

Package ‘coreheat’

May 8, 2026

Type Package

Title Correlation Heatmaps

Version 0.3.2

Date 2026-02-04

Author Vidal Fey [aut, cre],
Henri Sara [aut]

Maintainer Vidal Fey <vidal.fey@gmail.com>

Description Create correlation heatmaps from a numeric matrix. Ensembl Gene ID row names can be converted to Gene Symbols using, e.g., BioMart. Optionally, data can be clustered and filtered by correlation, tree cutting and/or number of missing values. Genes of interest can be highlighted in the plot and correlation significance be indicated by asterisks encoding corresponding P-Values. Plot dimensions and label measures are adjusted automatically by default. The plot features rely on the heatmap.n2() function in the 'heatmapFlex' package.

Depends Biobase

Imports WGCNA, heatmapFlex, convertid (>= 0.2.1), methods, graphics, grDevices, rappdirs

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.3

Suggests rmarkdown, knitr, BiocManager, org.Hs.eg.db, org.Mm.eg.db

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2026-02-09 13:10:16 UTC

Contents

clust_cormap	2
cormap2	3
cormap_filt	7
eset_cor	10

Index	12
--------------	-----------

clust_cormap	<i>Cluster a correlation matrix and return the sorted matrix for plotting.</i>
--------------	--

Description

Helper function to cluster the correlation matrix and return the sorted matrix for plotting.

Usage

```
clust_cormap(
  cormat,
  na.frac = 0.1,
  distfn = function(cm) (1 - cm),
  method = "complete",
  cor.cluster = 1,
  cor.window = NULL,
  cor.thr = 0.8,
  cor.mar = 0.05,
  cut.thr = 0.9,
  cut.size = 5,
  list.output = FALSE,
  verbose = FALSE
)
```

Arguments

cormat	(numeric or list). The correlation matrix or a list containing matrices of correlation values and P-Values as generated by <code>eset_cor</code> . If a list then the correlation matrix is expected to be in slot 'cor' and the matrix of P-Values in slot 'pvalues'.
na.frac	(numeric). Fraction of missing values allowed per row of the input matrix. Defaults to 0.1 which means LESS than 10per centof the values in one row are allowed to be NAs.
distfn	(function). Function to calculate the dissimilarity matrix for clustering. Defaults to <code>function(cm) (1-cm)</code> .
method	(character). The agglomeration method used for clustering. See help for hclust . Defaults to "complete".

<code>cor.cluster</code>	(numeric). The correlation cluster along the diagonal 'line' in the heatmap that should be zoomed into. A sliding window of size <code>cor.window</code> will be moved along the diagonal of the correlation matrix to find the cluster with the most correlation values meeting <code>core.thr</code> . Defaults to 1.
<code>cor.window</code>	(numeric). The size of the sliding window (see <code>cor.cluster</code>). Defaults to NULL. Note that this works only for positive correlations.
<code>cor.thr</code>	(numeric). Correlation threshold to filter the correlation matrix for plotting. Defaults to NULL meaning no filtering. See also ' <code>cor.mar</code> '. This value is sign-sensitive: a negative threshold will retain rows and columns of the correlation matrix with correlation values between -1 and the threshold, a positive value will retain rows and columns with values between the threshold and 1. Zero (0) is treated as positive.
<code>cor.mar</code>	(numeric). Margin of the values per row of the correlation matrix the ' <code>cor.thr</code> ' filter needs to meet. Defaults to 0.5 meaning at least 50 per cent of the values in a row need to meet the threshold in order to keep the row.
<code>cut.thr</code>	(numeric). Threshold at which dendrogram branches are to be cut. Passed on to argument <code>cutHeight</code> in <code>cutreeStatic</code> . Defaults to NULL meaning no cutting.
<code>cut.size</code>	(numeric). Minimum number of objects on a dendrogram branch considered a cluster. Passed on to argument <code>minSize</code> in <code>cutreeStatic</code> . Defaults to 5.
<code>list.output</code>	(logical). Should the output be a list of different object created in the call? Depends also on input type. If FALSE only the (filtered) correlation matrix is returned.
<code>verbose</code>	(logical). Should verbose output be written to the console? Defaults to FALSE.

Value

A correlation matrix or a list or the matrix and other values needed for plotting.

cormap2

Draw correlation maps from large datasets.

Description

`cormap2()` generates pair-wise correlations from an input `ExpressionSet` object, a `data.frame` or a numerical matrix. With the default options it also produces a heatmap.

Usage

```
cormap2(
  x,
  cormat = NULL,
  lab = NULL,
  convert = TRUE,
  biomart = FALSE,
  biom.data.set = "hsapiens_gene_ensembl",
```

```

biom.mart = "ensembl",
host = "https://www.ensembl.org",
biom.filter = "ensembl_gene_id",
biom.attributes = c("ensembl_gene_id", "hgnc_symbol"),
biom.cache = rappdirs::user_cache_dir("biomaRt"),
use.cache = TRUE,
cluster_correlations = TRUE,
main = "",
postfix = NULL,
cex = NULL,
na.frac = 0.1,
cor.cluster = 1,
cor.window = NULL,
cor.thr = NULL,
cor.mar = 0.5,
cut.thr = NULL,
cut.size = 5,
autoadj = TRUE,
labelheight = NULL,
labelwidth = NULL,
add.sig = FALSE,
genes2highl = NULL,
order.list = TRUE,
doPlot = TRUE,
updateProgress = NULL,
verbose = FALSE
)

```

Arguments

x	(ExpressionSet, data.frame or numeric). A numeric data frame, matrix or an ExpressionSet object.
cormat	(numeric). A correlation matrix. If this not NULL then 'x' is ignored. Defaults to NULL.
lab	(character). Optional row/column labels for the heatmap. Defaults to NULL meaning the row names of the input data are used. Note that the order of the labels must match the order of the row names of the input data!
convert	(logical). Should an attempt be made to convert IDs provided as row names of the input or in 'lab'? Defaults to TRUE. Conversion will be done using BioMart or an annotation package, depending on 'biomart'.
biomart	(logical). Should BioMart (or an annotation package) be used to convert IDs? If TRUE the todisp2 function in package convertid attempts to access the BioMart API to convert ENSG IDs to Gene Symbols Defaults to FALSE which will use the traditional AnnotationDbi Bimap interface.
biom.data.set	character of length one. Biomart data set to use. Defaults to 'hsapiens_gene_ensembl'
biom.mart	character vector. Biomart to use (uses the first element of the vector), defaults to "ensembl".

host	character of length one. Host URL.
biom.filter	character of length one. Name of biomart filter, i.e., type of query ids, defaults to "ensembl_gene_id".
biom.attributes	character vector. Biomart attributes, i.e., type of desired result(s); make sure query id type is included!
biom.cache	character. Path name giving the location of the cache getBM() uses if use.cache=TRUE. Defaults to the value in the <i>BIOMART_CACHE</i> environment variable.
use.cache	(logical). Should getBM() use the cache? Defaults to TRUE as in the getBM() function and is passed on to that.
cluster_correlations	(logical). Should the correlation matrix be clustered before plotting? Defaults to TRUE.
main	(character). The main title of the plot. Defaults to "".
postfix	(character of logical). A plot sub-title. Will be printed below the main title. Defaults to NULL.
cex	(numeric). Font size. Defaults to 0.5 if autoadj is FALSE. See 'Details'.
na.frac	(numeric). Fraction of missing values allowed per row of the input matrix. Defaults to 0.1 which means LESS than 10 per cent of the values in one row are allowed to be NAs.
cor.cluster	(numeric). The correlation cluster along the diagonal 'line' in the heatmap that should be zoomed into. A sliding window of size cor.window will be moved along the diagonal of the correlation matrix to find the cluster with the most correlation values meeting core.thr. Defaults to 1.
cor.window	(numeric). The size of the sliding window (see cor.cluster). Defaults to NULL. Note that this works only for positive correlations.
cor.thr	(numeric). Correlation threshold to filter the correlation matrix for plotting. Defaults to NULL meaning no filtering. Note that this value will be applied to margin 'cor.mar' of the values per row.
cor.mar	(numeric). Margin of the values per row of the correlation matrix the 'cor.thr' filter needs to meet. Defaults to 0.5 meaning at least 50 per cent of the values in a row need to meet the threshold in order to keep the row.
cut.thr	(numeric). Threshold at which dendrogram branches are to be cut. Passed on to argument cutHeight in <i>cutreeStatic</i> . Defaults to NULL meaning no cutting.
cut.size	(numeric). Minimum number of objects on a dendrogram branch considered a cluster. Passed on to argument minSize in <i>cutreeStatic</i> . Defaults to 5.
autoadj	(logical). Should plot measures be adjusted automatically? Defaults to TRUE.
labelheight	(numeric or lcm(numeric)). Relative or absolute height (using lcm, see layout) of the labels. Defaults to 0.2 if autoadj is FALSE. See 'Details'.
labelwidth	(numeric or lcm(numeric)). Relative or absolute width (using lcm, see layout) of the labels. Defaults to 0.2 if autoadj is FALSE. See 'Details'.
add.sig	(logical). Should significance asterisks be drawn? If TRUE P-Values for correlation significance are calculated and encoded as asterisks. See 'Details'.

genes2highl	(character). Vector of gene symbols (or whatever labels are used) to be highlighted. If not NULL will draw a semi-transparent rectangle around the labels and rows or columns in the heatmap labels.
order.list	(logical). Should the order of the correlation matrix, i.e. the 'list' of labels be reversed? Meaningful if the order of input variables should be preserved because image turns the input matrix. Defaults to TRUE.
doPlot	(logical). Draw the plot? Defaults to TRUE.
updateProgress	(function). Function for updating a progress bar in a Shiny web application. This was added here for the BioCPR application.
verbose	(logical). Should verbose output be written to the console? Defaults to FALSE.

Details

P-Values are calculated from the t-test value of the correlation coefficient: $t = rxsqrt(n-2)/sqrt(1-r^2)$, where r is the correlation coefficient, n is the number of samples with no missing values for each gene (row-wise `ncol(eset)` minus the number of columns that have an NA). P-Values are the calculated using [pt](#) and corrected account for the two-tailed nature of the test, i.e., the possibility of positive as well as negative correlation. The approach to calculate correlation significance was adopted from *Miles, J., & Banyard, P. (2007)* on "Calculating the exact significance of a Pearson correlation in MS Excel".

The asterisks encode significance as follows:

P < 0.05: *
 P < 0.01: **
 P < 0.001: ***

The label measures (`labelheight`, `labelwidth` and `cex`) are adjusted automatically by default with argument `autoadj=TRUE` and have default values which are hard coded into the helper function `heatmap.cor`. The values calculated by the helper function `plotAdjust` can be overridden by setting any of those arguments to a valid numeric or `lcm(numeric)` value.

Value

Invisibly returns the correlation matrix, though the function is mainly called for its side-effect of producing a heatmap (if `doPlot = TRUE` which is the default).

References

Miles, J., & Banyard, P. (2007). *Understanding and using statistics in psychology: A practical introduction*. Sage Publications Ltd. <https://psycnet.apa.org/record/2007-06525-000>.

See Also

[pt](#)
[tcrossprod](#)

Examples

```
# 1. Generate a random 10x10 matrix with two distinct sets and plot it with
# default settings without ID conversion since the IDs are made up:
set.seed(1234)
mat <- matrix(c(rnorm(100, mean = 1), rnorm(100, mean = -1)), nrow = 20)
rownames(mat) <- paste0("gene-", 1:20)
colnames(mat) <- paste0(c("A", "B"), rep(1:5, 2))
cormap2(mat, convert=FALSE, main="Random matrix")

# 2. Use a real-world dataset from TCGA (see README file in inst/extdata directory).
# Package 'convertid' is used to convert Ensembl Gene IDs to HGNC Symbols
## Read data and prepare input data frame
f1 <- system.file("extdata", "PrCaTCGASample.txt", package = "coreheat", mustWork = TRUE)
dat0 <- read.delim(f1, stringsAsFactors=FALSE)
dat1 <- data.frame(dat0[, grep("TCGA", names(dat0))], row.names=dat0$sensembl_gene_id)
cormap2(dat1, main="TCGA data frame + ID conversion")

# 3. Use separately supplied IDs with a matrix created from the data frame of the
# previous example and highlight genes of interest
dat2 <- as.matrix(dat0[, grep("TCGA", names(dat0))])
sym <- dat0$hgnc_symbol
cormap2(dat1, convert=FALSE, lab=sym, genes2highl=c("GNAS","NCOR1","AR", "ATM"),
main="TCGA matrix + custom labels")

# 4. Use an ExpressionSet object and add significance asterisks
## For simplicity reasons we create the ExpressionSet from a matrix created
## from the data frame in the second example
expr <- Biobase::ExpressionSet(as.matrix(dat1))
cormap2(expr, add.sig=TRUE, main="TCGA ExpressionSet object + ID conversion")

# More examples can be found in the vignette.
```

cormap_filt

Automatically split clusters based on noise level and hierarchy

Description

cormap_filt splits (cuts) the dendrogram at a given threshold dividing it into larger or smaller "sub-clusters". Correlation P-Values (see [eset_cor](#)) are converted to represent significance as a sub-cluster-wise signal metric used for filtering. Optionally, up to 3 plots are produced, the third one being a filtered heatmap based on significance and three height cutting.

Usage

```
cormap_filt(
  x,
  na.frac = 0.1,
  method = "ward.D",
  do.abs = TRUE,
```

```

main = "correlation map",
postfix = NULL,
p.thr = 0.01,
cex = 0.2,
cex.clust = cex,
cex.filt = cex,
cut.thr = NULL,
cor.thr = NULL,
cor.cluster = 1,
cor.window = NULL,
do.plots = c("dend", "full.heat", "filt.heat"),
genes2high1 = NULL,
order.list = TRUE,
convert = TRUE,
biomart = FALSE,
biom.data.set = "hsapiens_gene_ensembl",
biom.mart = "ensembl",
host = "https://www.ensembl.org",
biom.filter = "ensembl_gene_id",
biom.attributes = c("ensembl_gene_id", "hgnc_symbol"),
biom.cache = rappdirs::user_cache_dir("biomaRt"),
use.cache = TRUE,
add.sig = FALSE,
verbose = FALSE
)

```

Arguments

x	(ExpressionSet, data.frame or numeric). A numeric data frame, matrix or an ExpressionSet object.
na.frac	(numeric). Fraction of missing values allowed per row of the input matrix. Defaults to 0.1 which means LESS than 10 per cent of the values in one row are allowed to be NAs.
method	(character). The agglomeration method used for clustering. See help for hclust . Defaults to "ward.D".
do.abs	(logical). Should the distances for clustering be calculated based on the absolute correlation values? In other words, should the sign of the correlation be ignored in favor of its strength?
main	(character). The main title of the plot. Defaults to "".
postfix	(character or logical). A plot sub-title. Will be printed below the main title. Defaults to NULL.
p.thr	(numeric). P-Value threshold for filtering sub-clustered with significant correlations. Defaults to 0.01.
cex	(numeric). Font size for the heatmap of the unfiltered correlation matrix. Defaults to 0.2.
cex.clust	(numeric). Font size for the dendrogram plot of the unfiltered correlation matrix clusters. Defaults to cex.

cex.filt	(numeric). Font size for the heatmap of the filtered correlation matrix. Defaults to cex.
cut.thr	(numeric). Threshold at which dendrogram branches are to be cut. Passed on to argument h in <code>cut.dendrogram</code> . Defaults to NULL meaning no cutting.
cor.thr	(numeric). Correlation threshold to filter the correlation matrix for plotting. Defaults to NULL meaning no filtering. Note that this value will be applied to margin 'cor.mar' of the values per row.
cor.cluster	(numeric). The correlation cluster along the diagonal 'line' in the heatmap that should be zoomed into. A sliding window of size cor.window will be moved along the diagonal of the correlation matrix to find the cluster with the most correlation values meeting core.thr. Defaults to 1.
cor.window	(numeric). The size of the sliding window (see cor.cluster). Defaults to NULL. Note that this works only for positive correlations.
do.plots	(character). The plots to be produced. A character vector containing one or more of "dend" to produce the dendrogram plot, "full.heat" to produce the heatmap of the unfiltered correlation matrix, and "filt.heat" to produce the heatmap of the filtered correlation matrix. Defaults to all three plots.
genes2highl	(character). Vector of gene symbols (or whatever labels are used) to be highlighted. If not NULL will draw a semi-transparent rectangle around the labels and rows or columns in the heatmap labels.
order.list	(logical). Should the order of the correlation matrix, i.e. the 'list' of labels be reversed? Meaningful if the order of input variables should be preserved because <code>image</code> turns the input matrix. Defaults to TRUE.
convert	(logical). Should an attempt be made to convert IDs provided as row names of the input or in 'lab'? Defaults to TRUE. Conversion will be done using BioMart or an annotation package, depending on 'biomart'.
biomart	(logical). Should BioMart (or an annotation package) be used to convert IDs? If TRUE the <code>todisp2</code> function in package <code>convertid</code> attempts to access the BioMart API to convert ENSG IDs to Gene Symbols Defaults to FALSE which will use the traditional <code>AnnotationDbi</code> <code>Bimap</code> interface.
biom.data.set	character of length one. Biomart data set to use. Defaults to 'hsapiens_gene_ensembl'
biom.mart	character vector. Biomart to use (uses the first element of the vector), defaults to "ensembl".
host	character of length one. Host URL.
biom.filter	character of length one. Name of biomart filter, i.e., type of query ids, defaults to "ensembl_gene_id".
biom.attributes	character vector. Biomart attributes, i.e., type of desired result(s); make sure query id type is included!
biom.cache	character. Path name giving the location of the cache <code>getBM()</code> uses if <code>use.cache=TRUE</code> . Defaults to the value in the <code>BIOMART_CACHE</code> environment variable.
use.cache	(logical). Should <code>getBM()</code> use the cache? Defaults to TRUE as in the <code>getBM()</code> function and is passed on to that.

add.sig (logical). Should significance asterisks be drawn? If TRUE P-Values for correlation significance are calculated and encoded as asterisks. See 'Details'.

verbose (logical). Should verbose output be written to the console? Defaults to FALSE.

Details

P-Values are calculated from the t-test value of the correlation coefficient: $t = rxsqrt(n-2)/sqrt(1-r^2)$, where r is the correlation coefficient, n is the number of samples with no missing values for each gene (row-wise `ncol(eset)` minus the number of columns that have an NA). P-Values are then calculated using `pt` and corrected account for the two-tailed nature of the test, i.e., the possibility of positive as well as negative correlation. The approach to calculate correlation significance was adopted from *Miles, J., & Banyard, P. (2007)* on "Calculating the exact significance of a Pearson correlation in MS Excel".

To obtain a suitable metric for isolating significant sub-clusters, P-Values are represented as $-\log_{10}(\text{median}(pval))$ where `pval` is the *median of the parallel maximum of all P-Values belonging to the sub-cluster and* $1e-38$ to avoid values of zero (0).

Value

A list. If the dendrogram is being cut, i.e., `cut.thr` is not NULL, a list of

`clusters`: the list of cluster labels from lower component of the `cut.dendrogram` output which is list with the branches obtained
`filt`: the index of the cluster labels passing the signal metrics threshold
`filt_cluster`: the list of the filtered cluster labels
`h`: the cut threshold
`p.thr`: the P-Value threshold for filtering sub-clusters
`metric`: the signal metrics for all sub-clusters
`cormat`: the clustered (ordered) correlation matrix
`hclust`: a list of hierarchical clustering metrics (output of `hclust`)
`pvalues`: the correlation P-Value matrix

If no tree cutting is applied, a list of

`cormat`: the clustered (ordered) correlation matrix
`hclust`: a list of hierarchical clustering metrics (output of `hclust`)
`pvalues`: the correlation P-Value matrix

eset_cor

Helper function to calculate the correlation matrix.

Description

Helper function to calculate the correlation matrix.

Usage

```
eset_cor(x, with.pvalues = TRUE, order.list = TRUE, verbose = FALSE)
```

Arguments

x (ExpressionSet, data.frame or numeric). A numeric data frame, matrix or an ExpressionSet object.

with.pvalues (logical). Should P-Values be calculated for the correlations? If TRUE P-Values will be depicted in the heatmap by significance asterisks. See 'Details'. Defaults to FALSE.

order.list (logical). Is the input matrix column order reversed? Only applicable if input is correlation matrix. Defaults to TRUE.

verbose (logical). Should verbose output be written to the console? Defaults to FALSE.

Details

P-Values are calculated from the t-test value of the correlation coefficient: $t = r\sqrt{(n-2)/\sqrt{1-r^2}}$, where r is the correlation coefficient, n is the number of samples with no missing values for each gene (row-wise `ncol(eset)` minus the number of columns that have an NA). P-Values are then calculated using `pt` and corrected account for the two-tailed nature of the test, i.e., the possibility of positive as well as negative correlation. The approach to calculate correlation significance was adopted from *Miles, J., & Banyard, P. (2007)* on "Calculating the exact significance of a Pearson correlation in MS Excel".

Value

A correlation matrix or a list with three slots: the correlation matrix, the number of samples with no missing value for each gene and the P-Values matrix.

References

Miles, J., & Banyard, P. (2007). *Understanding and using statistics in psychology: A practical introduction*. Sage Publications Ltd. <https://psycnet.apa.org/record/2007-06525-000>.

Index

clust_cormap, [2](#)
cormap2, [3](#)
cormap_filt, [7](#)
cut.dendrogram, [9](#)
cutreeStatic, [3](#), [5](#)

eset_cor, [7](#), [10](#)

hclust, [2](#), [8](#), [10](#)

image, [6](#), [9](#)

layout, [5](#)
lcm, [5](#)

pt, [6](#), [10](#), [11](#)

tcrossprod, [6](#)