

Package ‘meaca’

January 3, 2017

Title What the Package Does (one line, title case)

Version 0.0.1.0

Description What the package does (one paragraph).

Depends R (>= 3.2.1)

License What license is it under?

LazyData true

RoxygenNote 5.0.1

R topics documented:

btw_gene_corr	2
Camera_multiple	2
Camera_multiple1	3
compare_test	3
data_simu	4
data_simu_meaca	4
estimate_sigma	5
GSEA	6
GSEA.Gct2Frame2	6
GSEA.ReadClsFile	7
GSEA1	7
meaca_multiple	8
meaca_multiple1	8
meaca_single	9
MRGSE_multiple	9
MRGSE_multiple1	10
prepare_go_term	10
prepare_simulation	11
read_gene_set	11
sig_path	12
simulate_expression_data	12
standardize_expression_data	13
Index	14

btw_gene_corr	<i>Estimate sample correlation.</i>
---------------	-------------------------------------

Description

Average correlations for genes

Usage

```
btw_gene_corr(expression_data, trt, geneset, standardize = T,
  minSetSize = 5)
```

Arguments

expression_data	the expressoin matrix.
trt	treatment labels
geneset	an object from read_gene_set()
standardize	whether the data should be standaridzed
minSetSize	the minimum number of genes contained for a gene set to be considered.

Value

a list	
set_name	The name of the gene set
testSetCor	Average correlation for genes in the test set
interCor	Average correlation between genes in the test set and those not in the test set
backSetCor	Average correlations for genes not in the test set.

Camera_multiple	<i>Camera multiple</i>
-----------------	------------------------

Description

multiple gene set enrichemnt analysis

Usage

```
Camera_multiple(expression_data, trt, geneset, use.rank = F)
```

Arguments

expression_data	the expression data matrix
trt	the treatment lables for each column of expression_data
geneset	a list containing all the gene sets to be tested for enrichment status
use.rank	If TRUE then it corresponds to "Camara-Rank", otherwise "Camara"

Value

a list that has exactly the same elements as that in camera of "limma" package (see user manual for more details)

Camera_multiple1	<i>CAMERA multiple</i>
------------------	------------------------

Description

multiple gene set enrichment analysis (specific for testing GSE64810)

Usage

```
Camera_multiple1(expression, trt, geneset, use.rank = F)
```

compare_test	<i>Compare meaca to existing methods</i>
--------------	--

Description

Compare meaca to other existing methods.

Usage

```
compare_test(dat)
```

Arguments

dat	the simulated data consisting of three lists: \itemdatthe expression matrix \itemtrtwhat treatment each column of data belongs to \itemgo_termwhether each row of data belongs to this go term
-----	--

Value

a matrix of p values

data_simu

Compare meaca to existing methods

Description

Produce p value matrix for simulation discussed in the paper.

Usage

```
data_simu(case, prop = c(0.1, 0.1), de = c(2, 1), nsim = 1000,
  size = 50, rho = c(0.1, 0.05, -0.05),
  dest = "/home/stats/zhuob/data/computing/", num_gene = c(500, 100),
  post_txt = "PCT.txt")
```

Arguments

prop	A vector of two, specifying the proportion of DE genes for the test set and the background set
de	A vector of two, specifying the (normally distributed) DE effect size and its std
nsim	number of simulations to run
size	number of samples to be simulated
rho	A vector of three, for correlation coefficients corresponding to rho1, rho2, rho3 in the paper
dest	where to store the results
num_gene	A vector of two, specifying the number of genes to be simulated in total and in the test set
post_txt	formatted name of output files

Value

a text file containing the p value matrix

data_simu_meaca

Power for meaca

Description

Produce p value matrix for simulation discussed in the paper (Figure 2).

Usage

```
data_simu_meaca(case, prop = c(0.1, 0.1), de = c(2, 1), nsim = 1000,
  size = 50, rho = c(0.1, 0.05, -0.05),
  dest = "/home/stats/zhuob/data/computing/", num_gene = c(500, 100))
```

Arguments

prop	A vector of two, specifying the proportion of DE genes for the test set and the background set
de	A vector of two, specifying the (normally distributed) DE effect size and its std
nsim	number of simulations to run
size	number of samples to be simulated
rho	A vector of three, for correlation coefficients corresponding to rho1, rho2, rho3 in the paper
dest	where to store the results
num_gene	A vector of two, specifying the number of genes to be simulated in total and in the test set
post_txt	formatted name of output files

Value

a text file containing the p value matrix

estimate_sigma	<i>Estimate sample covariance.</i>
----------------	------------------------------------

Description

Estimate sample covariance and calculate the gene-level statistics

Usage

```
estimate_sigma(expression_data, trt)
```

Arguments

expression_data	the expression matrix.
trt	sample labels. 0 for control and 1 for treatment

Value

a list	
sigma	a covariance matrix
t_val	a vector of gene level test statistics

GSEA	<i>GSEA.</i>
------	--------------

Description

GSEA for testing multiple gene sets.

Usage

```
GSEA(input.ds, input.cls, gene.ann = "", gs.db, gs.ann = "",
      output.directory = "", doc.string = "GSEA.analysis",
      non.interactive.run = F, reshuffling.type = "sample.labels",
      nperm = 1000, weighted.score.type = 1, nom.p.val.threshold = -1,
      fwer.p.val.threshold = -1, fdr.q.val.threshold = 0.25, topgs = 10,
      adjust.FDR.q.val = F, gs.size.threshold.min = 25,
      gs.size.threshold.max = 500, reverse.sign = F, preproc.type = 0,
      random.seed = 123456, perm.type = 0, fraction = 1, replace = F,
      save.intermediate.results = F, OLD.GSEA = F,
      use.fast.enrichment.routine = T)
```

GSEA.Gct2Frame2	<i>readling sample lables.</i>
-----------------	--------------------------------

Description

readling sample labels (from GSEA.1.0.R)
read the expression data of Gct format.

Usage

```
GSEA.Gct2Frame2(filename = "NULL")  
  
GSEA.Gct2Frame2(filename = "NULL")
```

Arguments

file the expression data files.

GSEA.ReadClsFile	<i>readling Cls files.</i>
------------------	----------------------------

Description

readling Cls file (from GSEA.1.0.R)
read the gene sets of the MsigDB format.

Usage

```
GSEA.ReadClsFile(file = "NULL")  
  
GSEA.ReadClsFile(file = "NULL")
```

Arguments

msigdb gene set ensemble downloaded from broad institute.

Value

a list
total
number of gene sets contained.
size
a numerical vector containing the size of each gene set.
gene_set
a list. The first element is the set name. From the third element each containing members of the gene set.

GSEA1	<i>GSEA multiple</i>
-------	----------------------

Description

multiple gene set enrichment analysis (specific for testing GSE64810)

Usage

```
GSEA1(input.ds, input.cls, gene.ann = "", gs.db, gs.ann = "",  
      output.directory = "", doc.string = "GSEA.analysis",  
      non.interactive.run = F, reshuffling.type = "sample.labels",  
      nperm = 1000, weighted.score.type = 1, nom.p.val.threshold = -1,  
      fwer.p.val.threshold = -1, fdr.q.val.threshold = 0.25, topgs = 10,  
      adjust.FDR.q.val = F, gs.size.threshold.min = 25,  
      gs.size.threshold.max = 500, reverse.sign = F, preproc.type = 0,  
      random.seed = 123456, perm.type = 0, fraction = 1, replace = F,  
      save.intermediate.results = F, OLD.GSEA = F,  
      use.fast.enrichment.routine = T)
```

meaca_multiple	<i>meaca-multiple.</i>
----------------	------------------------

Description

meaca for testing multiple gene sets.

Usage

```
meaca_multiple(expression_data, trt, geneset, standardize = T,
  minSetSize = 5, fdr_method = "BH")
```

Arguments

expression_data	the expressoin matrix.
trt	treatment labels.
geneset	gene sets to be tested, an object from read_gene_set.
standardize	whether the data should be standaridzed.
minSetSize	the minimum number of genes contained for a gene set to be considered.
fdr_method	which method is ued to adjust the p values. see arguments in function p.adjust.

Value

a data frame

meaca_multiple1	<i>meaca multiple</i>
-----------------	-----------------------

Description

multiple gene set enrichemnt analysis (specific for testing GSE64810)

Usage

```
meaca_multiple1(expression_data1, trt, geneset, standardize = T,
  minSetSize = 5, fdr_method = "BH")
```

meaca_single	<i>meaca-single.</i>
--------------	----------------------

Description

meaca for single gene set test.

Usage

```
meaca_single(expression_data, trt, go_term, standardize = F)
```

Arguments

expression_data	the expressoin matrix.
trt	treatment labels.
go_term	an indicator vector. 1 for genes in the test, 0 otherwise.
standardize	whether the data should be standaridzed.

Value

a list	
set_name	The name of the gene set
testSetCor	Average correlation for genes in the test set
interCor	Average correlation between genes in the test set and those not in the test set
backSetCor	Average correlations for genes not in the test set.

MRGSE_multiple	<i>MRGSE multiple</i>
----------------	-----------------------

Description

multiple gene set enrichemnt analysis

Usage

```
MRGSE_multiple(expression, trt, geneset, use.rank = TRUE)
```

Arguments

expression	the expression data matrix
trt	the treatment lables for each column of expression_data
geneset	a list containing all the gene sets to be tested for enrichment status
use.rank	If TRUE then it corresponds to "MRGSE", otherwise "Camara"

Value

a matrix

MRGSE_multiple1	<i>MRGSE multiple</i>
-----------------	-----------------------

Description

multiple gene set enrichment analysis (specific for testing GSE64810)

Usage

```
MRGSE_multiple1(expression, trt, geneset, use.rank = T)
```

prepare_go_term	<i>Simulate gene labels.</i>
-----------------	------------------------------

Description

Simulate test set gene labels

Usage

```
prepare_go_term(num_gene, prop, delta)
```

Arguments

num_gene	a vector of length 2, total number of genes to be simulated and number of genes in the test set.
prop	a vector of length 2, proportion of DE genes within go term and outside go_term, corresponding to \$p_t\$ and \$p_b\$.
delta	a vector of length num_gene, representing DE effect for each gene

Value

a list	
go_term	DE indicator for each simulated gene. 1 is for genes in the test set, and 0 otherwise.
delta	DE effect for each simulated gene.
n_gene	number of genes to be simulated.

prepare_simulation	<i>Simulate expression data</i>
--------------------	---------------------------------

Description

Prepare simulation parameters.

Usage

```
prepare_simulation(num_gene, prop, delta, case, rho)
```

Arguments

num_gene	number of genes to be simulated.
prop	a vector of length 2, corresponding to p_t and p_b .
delta	a vector of length num_gene, representing DE effect for each gene
case	which correlation structure to simulate data from (see simulation setup in the paper)
rho	a vector of length 3. Corresponds ρ_1 , ρ_2 and ρ_3 .

Value

a list	
go_term	DE indicator for each simulated gene. 1 is for genes in the test set, and 0 otherwise
delta	DE effect for each simulated gene
sigma	the true correlation, used to simulate the expression data

read_gene_set	<i>Convert gene sets to lists</i>
---------------	-----------------------------------

Description

read the gene sets of the MsigDB format.

Usage

```
read_gene_set(msigdb)
```

Arguments

msigdb	gene set ensemble downloaded from broad institute.
--------	--

Value

a list

total number of gene sets contained.

size a numerical vector containing the size of each gene set.

gene_set a list. The first element is the set name. From the third element each containing members of the gene set.

sig_path	<i>Sigpathway</i>
----------	-------------------

Description

implement the Sigpathay.

Usage

```
sig_path(index, statistics, nsim = 9999)
```

Arguments

num_gene number of genes to be simulated.

Value

p the p-values.

simulate_expression_data	<i>Simulate expression data.</i>
--------------------------	----------------------------------

Description

simulate normally distributed expression data with desired DE probabilities for genes in the test set and for those not in the test set..

Usage

```
simulate_expression_data(size, obj)
```

Arguments

size number of samples to be simulated

obj object returned by prepare_simulation.

Value

a list	
data	a expression matrix of $m \times n$ where m is the number of genes and n is the number of samples.
trt	sample labels of length n, 1 for treatment and 0 for control.
go_term	gene labels of length m, 1 for go_term genes and 0 otherwise.
sigma	true covariance matrix upon which data is simulated.

`standardize_expression_data`*standardize expression data, with method described in the paper.*

Description

Standardize the expression data.

Usage

```
standardize_expression_data(expression_data, trt)
```

Arguments

expression_data	the expression matrix.
trt	sample labels. 0 for control and 1 for treatment

Value

a matrix of the same dimension as input data.

Index

btw_gene_corr, [2](#)

Camera_multiple, [2](#)
Camera_multiple1, [3](#)
compare_test, [3](#)

data_simu, [4](#)
data_simu_meaca, [4](#)

estimate_sigma, [5](#)

GSEA, [6](#)
GSEA.Gct2Frame2, [6](#)
GSEA.ReadClsFile, [7](#)
GSEA1, [7](#)

meaca_multiple, [8](#)
meaca_multiple1, [8](#)
meaca_single, [9](#)
MRGSE_multiple, [9](#)
MRGSE_multiple1, [10](#)

prepare_go_term, [10](#)
prepare_simulation, [11](#)

read_gene_set, [11](#)

sig_path, [12](#)
simulate_expression_data, [12](#)
standardize_expression_data, [13](#)