# Package 'meaca'

January 4, 2017

**Title** Mixed-effects Enrichment Analysis with Correlation Adjusted.

**Version** 0.0.0.0009

**Author** Bin Zhuo

**Maintainer** Bin Zhuo <zhuob@oregonstate.edu>, Duo Jiang <jiangd@stat.oregonstate.edu>.

**Description** This package produces all results needed in the paper.

**Depends** R (>= 3.2.1)

**License** What license is it under?

**LazyData** true

**RoxygenNote** 5.0.1

## R topics documented:



---

`btw_gene_corr` *Estimate sample correlation.*

---

### 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 data frame with columns:

| | |
|---|---|
| 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. |

---

## estimate_sigma *Estimate test-statistic covariance.*

---

### Description

Estimate test-statistic covariance

### 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 | estimated covariance matrix for the gene-level test statistics. |
| t_val | the gene level statistics, one for each gene |

meaca_multiple *meaca-multiple.*

## 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, 1 for treatment 0 for control. |
| geneset | gene sets to be tested, an object having the same structure as that returned from read_gene_set. |
| standardize | whether the data should be standaridzed. |
| minSetSize | the minimum number of genes contained in a gene set for it to be considered as a test set. |
| fdr_method | which method is used to adjust the p values. Options are "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none", see arguments in function p.adjust for more details. |

## Value

a data frame with the following columns

| | |
|---|---|
| set_name | the name of the gene set |
| set_sizel | the size of the gene set |
| status | "Up" for up-regulated gene sets and "Down" for down-regulated gene sets |
| p1 | chi square test p value |
| p1_fdr | adjusted p1, using p.adjust with appropriate method |
| p2 | a two-sided t-test p value equivalent to the chi square test |
| p2_fdr | adjusted p2 |

## Examples

```
m <- 100   # number of rows (genes)
n <- 50    # number of columns (samples)
y <- matrix(rnorm(5000), 100, 50)              # expression  matrix
rownames(y) <- paste("gene", 1:m, sep = "")
colnames(y) <- paste("sample", 1:n, sep = "")
trt <- rep(c(0, 1), each = n/2)                # treatment labels
## create the gene set format, desired for this function.
total <- 15                     # number of gene sets in the database
set.seed(100)
```

```
size <- sample(1:30, replace=F, total)
go_term <- rep(0, m)
gs_line <- list()
set_name <- paste("set", 1:length(size), sep = "")
for (i in 1:total) {
    set.seed(i)
    gs_line[[i]] <- rownames(y)[sample(1:m, size[i])]
}
gs <- list(total=total, size=size,set_name = set_name, gene_set=gs_line)
# run the multiple gene set test
result <- meaca_multiple(y, trt, gs)
```

---

`meaca_single` *Testing the enrichment status of a single pre-defined gene set.*

---

## Description

meaca for single gene set test.

## Usage

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

## Arguments

| | |
|---|---|
| `expression_data` | the expressoin matrix. |
| `trt` | treatment labels, should be 0s and 1s. |
| `go_term` | an indicator vector with value 1 for genes in the test, 0 otherwise. |
| `standardize` | whether the data should be standardized. Set to be TRUE for real data analysis. |

## Value

a list

| | |
|---|---|
| `stat` | test statistic for meaca |
| `p1` | chi square test p value |
| `status` | "Up" for up-regulated gene sets and "Down" for down-regulated gene sets |
| `p2` | a two-sided p value for conducting t-test from the test statstic `stat` |

## Examples

```
m <- 100   # number of rows (genes)
n <- 50   # number of columns (samples)
y <- matrix(rnorm(m*n), m, n)   # expression  matrix
trt <- rep(c(0, 1), each = n/2)      # treatment labels
go_term <- rep(0, m)
go_term[sample(1:m, 20)] <- 1     # the rows in the test set are labeled as 1
result <- meaca_single(y, trt, go_term)
```

## read_gene_set *Convert gene sets to lists*

### 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 corresponds to a member gene of the gene set. |

## 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

the standardized expression, a matrix of the same dimension as input expression data.

# Index