Package ‘moduleColor’

December 11, 2007

Version 1.02

Date 2007-11-21

Title Basic module functions

Author Peter Langfelder <Peter.Langfelder@gmail.com> and Steve Horvath <SHorvath@mednet.ucla.edu>

Maintainer Peter Langfelder <Peter.Langfelder@gmail.com>

Depends R, stats, impute, grDevices, dynamicTreeCut

ZipData no

License GPL version 2 or newer

Description Methods for color labeling, calculation of eigengenes, merging of closely related modules.

URL http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/

R topics documented:

  checkSets .................................................. 2
  collectGarbage ........................................... 3
  consensusMEDissimilarity ............................... 3
  consensusOrderMEs ........................................ 4
  fixDataStructure .......................................... 5
  labels2colors ............................................. 6
  mergeCloseModules ....................................... 7
  moduleColor-package .................................... 9
  moduleEigengenes ....................................... 10
  moduleNumber ............................................ 11
  multiSetMEs .............................................. 12
  normalizeLabels ......................................... 14
  orderMEs .................................................. 14
  plotHclustColors ....................................... 16
  standardColors .......................................... 17
Description

Checks whether given sets have the correct format and retrieves dimensions.

Usage

\texttt{checkSets(data, checkStructure = FALSE, useSets = NULL)}

Arguments

\texttt{data} A vector of lists; in each list there must be a component named \texttt{data} whose content is a matrix or dataframe or array of dimension 2.

\texttt{checkStructure} If \texttt{FALSE}, incorrect structure of \texttt{data} will trigger an error. If \texttt{TRUE}, an appropriate flag (see output) will be set to indicate whether \texttt{data} has correct structure.

\texttt{useSets} Optional specification of entries of the vector \texttt{data} that are to be checked. Defaults to all components. This may be useful when \texttt{data} only contains information for some of the sets.

Details

For multiset calculations, many quantities (such as expression data, traits, module eigengenes etc) are presented by a common structure, a vector of lists (one list for each set) where each list has a component \texttt{data} that contains the actual (expression, trait, eigengene) data for the corresponding set in the form of a dataframe. This function checks whether \texttt{data} conforms to this convention and retrieves some basic dimension information (see output).

Value

A list with components

\texttt{nSets} Number of sets (length of the vector \texttt{data}).

\texttt{nGenes} Number of columns in the \texttt{data} components in the lists. This number must be the same for all sets.

\texttt{nSamples} A vector of length \texttt{nSets} giving the number of rows in the \texttt{data} components.

\texttt{structureOK} Only set if the argument \texttt{checkStructure} equals \texttt{TRUE}. The value is \texttt{TRUE} if the parameter \texttt{data} passes a few tests of its structure, and \texttt{FALSE} otherwise. The tests are not exhaustive and are meant to catch obvious user errors rather than be bulletproof.

Author(s)

Peter Langfelder, (Peter.Langfelder@gmail.com)
**collectGarbage**  
*Iterative garbage collection.*

**Description**
Performs garbage collection until free memory indicators show no change.

**Usage**
```
collectGarbage()
```

**Value**
None.

**Author(s)**
Steve Horvath

---

**consensusMEDissimilarity**  
*Consensus dissimilarity of module eigengenes.*

**Description**
Calculates consensus dissimilarity \((1 - \text{cor})\) of given module eigengenes realized in several sets.

**Usage**
```
consensusMEDissimilarity(MEs, useAbs = FALSE, useSets = NULL, method = "consensus")
```

**Arguments**
- **MEs**
  Module eigengenes of the same modules in several sets.
- **useAbs**
  Controls whether absolute value of correlation should be used instead of correlation in the calculation of dissimilarity.
- **useSets**
  If the consensus is to include only a selection of the given sets, this vector (or scalar in the case of a single set) can be used to specify the selection. If `NULL`, all sets will be used.
- **method**
  A character string giving the method to use. Allowed values are (abbreviations of) "consensus" and "majority". The consensus dissimilarity is calculated as the minimum of given set dissimilarities for "consensus" and as the average for "majority".
Details
This function calculates the individual set dissimilarities of the given eigengenes in each set, then takes the (parallel) maximum or average over all sets. For details on the structure of input data, see checkSets.

Value
A dataframe containing the matrix of dissimilarities, with names and rownames set appropriately.

Author(s)
Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

See Also
checkSets

consensusOrderMEs  Put close eigenvectors next to each other in several sets.

Description
Reorder given (eigen-)vectors such that similar ones (as measured by correlation) are next to each other. This is a multi-set version of orderMEs; the dissimilarity used can be of consensus type (for each pair of eigenvectors the consensus dissimilarity is the maximum of individual set dissimilarities over all sets) or of majority type (for each pair of eigenvectors the consensus dissimilarity is the average of individual set dissimilarities over all sets).

Usage
consensusOrderMEs(MEs, useAbs = FALSE, useSets = NULL,
greyLast = TRUE, greyName = "MEgrey", method = "consensus")

Arguments
MEs  Module eigengenes of several sets in a multi-set format (see checkSets). A vector of lists, with each list corresponding to one dataset and the module eigengenes in the component data, that is MEs[[set]]$data[sample, module] is the expression of the eigengene of module module in sample sample in dataset set. The number of samples can be different between the sets, but the modules must be the same.

useAbs  Controls whether vector similarity should be given by absolute value of correlation or plain correlation.

useSets  Allows the user to specify for which sets the eigengene ordering is to be performed.
greyLast  Normally the color grey is reserved for unassigned genes; hence the grey module
is not a proper module and it is conventional to put it last. If this is not desired,
set the parameter to FALSE.
greyName  Name of the grey module eigengene.
method    A character string giving the method to be used calculating the consensus dis-
similarity. Allowed values are (abbreviations of) "consensus" and "majority". The
consensus dissimilarity is calculated as the maximum of given set dissimi-
larities for "consensus" and as the average for "majority".

Details

Ordering module eigengenes is useful for plotting purposes. This function calculates the consensus
or majority dissimilarity of given eigengenes over the sets specified by useSets (defaults to all
sets). A hierarchical dendrogram is calculated using the dissimilarity and the order given by the
dendrogram is used for the eigengenes in all other sets.

Value

A vector of lists of the same type as MEs containing the re-ordered eigengenes.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

See Also

moduleEigengenes, multiSetMEs, orderMEs

fixDataStructure  Put single-set data into a form useful for multiset calculations.

Description

Encapsulates single-set data in a wrapper that makes the data suitable for functions working on
multiset data collections.

Usage

fixDataStructure(data, verbose = 0, indent = 0)

Arguments

data    A dataframe, matrix or array with two dimensions to be encapsulated.
verbose  Controls verbosity. 0 is silent.
indent   Controls indentation of printed progress messages. 0 means no indentation,
every unit adds two spaces.
Details

For multiset calculations, many quantities (such as expression data, traits, module eigengenes etc) are presented by a common structure, a vector of lists (one list for each set) where each list has a component `data` that contains the actual (expression, trait, eigengene) data for the corresponding set in the form of a dataframe. This function creates a vector of lists of length 1 and fills the component `data` with the content of parameter `data`.

Value

As described above, input data in a format suitable for functions operating on multiset data collections.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

See Also

`checkSets`

Examples

```r
singleSetData = matrix(rnorm(100), 10,10);
encapsData = fixDataStructure(singleSetData);
length(encapsData)
names(encapsData[[1]])
dim(encapsData[[1]]$data)
all.equal(encapsData[[1]]$data, singleSetData);
```

---

`labels2colors`  
*Convert numerical labels to colors.*

Description

Converts a vector of numerical labels into a vectors of colors corresponding to the labels.

Usage

```r
labels2colors(labels, zeroIsGrey = TRUE, colorSeq = NULL)
```

Arguments

- **labels**: Vector of non-negative integer labels.
- **zeroIsGrey**: If TRUE, labels 0 will be assigned color grey. Otherwise, labels below 1 will trigger an error.
- **colorSeq**: Color sequence corresponding to labels. If not given, a standard sequence will be used.
mergeCloseModules

Details

The standard sequence start with well-distinguishable colors, and after about 40 turns into a quasi-
random sampling of all colors available in R with the exception of all shades of grey (and gray).

Value

A vector of character strings of the same length as the input vector.

Author(s)

Peter Langfelder, (Peter.Langfelder@gmail.com)

Examples

labels = c(0:20);
labels2colors(labels);

mergeCloseModules  Merge close modules of gene expression data.

Description

Merges modules in gene expression networks that are too close as measured by the correlation of
their eigengenes.

Usage

mergeCloseModules(exprData, colors, cutHeight = 0.2,
MEs = NULL, useAbs = FALSE, iterate = TRUE, relabel = FALSE,
colorSeq = NULL, getNewMEs = TRUE, useSets = NULL,
checkDataFormat = TRUE, verbose = 1, indent = 0)

Arguments

exprData  Expression data in a multi-set format (see checkSets).
colors  A vector giving module colors for genes. The method only makes sense when
genes have the same color label in all sets, hence a single vector.
cutHeight  Maximum dissimilarity (i.e., 1-correlation) that qualifies modules for merging.
MEs  If module eigengenes have been calculated before, the user can save some com-
putational time by inputting them. If they are not given, they will be calculated.
useAbs  Specifies whether absolute value of correlation or plain correlation (of module
eigengenes) should be used in calculating module dissimilarity.
iterate  Controls whether the merging procedure should be repeated until there is no
change. If FALSE, only one iteration will be executed.
relabel  Controls whether, after merging, color labels should be ordered by module size.
mergeCloseModules

**colorSeq**  
Color labels to be used for relabeling. Defaults to the standard color order used in this package.

**getNewMEs**  
Controls whether module eigengenes of merged modules should be calculated and returned.

**useSets**  
A vector of scalar allowing the user to specify which sets will be used to calculate the consensus dissimilarity of module eigengenes. Defaults to all sets.

**checkDataFormat**  
If TRUE, the function will check `exprData` and `MEs` for correct multi-set structure. If single set data is given, it will be converted into a format useable for the function. If FALSE, incorrect structure of input data will trigger an error.

**verbose**  
Controls verbostity of printed progress messages. 0 means silent, up to (about) 5 the verbostity gradually increases.

**indent**  
A single non-negative integer controling indentation of printed messages. 0 means no indentation, each unit above that adds two spaces.

**Details**

This function returns the color labels for modules that are obtained from the input modules by merging ones that are closely related. The relationships are quantified by correlations of module eigengenes; a “consensus” measure is defined as the minimum over the corresponding relationship in each set. Once the (dis-)similarity is calculated, average linkage hierarchical clustering of the module eigengenes is performed, the dendrogram is cut at the height `cutHeight` and modules on each branch are merged. Optionally, the process is repeated until no more modules are merged.

**Value**

A list with components

- **colors**  
  Color labels for the genes corresponding to merged modules.

- **dendro**  
  Hierarchical clustering dendrogram (average linkage) of the eigengenes of the most recently computed tree. If `iterate` was set TRUE, this will be the dendrogram of the merged modules, otherwise it will be the dendrogram of the original modules.

- **oldDendro**  
  Hierarchical clustering dendrogram (average linkage) of the eigengenes of the original modules.

- **cutHeight**  
  The input `cutHeight`.

- **oldMEs**  
  Module eigengenes of the original modules in the sets given by `useSets`.

- **newMEs**  
  Module eigengenes of the merged modules in the sets given by `useSets`.

**Author(s)**

Peter Langfelder, (Peter.Langfelder@gmail.com)
moduleColor-package

Basic module functions

Description

Methods for color labeling, calculation of eigengenes, merging of closely related modules.

Details

Package: moduleColor
Version: 1.02
Date: 2007-11-21
Depends: R, stats, impute, grDevices, dynamicTreeCut
ZipData: no
License: GPL version 2 or newer
URL: http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/
Packaged: Wed Nov 21 13:02:08 2007; plangfelder
Built: R 2.4.1; ; 2007-11-21 13:02:57; unix

Index:

checkSets Retrieve basic sizes of a group of datasets.
collectGarbage Iterative garbage collection.
consensusMEDissimilarity Consensus dissimilarity of module eigengenes.
consensusOrderMEs Put close eigenvectors next to each other in several sets.
fixDataStructure Put single-set data into a form useful for multiset calculations.
labels2colors Convert numerical labels to colors.
mergeCloseModules Merge close modules of gene expression data.
moduleColor-package Basic module functions ~~ package title ~~
moduleEigengenes Calculate module eigengenes.
moduleNumber Fixed-height cut of a dendrogram.
multiSetMEs Calculate module eigengenes.
normalizeLabels Transform numerical labels into normal order.
orderMEs Put close eigenvectors next to each other
plotHclustColors Plot color bars corresponding to modules
standardColors Colors this library uses for labeling modules.

Author(s)

Peter Langfelder <Peter.Langfelder@gmail.com> and Steve Horvath <SHorvath@mednet.ucla.edu>
Maintainer: Peter Langfelder <Peter.Langfelder@gmail.com>
moduleEigengenes

Calculate module eigengenes.

Description

Calculates module eigengenes (1st principal component) of modules in a given single dataset.

Usage

moduleEigengenes(expr, colors, impute = TRUE, nPC = 1, align = "along average", verbose = 0, indent = 0)

Arguments

- **expr**: Expression data for a single set in the form of a dataframe where rows are samples and columns are genes (probes).
- **colors**: A vector of the same length as the number of probes in expr, giving module color for all probes (genes). Color "grey" is reserved for unassigned genes.
- **impute**: If TRUE, expression data will be checked for the presence of NA entries and if the latter are present, numerical data will be imputed, using function impute.knn and probes from the same module as the missing datum. The function impute.knn uses a fixed random seed giving repeatable results.
- **nPC**: Number of principal components to be calculated. If only eigengenes are needed, it is best to set it to 1 (default). If variance explained is needed as well, use value NULL. This will cause all principal components to be computed, which is slower.
- **align**: Controls whether eigengenes, whose orientation is undetermined, should be aligned with average expression (align = "along average", the default) or left as they are (align = " "). Any other value will trigger an error.
- **verbose**: Controls verbosity of printed progress messages. 0 means silent, up to (about) 5 the verbosity gradually increases.
- **indent**: A single non-negative integer controlling indentation of printed messages. 0 means no indentation, each unit above that adds two spaces.
moduleNumber

Details

Module eigengene is defined as the first principal component of the expression matrix of the corresponding module.

Value

A list with the following components:

- **eigengenes**: Module eigengenes in a dataframe, with each column corresponding to one eigengene. The columns are named by the corresponding color with an "ME" prepended, e.g., MEturquoise etc.
- **averageExpr**: If align == "along average", a dataframe containing average normalized expression in each module. The columns are named by the corresponding color with an "AE" prepended, e.g., AEturquoise etc.
- **varExplained**: A dataframe in which each column corresponds to a module, with the component varExplained[PC, module] giving the variance of module module explained by the principal component no. PC. This is only accurate if all principal components have been computed (input nPC = NULL). At most 5 principal components are recorded in this dataframe.
- **nPC**: A copy of the input nPC.

Author(s)

Steve Horvath (SHorvath@mednet.ucla.edu), Peter Langfelder (Peter.Langfelder@gmail.com)

References


Description

Fixed-height cut of a dendrogram.

Detects branches of on the input dendrogram by performing a fixed-height cut.

Usage

moduleNumber(dendro, cutHeight = 0.9, minSize = 50)

Arguments

dendro: a hierarchical clustering dendrogram such as one returned by hclust.
cutHeight: Maximum joining heights that will be considered.
minSize: Minimum cluster size.
multiSetMEs

Details

All contiguous branches below the height `cutHeight` that contain at least `minSize` objects are assigned unique positive numerical labels; all unassigned objects are assigned label 0.

Value

A vector of numerical labels giving the assignment of each object.

Note

The numerical labels may not be sequential. See `normalizeLabels` for a way to put the labels into a standard order.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

See Also

`hclust`, `cutree`, `normalizeLabels`

---

`multiSetMEs` *Calculate module eigengenes.*

Description

Calculates module eigengenes for several sets.

Usage

```r
multiSetMEs(exprData, colors, universalColors = NULL, useSets = NULL, impute = TRUE, nPC = 1, align = "along average", verbose = 1, indent = 0)
```

Arguments

- `exprData` Expression data in a multi-set format (see `checkSets`). A vector of lists, with each list corresponding to one microarray dataset and expression data in the component `data`, that is `expr[[set]]$data[sample, probe]` is the expression of probe `probe` in sample `sample` in dataset `set`. The number of samples can be different between the sets, but the probes must be the same.
- `colors` A matrix of dimensions (number of probes, number of sets) giving the module assignment of each gene in each set. The color "grey" is interpreted as unassigned.
- `universalColors` Alternative specification of module assignment. A single vector of length (number of probes) giving the module assignment of each gene in all sets (that is the modules are common to all sets). If given, takes precedence over `color`.
multisetMEs

useSets If calculations are requested in (a) selected set(s) only, the set(s) can be specified here. Defaults to all sets.

impute Logical. If TRUE, expression data will be checked for the presence of NA entries and if the latter are present, numerical data will be imputed, using function `impute.knn` and probes from the same module as the missing datum. The function `impute.knn` uses a fixed random seed giving repeatable results.

nPC Number of principal components to be calculated. If only eigengenes are needed, it is best to set it to 1 (default). If variance explained is needed as well, use value NULL. This will cause all principal components to be computed, which is slower.

align Controls whether eigengenes, whose orientation is undetermined, should be aligned with average expression (align = "along average", the default) or left as they are (align = ""). Any other value will trigger an error.

verbose Controls verbosity of printed progress messages. 0 means silent, up to (about) 5 the verbosity gradually increases.

indent A single non-negative integer controlling indentation of printed messages. 0 means no indentation, each unit above that adds two spaces.

Details

See `moduleEigengenes` for details on calculation of module eigengenes in individual sets. This function simply calls `moduleEigengenes` for each set represented in `exprData`.

Value

A vector of lists similar in spirit to the input `exprData`. For each set there is a list with the following components:

data Module eigengenes in a dataframe, with each column corresponding to one eigengene. The columns are named by the corresponding color with an "ME" prepended, e.g., MEturquoise etc.

averageExpr If `align` == "along average", a dataframe containing average normalized expression in each module. The columns are named by the corresponding color with an "AE" prepended, e.g., AEturquoise etc.

varExplained A dataframe in which each column corresponds to a module, with the component `varExplained[PC, module]` giving the variance of module `module` explained by the principal component no. `PC`. This is only accurate if all principal components have been computed (input `nPC = NULL`). At most 5 principal components are recorded in this dataframe.

nPC A copy of the input `nPC`.

Author(s)

Peter Langfelder, {Peter.Langfelder@gmail.com}

See Also

`moduleEigengenes`
normalizeLabels

*Transform numerical labels into normal order.*

Description

Transforms numerical labels into normal order, that is the largest group will be labeled 1, next largest 2 etc. Label 0 is optionally preserved.

Usage

```r
normalizeLabels(labels, keepZero = TRUE)
```

Arguments

- `labels` Numerical labels.
- `keepZero` If `TRUE` (the default), labels 0 are preserved.

Value

A vector of the same length as input, containing the normalized labels.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

orderMEs

*Put close eigenvectors next to each other*

Description

Reorder given (eigen-)vectors such that similar ones (as measured by correlation) are next to each other.

Usage

```r
orderMEs(MEs, greyLast = TRUE, greyName = "MEgrey", orderBy = 1, order = NULL, useSets = NULL, verbose = 0, indent = 0)
```
**Arguments**

- **MEs**
  Module eigengenes in a multi-set format (see `checkSets`). A vector of lists, with each list corresponding to one dataset and the module eigengenes in the component `data`, that is `MEs[[set]]$data[sample, module]` is the expression of the eigengene of module `module` in sample `sample` in dataset `set`. The number of samples can be different between the sets, but the modules must be the same.

- **greyLast**
  Normally the color grey is reserved for unassigned genes; hence the grey module is not a proper module and it is conventional to put it last. If this is not desired, set the parameter to `FALSE`.

- **greyName**
  Name of the grey module eigengene.

- **orderBy**
  Specifies the set by which the eigengenes are to be ordered (in all other sets as well). Defaults to the first set in `useSets` (or the first set, if `useSets` is not given).

- **order**
  Allows the user to specify a custom ordering.

- **useSets**
  Allows the user to specify for which sets the eigengene ordering is to be performed.

- **verbose**
  Controls verbostity of printed progress messages. 0 means silent, nonzero `verbose`.

- **indent**
  A single non-negative integer controlling indentation of printed messages. 0 means no indentation, each unit above zero adds two spaces.

**Details**

Ordering module eigengenes is useful for plotting purposes. For this function the order can be specified explicitly, or a set can be given in which the correlations of the eigengenes will determine the order. For the latter, a hierarchical dendrogram is calculated and the order given by the dendrogram is used for the eigengenes in all other sets.

**Value**

A vector of lists of the same type as `MEs` containing the re-ordered eigengenes.

**Author(s)**

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

**See Also**

`moduleEigengenes`, `multiSetMEs`, `consensusOrderMEs`
plotHclustColors  Plot color bars corresponding to modules

Description

Plot color bars corresponding to modules, usually beneath a dendrogram.

Usage

plotHclustColors(dendro, colors, rowLabels = NULL, cex.rowLabels = 0.9, ...)

Arguments

dendro A dendrogram such as returned by hclust.
colors Coloring of objects on the dendrogram. Either a vector (one color per object) or a matrix with each column giving one color per object. Each column will be plotted as a horizontal bar of colors under the dendrogram.
rowLabels Labels for the colorings given in color. If not given, sequential numbers starting from 1 will be used.
cex.rowLabels Font size scale factor for the row labels. See par.
... Other parameters to be passed on to the plotting method (such as main for the main title etc).

Details

It is often useful to plot module assignment (by color) that was obtained by cutting a hierarchical dendrogram, to visually check whether the obtained modules are meaningful, or which one of several possible module assignments looks best. One way to do it is to section the screen into two parts, plot the dendrogram (via plot(hclust)) in the upper section and use this function to plot colors corresponding to the dendrogram in the lower section.

Value

None.

Author(s)

Steve Horvath (SHorvath@mednet.ucla.edu) and Peter Langfelder (Peter.Langfelder@gmail.com)
standardColors

Colors this library uses for labeling modules.

Description

Returns the vector of color names in the order they are assigned by other functions in this library.

Usage

standardColors(n = NULL)

Arguments

n Number of colors requested. If NULL, all (approx. 450) colors will be returned. Any other invalid argument such as less than one or more than maximum (length(standardColors())) will trigger an error.

Value

A vector of character color names of the requested length.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

Examples

standardColors(10);
Index

*Topic cluster
  moduleNumber, 11

*Topic color
  labels2colors, 6
  standardColors, 16

*Topic hplot
  plotHclustColors, 15

*Topic misc
  checkSets, 1
  consensusMEDissimilarity, 3
  consensusOrderMEs, 4
  fixDataStructure, 5
  mergeCloseModules, 7
  moduleEigengenes, 9
  multiSetMEs, 12
  normalizeLabels, 13
  orderMEs, 14
  standardColors, 16

*Topic package
  moduleColor-package, 8

*Topic utilities
  collectGarbage, 2
  checkSets, 1, 3, 4, 6, 7, 12, 14
  collectGarbage, 2
  consensusMEDissimilarity, 3
  consensusOrderMEs, 4, 15
  cutree, 11
  fixDataStructure, 5
  hclust, 11, 15
  labels2colors, 6
  mergeCloseModules, 7
  moduleColor
    (moduleColor-package), 8
  moduleColor-package, 8
  moduleEigengenes, 5, 9, 13, 15
  moduleNumber, 11
  multiSetMEs, 5, 12, 15
  normalizeLabels, 11, 13
  orderMEs, 4, 5, 14
  par, 15
  plotHclustColors, 15
  standardColors, 16