discrete Colors* button is selected. A color map can optionally be loaded by clicking on *Load Color Map*. This color map is a file containing three colors with one color per line. The first color is the color for the minimum values and the last color is the color for the maximum values. The colors can be specified as either a decimal, octal, hexadecimal integer, or r:g:b triplet (e.g. #9900FF, 0:0:0, 171).
- 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 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.
- You can choose to hide (show) feature descriptions, feature names and sample names by deselecting (selecting) the corresponding checkbox.
- The *Feature Label Width* slider controls the size of feature labels. Similarly, the *Sample Label Height* slider controls the size of sample labels. These controls do

not appear unless you have features and/or samples loaded. To load features, select *File>Label Features.* To load samples, select *File>Label Samples*.

- You can control the sample or feature dendrogram height by editing the corresponding text box.
- You can set the sample or feature dendrogram line thickness by editing the corresponding text box.

Saving Images

Users can save the heat map image to a file by selecting *File>Save Image*. Supported formats are bmp, eps, jpeg, png, and tiff.

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.

Dendrogram Color

Note: *Feature dendrograms are displayed when a .gtr file is provided and Sample dendrograms are displayed with an .atr file is provided.*

You can set the color of a selected branch in the sample or feature dendrogram branch color by selecting *Edit > Sample Dendrogram Branch Color* or *Edit>Feature Dendrogram Branch Color*. You can also set the color of a dendrogram node by editing the atr or gtr file. Add a column to the atr or gtr file that contains the node color. For example (added column shown in bold):

NODE1X	GENE78X	GENE77X	0.964702	**#9900FF**
NODE2X	GENE25X	GENE8X	0.963298	**#990000**

Dendrogram Rotation

You can flip a selected branch in the sample or feature dendrogram by selecting *View > Flip Sample Dendrogram* or *View > Flip Feature Dendrogram.*

Feature Labels

Feature labels use color to annotate features in the heat map.

To use feature labels:

1. Create a grp, gmx, or gmt file. A complete description of these formats is available on the GenePattern website.
2. Select *File>Label Features* to open your feature set file. You have the option to restrict the view to features in your feature set if you did not cluster the features in your dataset. To return to the full view, select *View>Reset Dataset.* A color bar appears next to each feature in the feature set in the table.
3. Select E*dit>Feature Labels* to edit the color or close the feature set.

4. In the Feature Labels window, select your feature set 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 delete the feature set and remove the color bars from the table, click *Delete*.

The feature annotations legend can be saved to a file by selecting *File>Save Feature Labels Legend*.

Sample Labels

Sample labels use color to annotate sample in the heat map.

To use sample labels:

1. Create a sample info file. A complete description of this format is available on the GenePattern website
2. Select *File>Label Samples* to open your sample info file. A color bar appears below each sample name.
3. Select E*dit>Sample Labels* to edit the color or close the sample info file.
4. In the Sample Labels window, select your sample class from the drop-down list. The color assigned to that sample class appears in the box to the right.
    - To change the color, click the box and select a new color.
    - To delete the sample class and remove the color bars from the table, click *Delete*.

The sample annotations legend can be saved to a file by selecting *File>Save Sample Labels Legend*.

Loading Descriptions

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

To use GeneCruiser annotations:

1. Select *File>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.

You can also load descriptions from a csv (comma separated values) file. To load descriptions from a csv file, select *File>Load Descriptions From Csv File.*

Finding Features

To find a feature in the heat map, 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 which column to search from by selecting the desired column name in the dropdown box.

Profile Plot

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 column 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:
1. Select one or more features and optionally select one or more samples.
2. Click *View>Histogram.* Alternatively, right-click and select *Histogram* from the context menu.

Nearest Neighbors
The nearest neighbors item shows other features whose expression values follow similar trends to the selected feature. There are four choices for the distance metric: Cosine, Euclidean, Manhattan, and Pearson.
To display the nearest neighbors:
1. Select one feature.
2. Click *View>Nearest Neighbors.* Alternatively, right-click and select *Nearest Neighbors* from the context menu.

Scatter Plot
The scatter plot lets you compare expression values in two samples.
To display a scatter plot:
1. Select two samples.
2. Optionally select one or more features.
3. Click *View>Scatter Plot.* Alternatively, right-click and select *Scatter Plot* from the context menu.

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.file	Clustered data table, contains the original data, but reordered
gtr.file	Gene Tree file, records the order in which genes (rows) were joined (at least one of gtr filename and atr filename is required)

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