Package ‘RGSEA’

March 23, 2017

Type Package

Title Random Gene Set Enrichment Analysis

Version 1.9.0

Date 2014-04-22

Author Chengcheng Ma

biocViews GeneSetEnrichment, StatisticalMethod, Classification

Maintainer Chengcheng Ma <ccma@sibs.ac.cn>

Description Combining bootstrap aggregating and Gene set enrichment analysis (GSEA), RGSEA is a classification algorithm with high robustness and no over-fitting problem. It performs well especially for the data generated from different experiments.

License GPL(>=3)

Depends R(>= 2.10.0)

Imports BiocGenerics

VignetteBuilder knitr

Suggests BiocStyle, GEOquery, knitr, RUnit

NeedsCompilation no

R topics documented:

RGSEA-package ................................................. 2
cmap .......................................................... 2
e1 ............................................................. 3
e2 ............................................................. 4
RGSEAfix ..................................................... 5
RGSEApredict ................................................ 6
RGSEAsd ....................................................... 6

Index 8
**RGSEA-package**

**Random Gene Set Enrichment Analysis (RGSEA)**

**Description**

This is the package for similarity identification and classification of transcriptome data.

**Details**

- **Package:** RGSEA
- **Type:** Package
- **Version:** 1.0
- **Date:** 2014-04-22
- **License:** GPL(>=3)

~~ An overview of how to use the package, including the most important functions ~~

**Author(s)**

Chengcheng Ma Maintainer: tcmcc1986@gmail.com

**References**

~~ Literature or other references for background information ~~

**See Also**


**Examples**

```r
if(interactive()) {
  data(e1)
  data(e2)
  RGSEAf(e1,e2, queryclasses=colnames(e1), refclasses=colnames(e2),
  random=20000, featurenum=1000, iteration=100)->test
}
```

**cmap**

**Data from Connectivity map build 01**

**Description**

It is the sample data used for testing the function RGSEAsd.
Usage
data(cmap)

Format
The format is: num [1:22268, 1:6] 0.4892 -0.6137 3.5242 -0.0139 -2.0255 ... - attr(*, "dimnames")=List of 2 ..$ : chr [1:22268] "1007_s_at" "1053_at" "117_at" "121_at" ... ..$ : chr [1:6] "thioridazine" "tretinoin" "prochlorperazine" "chlorpromazine" ...

Details
The query data is the instance 5202764005791175120104.C08, treated with thioridazine.
The reference data is 5202764005789148112904.G05, X5202764005789148112904.F03, X5202764005789148112904.F05, X5202764005789148112904.E02, X5202764005789148112904.E04. They were treated with tretinoin, prochlorperazine, chlorpromazine, vorinostat, sirolimus respectively. All the data were generated by MCF7 cell line.

Source

References

Examples
data(cmap)

e1

Data from GDS4102

Description
This is the query data for testing the function RGSEAfix.

Usage
data(e1)

Format
The format is: num [1:54675, 1:2] 1012 44.32 43.03 36.65 4.92 ... - attr(*, "dimnames")=List of 2 ..$ : chr [1:54675] "1007_s_at" "1053_at" "117_at" "121_at" ... ..$ : chr [1:2] "tumor" "normal"

Details
The two data are GSM414924 GSM414975 repectively.
e2

Source

References
Pei H, Li L, Fridley BL, Jenkins GD et al. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. Cancer Cell 2009 Sep 8;16(3):259-66. PMID: 19732725

Examples
data(e1)

e2    Data from GDS4100

Description
This is the reference data for testing the function RGSEAfix.

Usage
data(e2)

Format
The format is: num [1:54675, 1:24] 1.48 1.19 0.67 2.75 NA 0.51 1.68 NA NA 1.99 ... - attr(*, "dimnames")=List of 2 ..$ : chr [1:54675] "1007_s_at" "1053_at" "117_at" "121_at" ... ..$ : chr [1:24] "tumor" "tumor" "tumor" "tumor" ...

Details
This dataset contains all the 4 data of GDS4100.

Source

References

Examples
data(e2)
**Random Gene Set Enrichment Analysis with fixed number of features**

**Description**

This is the function for classification and feature selection with fixed number of features from top and bottom of the subset features.

**Usage**

```r
RGSEAfix(query, reference, queryclasses, refclasses, random = 5000, featurenum = 500, iteration = 100)
```

**Arguments**

- **query**: A matrix. The query data. This is the data which the research wants to know the class.
- **reference**: A matrix. The reference data. Based on the reference data, the research infers the class of query data.
- **queryclasses**: A character vector. It contains the classes of query data. If you don’t know the classes of query data, just give it a character vector equal to the number of query data.
- **refclasses**: A character vector. It contains the classes of reference data. You must know it.
- **random**: A numeric variable. The number of features in the subset randomly sampled from the whole features each time.
- **featurenum**: A numeric variable. The number of features selected from top and bottom of the subset respectively.
- **iteration**: A numeric variable. The times of random sampling.

**Value**

1. The times of each sample in the reference dataset is the most similar to the query data.
2. The frequency of features selected from the top and bottom of the subsets from the query data, if the query data is correctly classified.

**Author(s)**

Chengcheng Ma

**Examples**

```r
if(interactive()) {
  data(e1)
  data(e2)
  RGSEAfix(e1, e2, queryclasses=colnames(e1), refclasses=colnames(e2),
            random=20000, featurenum=1000, iteration=100)->test
}
```
**RGSEApredict**

*Predict the class of the query data with the result of RGSEA functions*

**Description**

Predict the class of the query data with the result of RGSEA functions—RGSEAfix or RGSEAsd

**Usage**

\[
\text{RGSEApredict(RGSEAresult, refclasses)}
\]

**Arguments**

- **RGSEAresult**: The first item of the results generated by RGSEA functions.
- **refclasses**: A character vector. The classes of the reference data.

**Author(s)**

Chengcheng Ma

**Examples**

```r
if(interactive()) {
  data(e1)
  data(e2)
  RGSEAfix(e1, e2, queryclasses=colnames(e1), refclasses=colnames(e2),
            random=20000, featurenum=1000, iteration=100)->test
  RGSEApredict(test[[1]], colnames(e2))
}
```

**RGSEAsd**

*Random Gene Set Enrichment Analysis features selected based on standard deviation from the mean value*

**Description**

This is the function for classification with features from top and bottom deviating from the mean value of the whole transcriptome for a certain standard deviations of the subset features.

**Usage**

\[
\text{RGSEAsd(query, reference, queryclasses, refclasses, random = 5000, sd = 2, iteration = 100)}
\]
**Arguments**

- **query**: A matrix, The query data. This is the data which the research wants to know the class.
- **reference**: A matrix. The reference data. Based on the reference data, the research infers the class of query data.
- **queryclasses**: A character vector. It contains the classes of query data. If you don’t know the classes of query data, just give it a character vector equal to the number of query data.
- **refclasses**: A character vector. It contains the classes of reference data. You must know it.
- **random**: A numeric variable. The number of features in the subset randomly sampled from the whole features each time.
- **sd**: number of standard deviations the features selected from the subset deviate from the mean value of the subset.
- **iteration**: A numeric variable. The times of random sampling.

**Value**

1. The times of each sample in the reference dataset is the most similar to the query data. 
2. The frequency of features selected from the top and bottom of the subsets from the query data, if the query data is correctly classified.

**Author(s)**

Chengcheng Ma

**Examples**

```r
if(interactive()) {
  data(cmap)
  test <- RGSEAsd(cmap[,1], cmap[,2:6], queryclasses=colnames(cmap)[1],
                  refclasses=colnames(cmap)[2:6], random=5000, sd=2, iteration=100)
}
```
Index

*Topic datasets
  cmap, 2
  e1, 3
  e2, 4

*Topic package
  RGSEA-package, 2
  RGSEA (RGSEA-package), 2
  RGSEA-package, 2
  RGSEAf, 5
  RGSEApredict, 6
  RGSEAsd, 6