

Package ‘SISPA’

May 13, 2016

Type Package

Title SISPA: Method for Sample Integrated Set Profile Analysis

Version 1.2.1

Date 2015-10-22

Author Bhakti Dwivedi & Jeanne Kowalski

Maintainer Bhakti Dwivedi <bhakti.dwivedi@emory.edu>

Description Sample Integrated Gene Set Analysis (SISPA) is a method designed to define sample groups with similar gene set enrichment profiles.

Depends R (>= 3.2),GSVA,changepoint,data.table,ggplot2,plyr

License GPL-2

LazyData no

Collate 'SISPA.R' 'callZSCORE.R' 'callGSVA.R' 'cptSamples.R'
'waterfallplot.R' 'freqplot.R' 'sortData.R' 'gene_list.R'
'expression_data.R' 'variant_list.R' 'variant_data.R'

biocViews GeneSetEnrichment,GenomeWideAssociation

Suggests knitr

VignetteBuilder knitr

NeedsCompilation no

R topics documented:

callGSVA	2
callZSCORE	3
cptSamples	3
expression_data	4
freqplot	5
gene_list	6
SISPA	6
sortData	7
variant_data	8
variant_list	9
waterfallplot	9

callGSVA	<i>GSVA enrichment analysis</i>
----------	---------------------------------

Description

Estimates GSVA enrichment zscores.

Usage

```
callGSVA(x,y)
```

Arguments

x	A data frame or matrix of gene or probe expression values where rows correspond to genes and columns correspond to samples
y	A list of genes as data frame or vector

Details

This function uses "zscore" gene-set enrichment method in the estimation of gene-set enrichment scores per sample.

Value

A gene-set by sample matrix of GSVA enrichment zscores.

See Also

GSVA

Examples

```
g <- 10 ## number of genes
s <- 30 ## number of samples
## sample data matrix with values ranging from 1 to 10
rnames <- paste("g", 1:g, sep="")
cnames <- paste("s", 1:s, sep="")
expr <- matrix(sample.int(10, size = g*s, replace = TRUE), nrow=g, ncol=s, dimnames=list(rnames, cnames))
## genes of interest
genes <- data.frame(paste("g", 1:6, sep=""))
## Estimates GSVA enrichment zscores.
callGSVA(expr,genes)
```

`callZSCORE`*Row ZSCORES*

Description

Estimates the zscores for each row in the data matrix

Usage

```
callZSCORE(x)
```

Arguments

`x` A data frame or matrix of gene or probe expression values where rows correspond to genes and columns correspond to samples

Details

This function compute row zscores per sample when number of genes is less than 3

Value

A gene-set by sample matrix of zscores.

Examples

```
g <- 2 ## number of genes
s <- 60 ## number of samples
## sample data matrix with values ranging from 1 to 10
rnames <- paste("g", 1:g, sep="")
cnames <- paste("s", 1:s, sep="")
expr <- matrix(sample.int(10, size = g*s, replace = TRUE), nrow=g, ncol=s, dimnames=list(rnames, cnames))
## Estimates zscores
callZSCORE(expr)
```

`cptSamples`*Sample profile identifier analysis*

Description

Generate sample profile identifiers from sample zscores using change point model.

Usage

```
cptSamples(x, cpt_data, cpt_method, cpt_max)
```

Arguments

x	A matrix or data frame of sample GSVA enrichment zscores within which you wish to find a changepoint.
cpt_data	Identify changepoints for data using variance (cpt.var) or mean (cpt.mean). Default is cpt.var.
cpt_method	Choice of single or multiple changepoint model. Default is "BinSeg".
cpt_max	The maximum number of changepoints to search for using "BinSeg" method. Default is 60.

Details

This function assigns samples identified in the first changepoint with the active profile ("1") while the remaining samples are grouped under inactive profile ("0").

Value

The input data frame with added sample identifiers and estimated changepoints. A plot showing the changepoint locations estimated on the data

See Also

changepoint

Examples

```
g <- 10 ## number of genes
s <- 60 ## number of samples
## sample data matrix with values ranging from 1 to 10
rnames <- paste("g", 1:g, sep="")
cnames <- paste("s", 1:s, sep="")
expr <- matrix(sample.int(10, size = g*s, replace = TRUE), nrow=g, ncol=s, dimnames=list(rnames, cnames))
## genes of interest
genes <- data.frame(paste("g", 1:6, sep=""))
## Estimates GSVA enrichment zscores.
gsva_results <- callGSVA(expr,genes)
cptSamples(gsva_results,cpt_data="var",cpt_method="BinSeg",cpt_max=60)
```

expression_data

An example of RNAseq derived gene expression data

Description

This dataset contains the expression values of 8 probes (rows) in 125 samples (columns), as compiled by the CoMMpass study.

Usage

```
data(expression_data)
```

Details

This is data to be included in my package

Value

numeric expression dataset of 8 probes (rows) on 125 samples (column)

freqplot	<i>A plotting function for SISPA sample identifiers</i>
----------	---

Description

Given a sample changepoint data frame, will plot number of samples with and without profile activity

Usage

```
freqplot(x)
```

Arguments

x A data frame containing samples as rows followed by zscores and estimated changepoints to be plotted.

Details

This function expects the output from cptSamples function of SISPA package, and shows the number of samples with (orange filled bars) and without profile activity (grey filled bars).

Value

Bar plot pdf illustrating distribution of samples

Examples

```
samples <- c("s1", "s2", "s3", "s4", "s5", "s6", "s7", "s8", "s9", "s10")
zscores <- c(3.83, 2.70, 2.67, 2.31, 1.70, 1.25, -0.42, -1.01, -2.43, -3.37)
changepoints <- c(1, 1, 1, 2, 2, 3, 3, NA, NA, NA)
sample_groups <- c(1, 1, 1, 0, 0, 0, 0, 0, 0, 0)
my.data = data.frame(samples, zscores, changepoints, sample_groups)
freqplot(my.data)
```

gene_list	<i>an example file containing the list of input genes</i>
-----------	---

Description

A list of 8 genes for GSVA analysis

Usage

```
data(gene_list)
```

Details

This is data to be included in my package

Value

list of 8 genes

SISPA	<i>SISPA</i>
-------	--------------

Description

SISPA: Method for Sample Integrated Gene Set Analysis

Usage

```
SISPA(feature=1, f1.df, f1.genes, f1.profile, f2.df, f2.genes, f2.profile, cpt_data="var", cpt_method="BinS
```

Arguments

feature	Number of input feature or data types
f1.df	A data matrix of first feature (e.g., gene or probe expression values) where rows correspond to genes and columns correspond to samples
f1.genes	Gene sets for first feature provided as a vector or data frame
f1.profile	A flag to specify gene profile. If gene.profile="up" then samples with increased zscores are identified. If gene.profile="down" then samples with decreased zscores are identified. Default is "up".
f2.df	A data matrix of second feature (e.g., gene variant change) where rows correspond to genes and columns correspond to samples
f2.genes	Gene sets for second feature provided as a vector or data frame
f2.profile	A flag to specify gene profile. If gene.profile="up" then samples with increased zscores are identified. If gene.profile="down" then samples with decreased zscores are identified. Default is "up".

cpt_data	Identify changepoints for data using variance (cpt.var) or mean (cpt.mean). Default is cpt.var.
cpt_method	Choice of single or multiple changepoint model. Default is "BinSeg". See changepoint R package for details
cpt_max	The maximum number of changepoints to search for using "BinSeg" method. Default is 60.

Details

Sample Integrated Gene Set Analysis (SISPA) is a method designed to define sample groups with similar gene set enrichment profiles. The user specifies a gene list of interest and sample by gene molecular data (expression, methylation, variant, or copy change data) to obtain gene set enrichment scores by each sample. The score statistics is rank ordered by the desired profile (e.g., upregulated or downregulated) for samples. A change point model is then applied to the sample scores to identify groups of samples that show similar gene set profile patterns. Samples are ranked by desired profile activity score and grouped by samples with and without profile activity. Figure 1 shows the schematic representation of the SISPA method overview.

Value

The input molecular data frame with added sample identifiers and estimated changepoints. A plot showing the changepoint locations estimated on the data. Bar plots pdf illustrating distinct distribution of samples with and without profile activity

Examples

```
g <- 10 ## number of genes
s <- 60 ## number of samples
## sample data matrix with values ranging from 1 to 10
rnames <- paste("g", 1:g, sep="")
cnames <- paste("s", 1:s, sep="")
expr <- matrix(sample.int(10, size = g*s, replace = TRUE), nrow=g, ncol=s, dimnames=list(rnames, cnames))
## genes of interest
genes <- data.frame(paste("g", 1:6, sep=""))
SISPA(feature=1, f1.df=expr, f1.genes=genes, f1.profile="up")
```

sortData

Sorts the data by a column

Description

Sorts the data frame by a column index in the given order

Usage

```
sortData(x, i, b)
```

Arguments

x	A data frame
i	A numeric column index of the data frame to sort it by
b	User specified sorting order, ascending (FALSE) or descending (TRUE)

Details

defaults are used: i = 1, b = FALSE, if not specified

Value

sorted data by the input column index

Author(s)

Bhakti Dwivedi & Jeanne Kowalski

Examples

```
samples <- c("s1", "s2", "s3", "s4", "s5", "s6", "s7", "s8", "s9", "s10")
zscores <- c(3.83, 2.70, 2.67, 2.31, 1.70, 1.25, -0.42, -1.01, -2.43, -3.37)
my.data = data.frame(samples, zscores)
sortData(my.data, 2, TRUE)
```

variant_data

An example of RNAseq derived gene variant change data

Description

This dataset contains the variant proportion values of variants (n=380) associated with 8 genes (rows) in 125 samples (columns), as compiled by the CoMMpass study.

Usage

```
data(variant_data)
```

Details

This is data to be included in my package

Value

numeric variant dataset of 380 variants (rows) on 125 samples (column)

variant_list	<i>an example file containing the list of input genes variants</i>
--------------	--

Description

A list of variants associated with 8 genes for GSVA analysis

Usage

```
data(variant_list)
```

Details

This is data to be included in my package

Value

list of 380 gene variants

waterfallplot	<i>A plotting function for SISPA sample identifiers</i>
---------------	---

Description

Given a sample changepoint data frame, will plot all samples zscores from that data.

Usage

```
waterfallplot(x)
```

Arguments

x	A data frame containing samples as rows followed by zscores and estimated sample_groups to be plotted.
---	--

Details

This function expects the output from cptSamples function of SISPA package, and highlights the sample profile of interest in the changepoint 1 with orange filled bars.

Value

Bar plot pdf illustrating distinct SISPA sample profiles.

Examples

```
samples <- c("s1", "s2", "s3", "s4", "s5", "s6", "s7", "s8", "s9", "s10")
zscores <- c(3.83, 2.70, 2.67, 2.31, 1.70, 1.25, -0.42, -1.01, -2.43, -3.37)
changepts <- c(1, 1, 1, 2, 2, 3, 3, NA, NA, NA)
sample_groups <- c(1, 1, 1, 0, 0, 0, 0, 0, 0, 0)
my.data = data.frame(samples, zscores, changepts, sample_groups)
waterfallplot(my.data)
```

Index

*Topic **datasets**

expression_data, [4](#)

gene_list, [6](#)

variant_data, [8](#)

variant_list, [9](#)

callGSVA, [2](#)

callZSCORE, [3](#)

cptSamples, [3](#)

expression_data, [4](#)

freqplot, [5](#)

gene_list, [6](#)

SISPA, [6](#)

sortData, [7](#)

variant_data, [8](#)

variant_list, [9](#)

waterfallplot, [9](#)