Package 'pathVar'

April 16, 2019

Type Package

Title Methods to Find Pathways with Significantly Different Variability

Version 1.12.0

Date 2018-06-29

Author Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Maintainer Samuel Zimmerman <samuel.e.zimmerman@gmail.com>

Description This package contains the functions to find the pathways that have significantly different variability than a reference gene set. It also finds the categories from this pathway that are significant where each category is a cluster of genes. The genes are separated into clusters by their level of variability.

License LGPL (>= 2.0)

Collate classdef.R pipeline.final.R

LazyData true

Depends R (>= 3.3.0), methods, ggplot2, gridExtra

Imports EMT, mclust, Matching, data.table, stats, grDevices, graphics, utils

biocViews GeneticVariability, GeneSetEnrichment, Pathways

NeedsCompilation no

git_url https://git.bioconductor.org/packages/pathVar

git_branch RELEASE_3_8

git_last_commit 368f103

git_last_commit_date 2018-10-30

Date/Publication 2019-04-15

R topics documented:

`pathVar-package` *Detects pathways with different levels of variance than reference gene set variability.*

Description

This package contains functions used to determine pathways with significant differences in variability.

Details

Package:	pathVar
Type:	Package
Version:	1.11.2
Date:	2018-06-29
License:	LGPL
Depends:	R (>= 3.2.2), methods, ggplot2, gridExtra
Imports:	EMT, mclust, Matching, reshape, data.table

1. Compute the standard deviation for each gene. 2. Classify the genes with respect to sd in at most 4 clusters. 3. For each pathway, we extract the gene in our dataset and in which cluster they belong. 4. For each pathway, we look how its genes are distributed in each category and compare it to the expected number with all the gene from the dataset with the chisq. 5. Same as 4. but with the exact test 6. find significant pathway(s), which category(ies) from this pathway are significant and which gene(s) belongs to this(ese) category(ies)

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
results_kegg=pathVarOneSample(bock,pways.kegg,test="chisq",varStat="sd")
sig_kegg=sigPway(results_kegg,0.05)
```

---

bock *Gene Expression Matrix of Published Data*

Description

This is a matrix object containing genes as rows and the samples as columns, where each element in the matrix is an expression value. The dataset contains 12900 genes and 32 samples.

Usage

```
data(bock)
```

Value

A matrix with expression values for 12900 genes probes on 32 samples.

See Also

pways.kegg, pways.reactome

Examples

```
data(bock)
```

---

diagnosticsVarPlots *Plots the average expression against variability using different summary statistics.*

Description

Plots the average expression against variability using different summary statistics to help decide which statistic would be the best with your dataset.

Usage

```
diagnosticsVarPlots(dat.mat)
```

Arguments

dat.mat	a matrix with rows as genes and columns as samples.

Details

This function gives you 3 plots one for SD, one for MAD and one for CV against the mean to help you decide which one would be the best with your dataset. It also returns the correlation between each variability statistics and the mean.

Value

3 scatter plots where average expression is on the X-axis and SD, MAD, and CV are on the Y-axis.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
diagnosticsVarPlots(bock)
```

`diagnosticsVarPlotsTwoSample` *Plots the average expression against variability using different summary statistics when comparing 2 groups of samples to each other.*

Description

Plots the average expression against variability using different summary statistics to help decide which statistic would be the best with your dataset when comparing 2 groups of samples to each other.

Usage

```
diagnosticsVarPlotsTwoSample(dat.mat,groups)
```

Arguments

`dat.mat`	a `matrix` with rows as genes and columns as samples.
`groups`	`vector` indicating the amount of samples and replicates of each sample.

Details

This function gives you 3 plots one for SD, one for MAD and one for CV against the mean to help you decide which one would be the best with your dataset. It also returns the correlation between each variability statistics and the mean.

Value

3 scatter plots where average expression is on the X-axis and SD, MAD, and CV are on the Y-axis.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
diagnosticsVarPlotsTwoSample(bock[1:5000,], groups=as.factor(c(rep(1,10),rep(2,10))))
```

geneDistributionSet-class
Class "geneDistributionSet"

Description

This is a class representation for storing the output of the pathVarOneSample function.

Objects from the Class

Objects are output by the function pathVarOneSample. Objects can also be created by calls of the form new("geneDistributionSet", ...).

Slots

tablePway: A "data.table" of pathway name, pathway IDs, adjusted p-value from the chisq test or exact test, the number of genes from our dataset inside the pathway and the total number of genes inside the pathway.

NAPways: A character object that contains the pathway names of the pathway having less than 10 genes for the Chi-Squared or also more than 500 genes for the exact test.

genesInPway: A list object that contains each pathway with the genes from the datasets belonging to it and in which cluster they were classsify.

refProb: A table object that contains the probability of the reference in each cluster.

refCounts: A table object that contains the genes counts of the reference in each cluster.

pwayCounts: A list object that contains the genes counts of the each pathway in each cluster.

numOfClus: A numeric object that contains the number of clusters.

varStat: A character object that contains the statistics sd, mad, cv or mean chosen for the analysis.

genesInClus: A numeric object that contains all the genes from the dataset and in which cluster they belong.

var: A numeric object that contains the statistics value (sd, mad, cv or mean) for each gene.

Methods

No methods defined with class "geneDistributionSet" in the signature.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
out <- new("geneDistributionSet",tablePway=data.table::data.table(),NAPways=character(),genesInPway=list(),
showClass("geneDistributionSet")
```

geneDistributionSet2-class
Class "pathVarTwoSamplesCont"

Description

This is a class representation for storing the output of the pathVarTwoSamplesCont function.

Objects from the Class

Objects are output by the function pathVarTwoSamplesCont. Objects can also be created by calls of the form new("geneDistributionSet2", ...).

Slots

tablePway: A "data.table" of pathway name, pathway IDs, adjusted p-value from the boot KS test, the number of genes from our dataset inside the pathway and the total number of genes inside the pathway.

NAPways: A character object that contains the pathway names of the pathway having no genes inside the dataset.

genesInPway: A list object that contains the genes from the dataset belonging to each pathway.

groups: A factor object that contains the groups in which group each sample belongs to.

groupNames: A character object that contains the names of the two groups.

var1: A numeric object that contains the statistics (sd, mad, cv or mean) for each gene for group 1.

var2: A numeric object that contains the statistics (sd, mad, cv or mean) for each gene for group 2.

varStat: A character object that contains the statistics sd, mad, cv or mean chosen for the analysis.

Methods

No methods defined with class "geneDistributionSet2" in the signature.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
out <- new("geneDistributionSet2",tablePway=data.table::data.table(),NAPways=character(),genesInPway=list()
showClass("geneDistributionSet2")
```

geneDistributionSet3-class
Class "pathVarTwoSamplesDisc"

Description

This is a class representation for storing the output of the pathVarTwoSamplesDisc function.

Objects from the Class

Objects are output by the function pathVarTwoSamplesDisc. Objects can also be created by calls of the form new("geneDistributionSet3", ...).

Slots

tablePway: A "data.table" of pathway name, pathway IDs, adjusted p-value from the boot KS test, the number of genes from our dataset inside the pathway and the total number of genes inside the pathway.

NAPways: A character object that contains the pathway names of the pathway having no genes inside the dataset.

genesInPway1: A list object that contains the genes from the dataset belonging to each pathway in the first group.

genesInPway2: A list object that contains the genes from the dataset belonging to each pathway in the second group.

pwayCounts1: A list object that contains tables of the number of genes in each cluster per pathway for the first group

pwayCounts2: A list object that contains tables of the number of genes in each cluster per pathway for the second group

groups: A factor object that contains the groups in which group each sample belongs to.

groupNames: A character object that contains the names of the two groups.

var1: A numeric object that contains the statistics (sd, mad, cv or mean) for each gene for group 1.

var2: A numeric object that contains the statistics (sd, mad, cv or mean) for each gene for group 2.

varStat: A character object that contains the statistics sd, mad, cv or mean chosen for the analysis.

Methods

No methods defined with class "geneDistributionSet3" in the signature.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
out <- new("geneDistributionSet3",tablePway=data.table::data.table(),NAPways=character(),genesInPway1=list(
showClass("geneDistributionSet3")
```

geneSet-class *Class* "geneSet"

Description

This is a class representation for storing the output of the getGenes function.

Objects from the Class

Objects are output by the function getGenes. Objects can also be created by calls of the form new("geneSet", ...).

Slots

genes1: A character object that contains the genes belonging to the pathway in the given window for group 1.

genes2: A character object that contains the genes belonging to the pathway in the given window for group 2.

genesAll: A character object that contains the genes from the dataset belonging to the pathway.

Methods

No methods defined with class "geneSet" in the signature.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
out <- new("geneSet",genes1=character(),genes2=character(),genesAll=character())
showClass("geneSet")
```

getGenes *Gets significant genes within a certain window of variability.*

Description

Gets significant genes within a certain window of variability.

Usage

```
getGenes(pvalue_results,pathway,window)
```

Arguments

pvalue_results	output of pathVarTwoSamplesCont step.
pathway	A pathway name.
window	A vector with a min and max range to specify the window of genes to grab with a certain level of variance.

Details

It takes the result of `pathVarTwoSamplesCont`, a given pathway and "window". It will give you the genes having their variability value in the window for group 1 and another set of genes for group 2 corresponding to the given pathway. It also returns the set of all the genes from your dataset that belong to this pathway.

Value

An object of class `geneSet` with 3 properties. The first 2 are the genes in the window from each group. The third are all the genes.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
# we run the 2 samples analysis on the first 10 pathways from kegg
pways.kegg.10pways <- lapply(pways.kegg, function(x) x[1:10])
results_2samples=pathVarTwoSamplesCont(bock,pways.kegg.10pways,groups=as.factor(c(rep(1,10),rep(2,10))))
genes_window=getGenes(results_2samples,pways.kegg$PATHNAME[10],c(0.25,075))
```

`makeDBList` *Puts your own list of pathways and genes related to them into a list.*

Description

Puts your own list of pathways and genes related to them into a list.

Usage

```
makeDBList(file,pID=TRUE)
```

Arguments

`file`	a tab deimited text file where the first and second columns are pathwayID and pathway name. The third, or last column, is the genes associated with each pathway, seperated by commas. If no pathwayIDs are available, the first column should be the pathway name.
`pID`	boolean expression specifiying where the pathways have an ID or not.

Details

This function is given a tab delimited text file with the pathwayID (if available), the pathway name, and the genes associated with each pathway. It outputs a list where the first element is the pathway names, the second is the Pathway IDs and the third is a list of genes for each pathway. The first row should be the column names.

Value

A list of pathway names, IDs, and genes in each pathway is returned.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
## Not run:
pways.ipa <- makeDBList("fileName",pID=FALSE)

## End(Not run)
```

---

pathVarOneSample — *Compares the distribution of genes in each cluster to the distribution of genes in each cluster for every pathway.*

Description

Compares the distribution of genes in each cluster to the distribution of genes in each cluster for every pathway.

Usage

```
pathVarOneSample(dat.mat, pways,test=c("chisq", "exact"),varStat=c("sd", "mean", "mad", "cv"))
```

Arguments

dat.mat	matrix with the genes on the rows and the samples on the columns.
pways	list which contains a vector of pathway IDs, a vector of pathway names, and a list of genes in each pathway.
test	a string specifying the type of significance test to perform. The options are "exact" or "chisq".
varStat	a string specifying the type of variability summary statistic to perform. The options are "sd", "mean", "mad", or "cv".

Details

This function classifies your genes into one to four clusters with respect to the standard deviation (SD), median absolute deviation (MAD), coefficient of variation (CV) or mean. Then, it compares the counts of genes in each class from your dataset in one pathway with the counts of the genes in each class from the whole dataset. For that, it uses a Chi-square or an exact test. You can give your own list of pathways (using the output of makeDBList) or use Reactome and KEGG pathways that are already included.

Value

A geneDistributionSet object is returned.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
results_kegg=pathVarOneSample(bock,pways.kegg,test="chisq",varStat="sd")
```

pathVarTwoSamplesCont *Compares the distribution of genes in each pathway for two groups of samples that you define.*

Description

Compares the distribution of genes in each pathway for two groups of samples that you define.

Usage

```
pathVarTwoSamplesCont(dat.mat, pways,groups,boot=1000,varStat=c("sd","mean", "mad", "cv"))
```

Arguments

dat.mat	matrix with the genes on the rows and the samples on the columns.
pways	list which contains a vector of pathway IDs, a vector of pathway names, and a list of genes in each pathway.
groups	vector indicating the amount of samples and replicates of each sample.
boot	number of bootstraps to be performed.
varStat	a string specifying the type of variability summary statistic to perform. The options are "sd", "mean", "mad", or "cv".

Details

This function splits the samples into two groups that you define. It compares the density of the variability (SD, MAD, CV) or of the mean of the genes in a pathway from group 1 with the density from group 2. For that, it uses the bootstrap Kolmogorov-smirnov test. You can give your own list of pathways (using the output of makeDBList) or use Reactome and KEGG pathways that are already included.

Value

A geneDistributionSet2 object is returned.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
# we run the 2 samples analysis on the first 10 pathways from kegg
pways.kegg.10pways <- lapply(pways.kegg, function(x) x[1:10])
results_2samples=pathVarTwoSamplesCont(bock,pways.kegg.10pways,groups=as.factor(c(rep(1,10),rep(2,10))),boo
```

pathVarTwoSamplesDisc *Compares the number of genes in clusters in each pathway for two groups of samples that you define.*

Description

Compares the distribution of genes in each pathway for two groups of samples that you define.

Usage

```
pathVarTwoSamplesDisc(dat.mat,pways,groups,perc=c(1/3,2/3), test=c("chisq", "exact"),varStat=c("
```

Arguments

dat.mat	matrix with the genes on the rows and the samples on the columns.
pways	list which contains a vector of pathway IDs, a vector of pathway names, and a list of genes in each pathway.
groups	vector indicating the amount of samples and replicates of each sample.
perc	numeric vector of probabilities with values between 0 and 1. Used to put genes into clusters
test	a string, either "exact" or "chisq" which are tests to see if clusters in the 2 samples are sig. different from each other
varStat	a string specifying the type of variability summary statistic to perform. The options are "sd", "mean", "mad", or "cv".

Details

This function splits the samples into two groups that you define. It computes the variability (sd, mad, cv, or mean) for each gene in each group. Then it classifies the genes with respect to the variability in at most 4 clusters. For each pathway, we extract the gene in our dataset and in which cluster they belong. Then for each pathway we look at the gene counts in each category and compare the 2 samples to each other with all the genes from the data set with the Chi-Squared or exact test.

Value

A geneDistributionSet3 object is returned.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
# we run the 2 samples analysis on the first 10 pathways from kegg
pways.kegg.10pways <- lapply(pways.kegg, function(x) x[1:10])
results_2samples=pathVarTwoSamplesDisc(bock,pways.kegg.10pways,groups=as.factor(c(rep(1,10),rep(2,10))),per
```

plotAllTwoSampleDistributionCounts

Compares the distribution of genes in clusters for every gene in your data set for two groups of samples that you define and plot the counts.

Description

Compares the distribution of genes for every gene in your data set for two groups of samples that you define and plot the counts of each gene.

Usage

```
plotAllTwoSampleDistributionCounts(dat,pvalue_results,perc=c(1/3,2/3),pvalue,plotName)
```

Arguments

dat	matrix with the genes on the rows and the samples on the columns.
pvalue_results	output of pathVarTwoSamples step.
perc	numeric vector of probabilities with values between 0 and 1. Used to put genes into clusters
pvalue	the significance level to test.
plotName	a string specifying the file name of the output plot. Default value is NULL which prints, but does not save the graph in a pdf.

Details

This function looks at the gene counts in each category of 2 groups previously defined in the pathVarTwoSamples step and compare the 2 samples to each other with all the genes from the data set. Then it plots the counts for each group.

Value

A plot of results returned.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
# we run the 2 samples analysis on the first 10 pathways from kegg
pways.kegg.10pways <- lapply(pways.kegg, function(x) x[1:10])
results_2samples=pathVarTwoSamplesDisc(bock,pways.kegg.10pways,groups=as.factor(c(rep(1,10),rep(2,10))),per
plotAllTwoSampleDistributionCounts(bock, results_2samples, perc=c(1/3,2/3), pvalue=0.05, "bock.group1.group
```

plotPway *Checks if an object is from the one sample or two samples cases and then plots reference distribution and the chosen pathway.*

Description

Checks if an object is from the one sample or two samples cases and then plots reference distribution and the chosen pathway.

Usage

```
plotPway(pvalue_results,pathway,sig)
```

Arguments

pvalue_results	output of pathVarTwoSamples or pathVarOneSample step.
pathway	the chosen pathway you want to plot.
sig	output of sigPway or NULL.

Details

Plots the results of the one or two samples case for a chosen pathway.

Value

A plot of the results from sigPway.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
# we run the 2 samples analysis on the first 10 pathways from kegg
pways.kegg.10pways <- lapply(pways.kegg, function(x) x[1:10])
results_2samples=pathVarTwoSamplesCont(bock,pways.kegg.10pways,groups=as.factor(c(rep(1,10),rep(2,10))))
sigPways <- sigPway(results_2samples,0.05)
plotPway(results_2samples,"Glycolysis / Gluconeogenesis",sigPways)
```

pways.kegg *List containing pathway IDs, names, and genes in each pathway*

Description

This is a list containing a vector of pathway IDs, a vector of pathway names, and a list of the genes in each pathway. There are a total of 272 KEGG pathways.

Usage

```
data(pways.kegg)
```

Value

A list containing a vector of pathway IDs, a vector of pathway names, and a list of the genes in each pathway for 272 pathways.

See Also

bock, pways.reactome

Examples

```
data(pways.kegg)
```

---

pways.reactome *List containing pathway IDs, names, and genes in each pathway*

Description

This is a list containing a vector of pathway IDs, a vector of pathway names, and a list of the genes in each pathway. There are a total of 946 reactome pathways.

Usage

```
data(pways.reactome)
```

Value

A list containing a vector of pathway IDs, a vector of pathway names, and a list of the genes in each pathway for 946 pathways.

See Also

bock, pways.kegg

Examples

```
data(pways.reactome)
```

---

saveAsPDF *Save the plots of the significant pathway or a chosen list of pathways as a pdf.*

Description

Save the plots of the significant pathway or a chosen list of pathways as a pdf.

Usage

```
saveAsPDF(pvalue_results,sig,listPath="significant")
```

Arguments

pvalue_results	output of pathVarTwoSamples or pathVarOneSample step.
sig	output of sigPway or NULL.
listPath	the string "significant" if you want to save the plots of the significant pathways or a list of names of pathways of interest.

Details

Save the plots of the significant pathway or a chosen list of pathways as a pdf.

Value

A pdf of the results from sigPway.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
# we run the 2 samples analysis on the first 10 pathways from kegg
pways.kegg.10pways <- lapply(pways.kegg, function(x) x[1:10])
results_2samples=pathVarTwoSamplesCont(bock, pways.kegg.10pways,groups=as.factor(c(rep(1,10),rep(2,10))))
sigPways <- sigPway(results_2samples,0.05)
saveAsPDF(results_2samples,sigPways,listPath="significant")
```

---

significantPathway-class

Class "significantPathway"

Description

This is a class representation for storing the output of the sigPway function in the one sample case.

Objects from the Class

Objects are output by the function sigPway in the one sample case. Objects can be created by calls of the form new("significantPathway", ...).

Slots

genesInSigPways1: A "list" object that contains the genes per significant pathway belonging to the significant category.

sigCatPerPway: A "list" object that contains the category(ies) per pathway that are significant.

thresPValue: A "numeric" object that contains the chosen p-value for the significance.

Methods

No methods defined with class "significantPathway" in the signature.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
sig <- new("significantPathway",genesInSigPways1=list(),sigCatPerPway=list(),thresPValue=numeric())
showClass("significantPathway")
```

---

significantPathway2-class

Class "significantPathway2"

Description

This is a class representation for storing the output of the sigPway function in the two sample case.

Objects from the Class

Objects are output by the function sigPway in the two sample case. Objects can be created by calls of the form new("significantPathway2", ...).

Slots

genesInSigPways1: A "list" ojbect that contains the genes belonging to each significant pathway.

thresPValue: A "numeric" object that contains the chosen p-value for the significance.

Methods

No methods defined with class "significantPathway2" in the signature.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
sig <- new("significantPathway2",genesInSigPways1=list(),thresPValue=numeric())
showClass("significantPathway2")
```

significantPathway3-class *Class* "significantPathway3"

Description

This is a class representation for storing the output of the sigPway function in the two sample case.

Objects from the Class

Objects are output by the function sigPway in the two sample case. Objects can be created by calls of the form new("significantPathway3", ...).

Slots

genesInSigPways1: A "list" ojbect that contains the genes belonging to each significant pathway in group 1.

genesInSigPways2: A "list" ojbect that contains the genes belonging to each significant pathway in group 2.

sigCatPerPway: A "list" ojbect that contains the significant categories in each pathway.

thresPValue: A "numeric" object that contains the chosen p-value for the significance.

Methods

No methods defined with class "significantPathway3" in the signature.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
sig <- new("significantPathway3",genesInSigPways1=list(),genesInSigPways2=list(),sigCatPerPway=list(),thres
showClass("significantPathway3")
```

sigPway *A function checks if an object is from the one sample or two samples cases and then finds the significant pathways.*

Description

A function checks if an object is from the one sample or two samples cases and then finds the significant pathways.

Usage

```
sigPway(pvalue_results, pvalue)
```

Arguments

pvalue_results	geneDistributionSet2 object and the output of the pathVarTwoSamples function or geneDistributionSet object and the output of the pathVarOneSample.
pvalue	the significance level to test.

Details

This function takes the output pathVarTwoSamples or pathVarOneSample and returns a significantPathway2 or significantPathway object specifying the significant pathway(s) and which gene(s) belongs to the significant pathway(s).

Value

A significantPathway2 object or significantPathway object is returned.

Author(s)

Laurence de Torrente, Samuel Zimmerman, Jessica Mar

Examples

```
results_2samples=pathVarTwoSamplesCont(bock,pways.kegg,groups=as.factor(c(rep(1,10),rep(2,10))),boot=1000,v
sig_kegg=sigPway(results_2samples,0.05)
```

Index