

# Package ‘SISPA’

April 15, 2020

**Type** Package

**Title** SISPA: Method for Sample Integrated Set Profile Analysis

**Version** 1.16.0

**Date** 2018-04-03

**Author** Bhakti Dwivedi and Jeanne Kowalski

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

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

**Depends** R (>= 3.2),genefilter,GSVA,changeoint

**Imports** data.table, plyr, ggplot2

**License** GPL-2

**LazyData** true

**Collate** 'SISPA.R' 'callZSCORE.R' 'callGSVA.R' 'cptSamples.R'  
'waterfallplot.R' 'freqplot.R' 'sortData.R' 'filterVars.R'  
'expression\_data.R' 'variant\_data.R'

**biocViews** GeneSetEnrichment,GenomeWideAssociation

**Suggests** knitr

**VignetteBuilder** knitr

**NeedsCompilation** no

**git\_url** <https://git.bioconductor.org/packages/SISPA>

**git\_branch** RELEASE\_3\_10

**git\_last\_commit** dd721df

**git\_last\_commit\_date** 2019-10-29

**Date/Publication** 2020-04-14

## R topics documented:

callGSVA . . . . .	2
callZSCORE . . . . .	3
cptSamples . . . . .	3
expression_data . . . . .	4
filterVars . . . . .	5
freqplot . . . . .	5

SISPA . . . . .	6
sortData . . . . .	7
variant_data . . . . .	8
waterfallplot . . . . .	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

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 <- paste("g", 1:g, sep="")
## Estimates GSVA enrichment zscores.
callGSVA(expr,genes)
```

---

callZSCORE	<i>Row ZSCORES</i>
------------	--------------------

---

**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	<i>Sample profile identifier analysis</i>
------------	-------------------------------------------

---

**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), mean (cpt.mean) or both (cpt.meanvar). 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 <- paste("g", 1:g, 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)

---

filterVars	<i>A filter function for the data</i>
------------	---------------------------------------

---

**Description**

Filter rows with zero values

**Usage**

```
filterVars(x,y)
```

**Arguments**

**x** : A data frame or matrix where rows represent gene and columns represent samples

**y** : A vector of a sample column values to apply the filtering on.

**Details**

This function filter out rows with zero data value for a given sample. Both input arguments (x and y) must be of the same length

**Value**

The returned value is a list containing an entry for each row filtered out by zero data value

**Examples**

```
x = matrix(runif(3*10, 0, 1), ncol=3)
y <- x[,1]
filterVars(x,y)
```

---

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)
changeoints <- 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, changeoints, sample_groups)
freqplot(my.data)
```

---

SISPA

*SISPA*


---

**Description**

SISPA: Method for Sample Integrated Gene Set Analysis

**Usage**

```
SISPA(feature=1, f1.df, f1.profile, f2.df, f2.profile, cpt_data="var", cpt_method="BinSeg", cpt_max=60)
```

**Arguments**

<code>feature</code>	Number of input feature or data types
<code>f1.df</code>	A data matrix of first feature (e.g., gene or probe expression values) where rows correspond to genes and columns correspond to samples
<code>f1.profile</code>	A flag to specify gene profile. If <code>gene.profile="up"</code> then samples with increased zscores are identified. If <code>gene.profile="down"</code> then samples with decreased zscores are identified. Default is "up".
<code>f2.df</code>	A data matrix of second feature (e.g., gene variant change) where rows correspond to genes and columns correspond to samples
<code>f2.profile</code>	A flag to specify gene profile. If <code>gene.profile="up"</code> then samples with increased zscores are identified. If <code>gene.profile="down"</code> then samples with decreased zscores are identified. Default is "up".
<code>cpt_data</code>	Identify changepoints for data using variance ( <code>cpt.var</code> ), mean ( <code>cpt.mean</code> ), or both ( <code>cpt.meanvar</code> ). Default is <code>cpt.var</code> .
<code>cpt_method</code>	Choice of single or multiple changepoint model. Default is "BinSeg". See changepoint R package for details
<code>cpt_max</code>	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))
SISPA(feature=1, f1.df=expr, 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)

---

waterfallplot

*A plotting function for SISPA sample identifiers*

---

**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)
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)
waterfallplot(my.data)
```

# Index

## \*Topic **datasets**

expression\_data, [4](#)  
variant\_data, [8](#)

callGSVA, [2](#)  
callZSCORE, [3](#)  
cptSamples, [3](#)

expression\_data, [4](#)

filterVars, [5](#)  
freqplot, [5](#)

SISPA, [6](#)  
sortData, [7](#)

variant\_data, [8](#)

waterfallplot, [8](#)