Package 'SPONGE'

April 16, 2019

Type Package

Title Sparse Partial Correlations On Gene Expression

Version 1.4.0

Author Markus List, Azim Dehghani Amirabad, Dennis Kostka, Marcel H. Schulz

Maintainer Markus List <markus.list@wzw.tum.de>

Description This package provides methods to efficiently detect competitive endogeneous RNA interactions between two genes. Such interactions are mediated by one or several miRNAs such that both gene and miRNA expression data for a larger number of samples is needed as input.

License GPL (>=3)

LazyData TRUE

RoxygenNote 6.0.1

Depends R (>= 3.4)

Suggests testthat, knitr, rmarkdown, visNetwork, ggplot2, ggrepel, gridExtra, digest, doParallel, bigmemory

Imports Biobase, stats, ppcor, logging, foreach, doRNG, data.table, MASS, expm, gRbase, glmnet, igraph, iterators,

VignetteBuilder knitr

biocViews GeneExpression, Transcription, GeneRegulation, NetworkInference, Transcriptomics, SystemsBiology, Regression

git_url https://git.bioconductor.org/packages/SPONGE

git_branch RELEASE_3_8

git_last_commit 4123d4f

git_last_commit_date 2018-10-30

Date/Publication 2019-04-15

R topics documented:

`ceRNA_interactions` *ceRNA interactions*

Description

ceRNA interactions

Usage

```
ceRNA_interactions
```

Format

A data table of ceRNA interactions typically provided by sponge

check_and_convert_expression_data

Checks if expression data is in matrix or ExpressionSet format and converts the latter to a standard matrix. Alternatively, a big.matrix descriptor object can be supplied to make use of shared memory between parallelized workers through the bigmemory package.

Description

Checks if expression data is in matrix or ExpressionSet format and converts the latter to a standard matrix. Alternatively, a big.matrix descriptor object can be supplied to make use of shared memory between parallelized workers through the bigmemory package.

Usage

```
check_and_convert_expression_data(expr_data)
```

Arguments

expr_data	expr_data as matrix or ExpressionSet

Value

expr_data as matrix

Examples

```
## Not run: check_and_convert_expression_data(gene_expr)
```

fn_elasticnet

Computes an elastic net model

Description

Computes an elastic net model

Usage

```
fn_elasticnet(x, y, alpha.step = 0.1)
```

Arguments

x	miRNA expression matrix
y	gene expression vector
alpha.step	Step size for alpha, the tuning parameter for elastic net.

Value

The best model, i.e. the one for which the selected alpha yielded the smallest residual sum of squares error

fn_gene_miRNA_F_test *Perform F test for gene-miRNA elastic net model*

Description

Perform F test for gene-miRNA elastic net model

Usage

```
fn_gene_miRNA_F_test(g_expr, m_expr, model, p.adj.threshold = NULL)
```

Arguments

g_expr	A gene expression matrix with samples in rows and genes in columns
m_expr	A miRNA expression matrix with samples in rows and genes in columns. Sample number and order has to agree with above gene expression matrix
model	A nested elastic net model to be tested
p.adj.threshold	Threshold for FDR corrected p-value

Value

return data frame with miRNA, fstat and adjusted p.value (BH).

fn_get_model_coef *Extract the model coefficients from an elastic net model*

Description

Extract the model coefficients from an elastic net model

Usage

```
fn_get_model_coef(model)
```

Arguments

model	An elastic net model

Value

A data frame with miRNAs and coefficients

fn_get_rss *Compute the residual sum of squares error for an elastic net model*

Description

Compute the residual sum of squares error for an elastic net model

Usage

```
fn_get_rss(model, x, y)
```

Arguments

model	The elastic net model
x	The miRNA expression
y	The gene expression

Value

the RSS

fn_get_shared_miRNAs *Identify miRNAs for which both genes have miRNA binding sites aka miRNA response elements in the competing endogeneous RNA hypothesis*

Description

Identify miRNAs for which both genes have miRNA binding sites aka miRNA response elements in the competing endogeneous RNA hypothesis

Usage

```
fn_get_shared_miRNAs(geneA, geneB, mir_interactions)
```

Arguments

geneA	The first gene
geneB	The second gene
mir_interactions	A named list of genes, where for each gene all miRNA interacting partners are listed

Value

A vector with shared RNAs of the two genes.

genes_pairwise_combinations *Compute all pairwise interactions for a number of genes as indices*

Description

Compute all pairwise interactions for a number of genes as indices

Usage

```
genes_pairwise_combinations(number.of.genes)
```

Arguments

number.of.genes
: Number of genes for which all pairwise interactions are needed

Value

data frame with one row per unique pairwise combination. To be used as input for the sponge method.

gene_expr *Gene expression test data set*

Description

Gene expression test data set

Usage

```
gene_expr
```

Format

A data frame of expression values with samples in columns and genes in rows

mircode_ensg *mircode predicted miRNA gene interactions*

Description

mircode predicted miRNA gene interactions

Usage

```
mircode_ensg
```

Format

A matrix gene ensembl ids vs miRNA family names. >=1 if interaction is predicted, 0 otherwise

Source

http://www.mircode.org/download.php

mircode_symbol *mircode predicted miRNA gene interactions*

Description

mircode predicted miRNA gene interactions

Usage

```
mircode_symbol
```

Format

A matrix gene symbols vs miRNA family names. >=1 if interaction is predicted, 0 otherwise

Source

http://www.mircode.org/download.php

mir_expr *miRNA expression test data set*

Description

miRNA expression test data set

Usage

```
mir_expr
```

Format

A data frame of expression values with samples in columns and miRNA in rows

mir_interactions *miRNA / gene interactions*

Description

miRNA / gene interactions

Usage

```
mir_interactions
```

Format

A data frame of regression coefficients typically provided by sponge_gene_miRNA_interaction_filter

precomputed_cov_matrices *covariance matrices under the null hypothesis that sensitivity correlation is zero*

Description

covariance matrices under the null hypothesis that sensitivity correlation is zero

Usage

```
precomputed_cov_matrices
```

Format

A list (different gene-gene correlations k) of lists (different number of miRNAs m) of covariance matrices

precomputed_null_model *A null model for testing purposes*

Description

A null model for testing purposes

Usage

```
precomputed_null_model
```

Format

A list (different gene-gene correlations k) of lists (different number of miRNAs m) of sampled mscor values (100 each, computed from 100 samples)

sample_zero_mscor_cov *Sampling zero multiple miRNA sensitivity covariance matrices*

Description

Sampling zero multiple miRNA sensitivity covariance matrices

Usage

```
sample_zero_mscor_cov(m, number_of_solutions, number_of_attempts = 1000,
  gene_gene_correlation = NULL, random_seed = NULL, log.level = "ERROR")
```

Arguments

m	number of miRNAs, i.e. number of columns of the matrix
number_of_solutions	stop after this many instances have been samples
number_of_attempts	give up after that many attempts
gene_gene_correlation	optional, define the correlation of the first two elements, i.e. the genes.
random_seed	A random seed to be used for reproducible results
log.level	the log level, typically set to INFO, set to DEBUG for verbose logging

Value

a list of covariance matrices with zero sensitivity correlation

Examples

```
sample_zero_mscor_cov(m = 1,
number_of_solutions = 1,
gene_gene_correlation = 0.5)
```

sample_zero_mscor_data *Sample mscor coefficients from pre-computed covariance matrices*

Description

Sample mscor coefficients from pre-computed covariance matrices

Usage

```
sample_zero_mscor_data(cov_matrices, number_of_samples = 100,
  number_of_datasets = 100)
```

Arguments

cov_matrices	a list of pre-computed covariance matrices
number_of_samples	the number of samples available in the expression data
number_of_datasets	the number of mscor coefficients to be sampled from each covariance matrix

Value

a vector of mscor coefficients

See Also

sample_zero_mscor_cov

Examples

```
#we select from the pre-computed covariance matrices in SPONGE
#100 for m = 5 miRNAs and gene-gene correlation 0.6
cov_matrices_selected <- precomputed_cov_matrices[["5"]][["0.6"]]
sample_zero_mscor_data(cov_matrices = cov_matrices_selected,
number_of_samples = 200, number_of_datasets = 10)
```

---

sponge *Compute competing endogeneous RNA interactions using Sparse Partial correlations ON Gene Expression (SPONGE)*

Description

Compute competing endogeneous RNA interactions using Sparse Partial correlations ON Gene Expression (SPONGE)

Usage

```
sponge(gene_expr, mir_expr, mir_interactions = NULL, log.level = "ERROR",
  log.every.n = 1e+05, log.file = NULL, selected.genes = NULL,
  gene.combinations = NULL, each.miRNA = FALSE, min.cor = 0.1,
  parallel.chunks = 1000, random_seed = NULL, result_as_dt = FALSE)
```

Arguments

gene_expr	A gene expression matrix with samples in rows and featurs in columns. Alternatively an object of class ExpressionSet.
mir_expr	A miRNA expression matrix with samples in rows and features in columns. Alternatively an object of class ExpressionSet.
mir_interactions	A named list of genes, where for each gene we list all miRNA interaction partners that should be considered.
log.level	The log level, can be one of "info", "debug", "error"
log.every.n	write to the log after every n steps

log.file	write log to a file, particularly useful for paralleliyzation
selected.genes	Operate only on a subset of genes, particularly useful for bootstrapping
gene.combinations	A data frame of combinations of genes to be tested. Gene names are taken from the first two columns and have to match the names used for gene_expr
each.miRNA	Whether to consider individual miRNAs or pooling them.
min.cor	Consider only gene pairs with a minimum correlation specified here.
parallel.chunks	Split into this number of tasks if parallel processing is set up. The number should be high enough to guarantee equal distribution of the work load in parallel execution. However, if the number is too large, e.g. in the worst case one chunk per computation, the overhead causes more computing time than can be saved by parallel execution. Register a parallel backend that is compatible with foreach to use this feature. More information can be found in the documentation of the foreach / doParallel packages.
random_seed	A random seed to be used for reproducible results
result_as_dt	whether to return results as data table or data frame

Value

A data frame with significant gene-gene competetive endogenous RNA or 'sponge' interactions

Examples

```
#First, extract miRNA candidates for each of the genes
#using sponge_gene_miRNA_interaction_filter. Here we use a prepared
#dataset mir_interactions.

#Second we compute ceRNA interactions for all pairwise combinations of genes
#using all miRNAs remaining after filtering through elasticnet.
ceRNA_interactions <- sponge(
gene_expr = gene_expr,
mir_expr = mir_expr,
mir_interactions = mir_interactions)
```

---

sponge_build_null_model

Build null model for p-value computation

Description

Build null model for p-value computation

Usage

```
sponge_build_null_model(number_of_datasets = 1e+05, number_of_samples,
  cov_matrices = precomputed_cov_matrices, ks = seq(0.2, 0.9, 0.1),
  m_max = 8, log.level = "ERROR")
```

Arguments

number_of_datasets	the number of datesets defining the precision of the p-value
number_of_samples	the number of samples in the expression data
cov_matrices	pre-computed covariance matrices
ks	a sequence of gene-gene correlation values for which null models are computed
m_max	null models are build for each elt in ks for 1 to m_max miRNAs
log.level	The log level of the logging package

Value

a list (for various values of m) of lists (for various values of k) of lists of simulated data sets, drawn from a set of precomputed covariance matrices

Examples

```
sponge_build_null_model(100, 100,
cov_matrices = precomputed_cov_matrices[1:3], m_max = 3)
```

---

sponge_compute_p_values

Compute p-values for SPONGE interactions

Description

This method uses pre-computed covariance matrices that were created for various gene-gene correlations (0.2 to 0.9 in steps of 0.1) and number of miRNAs (between 1 and 8) under the null hypothesis that the sensitivity correlation is zero. Datasets are sampled from this null model and allow for an empirical p-value to be computed that is only significant if the sensitivity correlation is higher than can be expected by chance given the number of samples, correlation and number of miRNAs. p-values are adjusted indepdenently for each parameter combination using Benjamini-Hochberg FDR correction.

Usage

```
sponge_compute_p_values(sponge_result, null_model, log.level = "ERROR")
```

Arguments

sponge_result	A data frame from a sponge call
null_model	optional, pre-computed simulated data
log.level	The log level of the logging package

Value

A data frame with sponge results, now including p-values and adjusted p-value

See Also

sponge_build_null_model

Examples

```
sponge_compute_p_values(ceRNA_interactions,
null_model = precomputed_null_model)
```

`sponge_edge_centralities` *Computes edge centralities*

Description

Computes edge betweenness centrality for the ceRNA interaction network induced by the results of the SPONGE method.

Usage

```
sponge_edge_centralities(sponge_result)
```

Arguments

`sponge_result`	The output generated by the sponge method.

Value

data table or data frame with gene, degree, eigenvector and betweenness

See Also

sponge

Examples

```
sponge_edge_centralities(ceRNA_interactions)
```

`sponge_gene_miRNA_interaction_filter` *Determine miRNA-gene interactions to be considered in SPONGE*

Description

The purpose of this method is to limit the number of miRNA-gene interactions we need to consider in SPONGE. There are 3 filtering steps: 1. variance filter (optional). Only considre genes and miRNAs with variance > var.threshold. 2. miRNA target database filter (optional). Use a miRNA target database provided by the user to filter for those miRNA gene interactions for which evidence exists. This can either be predicted target interactions or experimentally validated ones. 3. For each remaining interaction of a gene and its regulating miRNAs use elastic net regression to achieve a) Feature selection: We only retain miRNAs that influence gene expression b) Effect strength: The sign of the coefficients allows us to filter for miRNAs that down-regulate gene expression. Moreover, we can use the coefficients to rank the miRNAs by their relative effect strength. We strongly recommend setting up a parallel backend compatible with the foreach package. See example and the documentation of the foreach and doParallel packages.

Usage

```
sponge_gene_miRNA_interaction_filter(gene_expr, mir_expr, mir_predicted_targets,
  elastic.net = TRUE, log.level = "ERROR", log.file = NULL,
  var.threshold = NULL, F.test = FALSE, F.test.p.adj.threshold = 0.05,
  coefficient.threshold = -0.05, coefficient.direction = "<",
  select.non.targets = FALSE, random_seed = NULL, parallel.chunks = 100)
```

Arguments

`gene_expr`	A gene expression matrix with samples in rows and featurs in columns. Alternatively an object of class ExpressionSet.
`mir_expr`	A miRNA expression matrix with samples in rows and features in columns. Alternatively an object of class ExpressionSet.
`mir_predicted_targets`	A data frame with miRNA in cols and genes in rows. A 0 indicates the miRNA is not predicted to target the gene, >0 otherwise. If this parameter is NULL all miRNA-gene interactions are tested
`elastic.net`	Whether to apply elastic net regression filtering or not.
`log.level`	One of 'warn', 'error', 'info'
`log.file`	Log file to write to
`var.threshold`	Only consider genes and miRNA with variance > var.threshold. If this parameter is NULL no variance filtering is performed.
`F.test`	If true, an F-test is performed on each model parameter to assess its importance for the model based on the RSS of the full model vs the RSS of the nested model without the miRNA in question. This is time consuming and has the potential disadvantage that correlated miRNAs are removed even though they might play a role in ceRNA interactions. Use at your own risk.
`F.test.p.adj.threshold`	If F.test is TRUE, threshold to use for miRNAs to be included.

`coefficient.threshold`
: threshold to cross for a regression coefficient to be called significant. depends on the parameter coefficient.direction.

`coefficient.direction`
: If "<", coefficient has to be lower than coefficient.threshold, if ">", coefficient has to be larger than threshold. If NULL, the absolute value of the coefficient has to be larger than the threshold.

`select.non.targets`
: For testing effect of miRNA target information. If TRUE, the method determines as usual which miRNAs are potentially targeting a gene. However, these are then replaced by a random sample of non-targeting miRNAs (without seeds) of the same size. Useful for testing if observed effects are caused by miRNA regulation.

`random_seed`
: A random seed to be used for reproducible results

`parallel.chunks`
: Split into this number of tasks if parallel processing is set up. The number should be high enough to guarantee equal distribution of the work load in parallel execution. However, if the number is too large, e.g. in the worst case one chunk per computation, the overhead causes more computing time than can be saved by parallel execution. Register a parallel backend that is compatible with foreach to use this feature. More information can be found in the documentation of the foreach / doParallel packages.

Value

A list of genes, where for each gene, the regulating miRNA are included as a data frame. For F.test = TRUE this is a data frame with fstat and p-value for each miRNA. Else it is a data frame with the model coefficients.

See Also

sponge

Examples

```
#library(doParallel)
#cl <- makePSOCKcluster(2)
#registerDoParallel(cl)
genes_miRNA_candidates <- sponge_gene_miRNA_interaction_filter(
gene_expr = gene_expr,
mir_expr = mir_expr,
mir_predicted_targets = targetscan_symbol)
#stopCluster(cl)

#If we also perform an F-test, only few of the above miRNAs remain
genes_miRNA_candidates <- sponge_gene_miRNA_interaction_filter(
gene_expr = gene_expr,
mir_expr = mir_expr,
mir_predicted_targets = targetscan_symbol,
F.test = TRUE,
F.test.p.adj.threshold = 0.05)
```

sponge_network *Prepare a sponge network for plotting*

Description

Prepare a sponge network for plotting

Usage

```
sponge_network(sponge_result, mir_data, target.genes = NULL,
  show.sponge.interaction = TRUE, show.mirnas = "none",
  min.interactions = 3)
```

Arguments

sponge_result	ceRNA interactions as produced by the sponge method.
mir_data	miRNA interactions as produced by sponge_gene_miRNA_interaction_filter
target.genes	a character vector to select a subset of genes
show.sponge.interaction	whether to connect ceRNAs
show.mirnas	one of none, shared, all
min.interactions	minimum degree of a gene to be shown

Value

a list of nodes and edges

Examples

```
sponge_network(ceRNA_interactions, mir_interactions)
```

sponge_node_centralities *Computes various node centralities*

Description

Computes degree, eigenvector centrality and betweenness centrality for the ceRNA interaction network induced by the results of the SPONGE method

Usage

```
sponge_node_centralities(sponge_result, directed = FALSE)
```

Arguments

sponge_result	output of the sponge method
directed	Whether to consider the input network as directed or not.

Value

data table or data frame with gene, degree, eigenvector and betweenness

See Also

sponge

Examples

```
sponge_node_centralities(ceRNA_interactions)
```

`sponge_plot_network` *Plot a sponge network*

Description

Plot a sponge network

Usage

```
sponge_plot_network(sponge_result, mir_data,
  layout = "layout.fruchterman.reingold", force.directed = FALSE, ...)
```

Arguments

`sponge_result`	ceRNA interactions as produced by the sponge method.
`mir_data`	miRNA interactions as produced by sponge_gene_miRNA_interaction_filter
`layout`	one of the layout methods supported in the visNetwork package
`force.directed`	whether to produce a force directed network, gets slow for large networks
`...`	further params for sponge_network

Value

shows a plot

Examples

```
sponge_plot_network(ceRNA_interactions, mir_interactions)
```

sponge_plot_network_centralities *plot node network centralities*

Description

plot node network centralities

Usage

```
sponge_plot_network_centralities(network_centralities, measure = "all",
  x = "degree", top = 5, base_size = 18)
```

Arguments

network_centralities	a result from sponge_node_centralities()
measure	one of 'all', 'degree', 'ev' or 'btw'
x	plot against another column in the data table, defaults to degree
top	label the top x samples in the plot
base_size	size of the text in the plot

Value

a plot

Examples

```
## Not run:
network_centralities <- sponge_node_centralities(ceRNA_interactions)
sponge_plot_network_centralities(network_centralities)
## End(Not run)
```

sponge_plot_simulation_results *Plot simulation results for different null models*

Description

Plot simulation results for different null models

Usage

```
sponge_plot_simulation_results(null_model_data)
```

Arguments

null_model_data	the output of sponge_build_null_model

Value

a ggplot2 object

Examples

```
sponge_plot_simulation_results(precomputed_null_model)
```

---

sponge_run_benchmark *run sponge benchmark where various settings, i.e. with or without regression, single or pooled miRNAs, are compared.*

Description

run sponge benchmark where various settings, i.e. with or without regression, single or pooled miRNAs, are compared.

Usage

```
sponge_run_benchmark(gene_expr, mir_expr, mir_predicted_targets,
  number_of_samples = 100, number_of_datasets = 100,
  number_of_genes_to_test = c(25), compute_significance = FALSE,
  folder = NULL)
```

Arguments

gene_expr	A gene expression matrix with samples in rows and featurs in columns. Alternatively an object of class ExpressionSet.
mir_expr	A miRNA expression matrix with samples in rows and features in columns. Alternatively an object of class ExpressionSet.
mir_predicted_targets	(a list of) mir interaction sources such as targetscan, etc.
number_of_samples	number of samples in the null model
number_of_datasets	number of datasets to sample from the null model
number_of_genes_to_test	a vector of numbers of genes to be tested, e.g. c(250,500)
compute_significance	whether to compute p-values
folder	where the results should be saved, if NULL no output to disk

Value

a list (regression, no regression) of lists (single miRNA, pooled miRNAs) of benchmark results

Examples

```
sponge_run_benchmark(gene_expr = gene_expr, mir_expr = mir_expr,
mir_predicted_targets = targetscan_symbol,
number_of_genes_to_test = c(10), folder = NULL)
```

sponge_subsampling *Sponge subsampling*

Description

Sponge subsampling

Usage

```
sponge_subsampling(subsample.n = 100, subsample.repeats = 10,
  subsample.with.replacement = FALSE, subsample.plot = FALSE, gene_expr,
  mir_expr, ...)
```

Arguments

subsample.n	the number of samples to be drawn in each round
subsample.repeats	how often should the subsampling be done?
subsample.with.replacement	logical, should we allow samples to be used repeatedly
subsample.plot	logical, should the results be plotted as box plots
gene_expr	A gene expression matrix with samples in rows and featurs in columns. Alternatively an object of class ExpressionSet.
mir_expr	A miRNA expression matrix with samples in rows and features in columns. Alternatively an object of class ExpressionSet.
...	parameters passed on to the sponge function

Value

a summary of the results with mean and standard deviations of the correlation and sensitive correlation.

References

sponge

Examples

```
sponge_subsampling(gene_expr = gene_expr,
mir_expr = mir_expr, mir_interactions = mir_interactions,
subsample.n = 10, subsample.repeats = 1)
```

targetscan_ensg *targetscan predicted miRNA gene interactions*

Description

targetscan predicted miRNA gene interactions

Usage

```
targetscan_ensg
```

Format

A matrix gene ensembl ids vs miRNA family names. >=1 if interaction is predicted, 0 otherwise

Source

http://www.targetscan.org/vert_71/

targetscan_symbol *targetscan predicted miRNA gene interactions*

Description

targetscan predicted miRNA gene interactions

Usage

```
targetscan_symbol
```

Format

A matrix gene symbols vs miRNA family names. >=1 if interaction is predicted, 0 otherwise

Source

http://www.targetscan.org/vert_71/

Index