

Dynamic Heatmap Viewer

BRB-ArrayTools Development Team

2019-07-09

This package creates a Dynamic Heatmap based on the Hierarchical Clustering method for visualizing microarray expression data. You can easily zoom in and out, change the color preferences, and mark genes and sample classes of the heatmap.

Quick Start

This package provides `test.dynamicHeatmap` for a quick heatmap viewing for the built-in sample dataset (i.e., “Brca”, “Perou”, and “Pomeroy”).

```
library(dynamicHeatmap)
test.dynamicHeatmap("Brca")
```

It launches a new window with a heatmap and a gene clustering dendrogram. If the samples are clustered, a sample clustering dendrogram will be shown.

Main Functions

Dynamic Heatmap Viewer is a powerful and versatile tool to help scientists to interactively visualize the microarray gene expression data. This guide provides a summary of options in the graphical user interface.

Gene ID/Array Label

The drop-down list displays the gene id (array label) from mouse-over gene (array).

Value

This shows the gene expression of the mouse-over array and gene combination. If centering and scaling option is checked in the main dialog, it will be applied on this value.

Class

A color bar representing different classes in arrays by different colors can be shown with this option.

Color Pref

Several predefined color preferences representing the expression data are available for users to select. Gray color represents missing values.

Color Scheme

Two color schemes are provided. The Saturated method will truncate gene expression at color limits in heatmap while Quantile method will use expression data quantiles as bins to represent the colors. Gray color represents missing values.

Color Limit

If Saturated method was selected (default), the heatmap displays gene expression data by truncating the expression at lower and upper thresholds. Users can adjust the upper/lower threshold by clicking the +/- buttons. The color legend on the right hand side of heatmap shows the current threshold values. By default,

if the minimum value of expression is less than 0, the limit is determined by the data; otherwise [max(-2, smallest data value), min(2, largest data value)] will be used. Summary Statistics: These statistics provide basic information about the data which could be useful for the selection of thresholds if 'Saturated' color scheme is used. The gene distance information is useful when cutting gene tree is performed.

Link to GeneCards

Right click on the heatmap, a browser will be opened and connected to GeneCards that contains more relevant information about the gene.

Zoom In

To zoom-in to a subset of genes, users can left click mouse button and select an area on the heatmap. The heatmap will be re-drawn the desired region after the mouse is released. If the plot was zoomed in to show less than 50 genes, gene symbols or unique IDs will be shown along the plot on the right hand side.

Reset Zoom-In

The heatmap will be re-drawn using all genes. The keyboard shortcut is Alt+R.

Zoom Out

To go back to the heatmap before zooming in, click on the Zoom Out button.

Print to PDF

This function allows users to save the plot in a high-quality Portable Document Format (PDF) format. Alternatively, users can use Windows keyboard shortcut Alt + PrtSc to copy the current heatmap to a clipboard. Then the graph in the clipboard can be pasted to a MS-WORD document or other image processing software.

Highlight Genes

To highlight genes in the heatmap, users can choose from one of three options (the selected gene identifier will be used to match the gene labels if users choose the option #1).

- Enter gene labels to search for. The comma sign can be used to separate multiple genes.
- Browse a genelist file. The first column in the genelist file will be used to match.
- Browse a Biocarta/KEGG pathway.

After clicking the OK button, genes containing the desired labels will be highlighted using purple circle dots located on the right hand side of the heatmap. To erase the dots, select either option 1 or 2. Then press the Clear and then the OK button.

Cut Gene Tree

This will cut gene dendrogram and save the gene IDs in each individual cluster to a text file. For example, if users select 3 clusters or a distance which results in 3 clusters, there will be 3 tab-delimited text files generated (cluster1.txt, cluster2.txt and cluster3.txt) under the output folder where each text file contains gene IDs from each individual cluster. This function will be disabled if the zoom-in has been applied.

Save Genes

This will save the gene IDs from currently selected genes to a text file.

Options

- Show gene dendrogram: check/uncheck the box to show/hide the gene dendrogram.
- Save color preference and scheme: check the box to save the color preference and scheme for the next launch.
- Choice of Y label: which gene label will be used on the Y-axis.
- Choice of X label: which array ID will be shown on the X-axis.

Tips

1. The heatmap displays array IDs at the bottom of the heatmap. The array IDs are obtained from the first column of the Experiment Descriptor (this can be changed from the Options dialog). They are displayed only if the total number of arrays is less than or equal to 200.
2. The heatmap displays gene symbol at the right hand side of the heatmap. If gene symbols are not available, the first column of gene identifiers worksheet will be used (this can be changed from the Options dialog). The gene symbols are only displayed if the number of genes is less than or equal to 100.
3. If the number of classes is more than 10, the colors for the class color bar will be re-used.
4. The executable program Qheatmap.exe can be used as a stand alone application. When users run the program without specifying a working directory, the program will pop up a dialog asking the working directory. The working directory should contain input files expression.txt (gene expression), GeneIDs.txt (gene annotation), ExpDescWkSht.txt (experiment descriptor), options.txt (options such as linkage method and distance), order.txt (orderGene.txt), merge.txt (mergeGene.txt), and height.txt (heightGene.txt) for plotting array and gene dendrograms.

Clustering Methods

`dynamicHeatmap` is the main R function to perform the clustering calculations with the function `hclust()`.

Hierarchical clustering parameters:

Analysis options

- Center and scale genes: If this option is selected then each gene will be mean-centered and standard-deviation-scaled across all the experiments which have been selected for this analysis.
 - Center genes: If this option is selected then each gene will be mean-centered across all the experiments which have been selected for this analysis.
 - None: If this option is selected then each gene will not be mean-centered and standard-deviation-scaled across all the experiments which have been selected for this analysis.

Metric

Choose “1-correlation” or Euclidean distance as the distance matrix.

Linkage

Choose from “Average linkage” Complete linkage Single linkage or “Ward linkage (ward.D2)” to determine the method of computing the distance between two clusters. Average linkage takes the average of the distances between each possible pair of genes from the two clusters whereas complete linkage takes the maximum and single linkage takes the minimum of the distances between each possible pair of genes from the two clusters.

Select genes for analysis

The analysis may be based on all genes in the dataset or a subset of those genes. Subsets of genes may be selected for inclusion or exclusion from the analysis. Gene subsets can be specified in the ‘Genelists’ folder.

Select experiments for analysis

The analysis may be based on all experiments in the dataset or a subset of those experiments. A subset of experiments may be specified by specifying an experiment descriptor variable and any experiment for which that experiment descriptor variable is blank will be excluded from the analysis.

Plotting-order of experiments is based on

Experiments may be ordered by the dendrogram-order of a hierarchical cluster analysis performed on the entire filtered set of log-ratios. In this case all the cluster lineplots will have the same experiment order. Another option is to order the experiments based on an ordering variable selected from the experiment descriptors. If the ordering variable is categorical then the experiments will be sorted by alphabetical-order of the categorical values.

Data Input

`dynamicHeatmap` is the main R function to generate the heatmaps. In this section, we will look into details about how to prepare inputs for `dynamicHeatmap`. Once again, we use the “Brca” sample data for an example. The package contains the following “Brca” sample information:

*Brca_LOGRAT.txt : a table of expression data with rows representing genes and columns representing samples;

*Brca_FILTER.TXT: a list of filtering information, where 1 means the corresponding gene passes the filters while 0 means it is excluded from analysis;

*Brca_GENEID.txt: a table of gene information corresponding to row information of Brca_LOGRAT.txt and Brca_FILTER.TXT;

*Brca_EXPDESIGN.txt: a table with class information AND/OR separate test set information.

There are a total of 22 samples in the dataset. We run the following code to obtain objects like `exprData` as inputs to `dynamicHeatmap`.

```
dataset<-"Brca"
# Gene IDs
geneId <- read.delim(system.file("extdata", paste0(dataset, "_GENEID.txt")
, package = "dynamicHeatmap"), as.is = TRUE, colClasses = "character")
# Expression data, and here are log ratio.
x <- read.delim(system.file("extdata", paste0(dataset, "_LOGRAT.TXT")
, package = "dynamicHeatmap"), header = FALSE)
# Gene filter information, 1 - pass the filter, 0 - filtered
geneFilter <- scan(system.file("extdata", paste0(dataset, "_FILTER.TXT")
, package = "dynamicHeatmap"), quiet = TRUE)
# Class information
expDesign <- read.delim(system.file("extdata", paste0(dataset, "_EXPDESIGN.txt")
, package = "dynamicHeatmap"), as.is = TRUE)
# Filter out genes.
geneId <- geneId[geneFilter == 1, ]
x <- x[geneFilter == 1, ]
# Pick the first column as the array IDs.
exprData <- x
colnames(exprData) <- expDesign[, 1]
```

`exprData` is a 2009*22 log ratio matrix with rows representing 2009 genes and columns representing 22 samples. Before filtering, it has 3226 rows.

```
##          s1321      s1996      s1822      s1714      s1224      s1252      s1510
s1900
## 1 -1.39854932 -3.0817938 -2.7303929 -1.8744690 -2.2882450 -0.3453870 -1.42321134
```

```

-1.7776077
## 2 0.39940688 0.2781018 -0.2011399 -0.5334322 -0.5792937 -0.2874397 -0.88264304
-0.4150376
## 3 -0.02509096 0.4375801 0.1047962 0.9533499 -0.2205003 0.3532323 -0.67318958
0.5109619
## 4 -0.13006058 -0.8389376 -0.2356283 0.6195197 0.8122152 -0.4181434 -0.52509099
0.2630344
## 6 -0.46566358 -0.6667566 -0.6199690 0.4760281 0.1093474 -0.6036991 0.04809438
-0.6214885
## .....

```

geneFilter is a 1/0 integer vector with length 3226 for genes. Here 2009 genes are 1, and 217 genes are 0. After filtering, 2009 genes will be left.

```

## [1] 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 0 1 0 1 0 1 1 1 1 0 0 0 0 0 1 1 0 1 1 1 1 0 1 1
1 0 0
## [46] 1 1 0 0 0 0 0 1 1 1 1 0 0 1 1 1 0 1 0 0 1 0 1 1 1 0 0 1 1 1 1 1 1 1 0 1 1 1 0 0 0 0
1 1 1
## [91] 0 1 1 0 1 0 0 0 1 1 1 1 1 0 0 1 1 0 1 1 1 1 1 1 1 0 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1 1
0 0 0
## [136] 1 0 0 1 1 0 0 1 0 0 1 1 1 1 0 0 0 1 0 1 1 1 1 1 0 1 0 0 1 1 0 0 1 1 1 0 0 1 1 1 0
1 1 0
## [181] 1 0 1 1 1 0 1 1 0 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1
1 1 1
## .....

```

Then generate the heatmap:

```

projectPath <- tempdir()
outputName = "dynamicHeatmapBrca"
dynamicHeatmap(exprData, expDesign, geneId,
               analysisOptions = "CenterAndScaleGenes",
               # "CenterAndScaleGenes", "CenterGenes", "None"
               metric = "1-Correlation", # "1-Correlation", "Euclidean"
               linkage = "Average", # "Average", "Complete", "Single", "Ward"
               sortSamplesByClass = FALSE,
               sortSamplesClassName = "BRCA1.v.notBRCA1",
               useSamplesCenteredCorrelation = FALSE,
               projectPath,
               outputName)

```

A new window with a heatmap and a gene clustering dendrogram will pop up. For more details about dynamicHeatmap, please type help("dynamicHeatmap") in the R console.