Package 'psSubpathway'

March 18, 2024

Type Package

Title Flexible Identification of Phenotype-Specific Subpathways

Version 0.1.3

Date 2024-3-18

Author Xudong Han,
Junwei Han,
Qing Fei

Maintainer Junwei Han `<hanjunwei1981@163.com>`

Description A network-based systems biology tool for flexible identification of phenotype-specific subpathways in the cancer gene expression data
with multiple categories (such as multiple subtype or developmental stages of cancer). Subtype Set Enrichment Analysis (SubSEA) and Dynamic Changed
Subpathway Analysis (DCSA) are developed to flexible identify subtype specific and dynamic changed subpathways respectively. The operation modes
include extraction of subpathways from biological pathways, inference of subpathway activities in the context of gene expression data, identification
of subtype specific subpathways with SubSEA, identification of dynamic changed subpathways associated with the cancer developmental stage with DCSA,
and visualization of the activities of resulting subpathways by using box plots and heat maps. Its capabilities render the tool could find the specific
abnormal subpathways in the cancer dataset with multi-phenotype samples.

Depends R (>= 3.5.0)

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.0

Imports GSVA, igraph, mpmi, pheatmap

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation yes

Repository CRAN

Date/Publication 2024-03-18 08:40:06 UTC

R **topics documented:**

---

compute_rank_score   *psSubpathway internal functions*

Description

Compute rank score.

Usage

```
SubSEA

DCSA
```

Details

compute_rank_score

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

DCSA   *Dynamic Changed Subpathway Analysis (DCSA)*

Description

This function will perform the Dynamic Changed Subpathway Analysis (DCSA) method to estimate the dynamic changed subpathways associated with the sample phenotypes (like the developmental stage of cancer).

Usage

```
DCSA(
  expr,
  input.cls = "",
  subpathwaylist = "Symbol",
  kcdf = "Gaussian",
  method = "gsva",
  min.sz = 5,
  max.sz = 1000,
  nperm = 100,
  fdr.th = 1,
  mx.diff = TRUE,
  parallel.sz = 0
)
```

Arguments

expr	Matrix of gene expression values (rows are genes, columns are samples).
input.cls	Input sample phenotype class vector file in CLS format.
subpathwaylist	Character string denoting the gene label of the subpathway list is 'Entrezid' (default) or 'Symbol'. Users can also enter their own subpathway list data. This list should be consistent with the gene label in the input gene expression profile.
kcdf	Character string denoting the kernel to use during the non-parametric estimation of the cumulative distribution function of expression levels across samples when method="gsva". By default, 'kcdf="Gaussian"' which is suitable when input expression values are continuous, such as microarray fluorescent units in logarithmic scale, RNA-seq log-CPMs, log-RPKMs or log-TPMs. When input expression values are integer counts, such as those derived from RNA-seq experiments, then this argument should be set to 'kcdf="Poisson"'.
method	Method to employ in the estimation of subpathway enrichment scores per sample. By default, this is set to 'gsva' (Hänzelmann et al, 2013) and other options are 'ssgsea' (Barbie et al, 2009).
min.sz	Minimum size of the resulting subpathway.
max.sz	Maximum size of the resulting subpathway.

nperm	Number of random permutations (default: 100).In practice, the users can set their own values as needed, and more than 1000 times may be fine in general.
fdr.th	Cutoff value for FDR. The only subpathway with lower FDR.th are listed (default: 1).
mx.diff	Offers two approaches to calculate the sample enrichment score (SES) from the KS random walk statistic. 'mx.diff=FALSE': SES is calculated as the maximum distance of the random walk from 0. 'mx.diff=TRUE' (default): SES is calculated as the magnitude difference between the largest positive and negative random walk deviations.
parallel.sz	Number of processors to use when doing the calculations in parallel. If this argument is left with its default value (parallel.sz=0) then it will use all available core processors unless we set this argument with a smaller number.

Details

DCSA

This function calculates the subpathway activity profile based on the gene expression profile and subpathway list by 'gsva' or 'ssgssea'. Next, we used the information-theoretic measure of statistical dependence, mutual information (MI), to estimate the dynamically changed subpathways associated with the sample phenotypes. Finally we used the perturbation analysis of the gene label rearrangement to estimating the statistical significance.

Value

A list containing the results of DCSA and subpathway activity profile.

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

Examples

```
# load depend package.
require(GSVA)
require(parallel)
require(mpmi)
# get ACC disease stage gene expression profiling.
# ACCgenematrix<-get("DCgenematrix")
# get path of the sample disease stage phenotype files.
# Stagelabels<-system.file("extdata", "DClabels.cls", package = "psSubpathway")
# perform the DCSA method.
# DCSA(ACCgenematrix,input.cls=Stagelabels,nperm=50,fdr.th=0.01,parallel.sz=2)
# get the result of the SubSEA function
# DCSAresult<-get("DCspwresult")
# str(DCSAresult)
# head(DCSAresult$DCSA)

# Simulated gene matrix.
genematrix <- matrix(rnorm(500*40), nrow=500, dimnames=list(1:500, 1:40))
# Construct subpathway list data.
```

```
subpathwaylist <- as.list(sample(2:100, size=20, replace=TRUE))
subpathwaylist <- lapply(subpathwaylist, function(n) sample(1:500, size=n, replace=FALSE))
names(subpathwaylist)<-c(paste(rep("spw",20),c(1:20)))
# Construct sample labels data.
stagelabel<-list(phen=c("stage1","stage2","stage3","stage4"),
                  class.labes=c(rep("stage1",10),rep("stage2",10),
                  rep("stage3",10),rep("stage4",10)))
DCSAcs<-DCSA(genematrix,stagelabel,subpathwaylist,nperm=0,parallel.sz=1)
```

---

envData	*The variables in the environment include subpathway list data, information of subpathway,an expression profile and a example result*

Description

We used the k-clique algorithm to divide the human pathway of the KEGG database into subpathways and eliminated the smaller module that had a overlap above 80 These subpathway data will be divided into two parts: subpathway information(spwnetworkdata,spwtitle) and subpathway list data(spwentrezidlist: the gene Entrezid contained in each subpathway and spwsymbollist:the gene symbol contained in each subpathway). Subgenematrix is the gene expression profile of the breast cancer we selected from GDCTCGA and Subspwresult is the result of applying Subgenematrix to Subtype Set Enrichment Analysis (SubSEA).In order to reduce the memory, we delete the gene in the gene expression profile that is not in the gene of the subpathway list. We also selected the gene expression profile of the adrenocortical cancer (ACC) from GDCTCGA and performed Dynamic Changed Subpathway Analysis (DCSA). The data are DCgenematrix and DCspwresult In order to reduce the memory, we delete the gene in the gene expression profile that is not in the gene of the subpathway list. respectively. DClabes.cls and Sublabels.cls are the label vectors of the phenotypes of the samples of the two gene expression profiles.

Format

An environment variable

Author(s)

Junwei Han<hanjunwei1981@163.com>,Xudong Han <HanXD1217@163.com>

FastSEAscore   *psSubpathway internal functions*

Description

Fast calculate phenotypic set enrichment score.

Usage

    SubSEA

    DCSA

Details

FastSEAscore

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

---

getgenedensity   *psSubpathway internal functions*

Description

Calculating the Kernel estimation of gene.

Usage

    SubSEA

    DCSA

Details

getgenedensity

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

---

isPackageLoaded *psSubpathway internal functions*

Description

Determine if the package is loaded, if no package is loaded.

Usage

SubSEA

DCSA

Details

isPackageLoaded

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

---

ks_test_m *psSubpathway internal functions*

Description

Calculating subpathway Variation score.

Usage

SubSEA

DCSA

Details

ks_test_m

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

---

plotheatmap   *Plot a heatmap*

Description

Plot a heatmap of subpathway activity profile based on the parameters set by the user.

Usage

```
plotheatmap(
  inputdata,
  plotSubSEA = TRUE,
  fdr.th = 1,
  SES = "positive",
  phenotype = "all"
)
```

Arguments

inputdata	A list of result data generated by function 'SubSEA' or 'DCSA'.
plotSubSEA	Determine the inputdata is the result data of function 'SubSEA' (default:plotSubSEA=TRUE) or 'DCSA' (plotSubSEA=FLASE).
fdr.th	Cutoff value for FDR. The only subpathway with lower FDR is plotted. (default: 1).
SES	Parameter 'SES' is useful only when 'plotSubSEA' is TRUE. When 'plotSub-SEA=TRUE',if 'SES' is positive, the subpathway with high-expression will be plotted. when it is negative, plot low-expression subpathways.
phenotype	Parameter 'phenotype' is useful only when 'plotSubSEA' is TRUE. 'phenotype' decides which phenotypic significant subpathways to screen (which phenotypic result is applied to parameter 'fdr.th' and 'SES'.) and plot a heat map of these subpathways.By default,'phenotype="all"' which will screen the subpathways of all phenotypes and plot a heat map. When the user wants to plot a subpathway heat map of the specified phenotype, this parameter should be set to the name of the phenotype.

Details

plotheatmap

The subpathways are screened according to the conditions set by the user and a heat map of the activity of these subpathways is drawn.

Value

a heatmap

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

Examples

```
# load depend package.
library(pheatmap)
# get the Subspwresult which is the result of SubSEA function.
Subspwresult<-get("Subspwresult")
# get the DCspwresult which is the result of DCSA function.
DCspwresult<-get("DCspwresult")
# plot significant up-regulation subpathway heat map specific for each breast cancer subtype.
plotheatmap(Subspwresult,plotSubSEA=TRUE,fdr.th=0.01,SES="positive",phenotype="all")
# plot significant down-regulation subpathway heat map specific for each breast cancer subtype.
plotheatmap(Subspwresult,plotSubSEA=TRUE,fdr.th=0.01,SES="negative",phenotype="all")
# plot basal subtype specific significant subpathway heat map.
plotheatmap(Subspwresult,plotSubSEA=TRUE,fdr.th=0.01,SES="all",phenotype="Basal")
# plot adrenocortical cancer disease stages specific significant subpathway heat map.
plotheatmap(DCspwresult,plotSubSEA=FALSE,fdr.th=0.01)
```

---

plotSpwACmap   *Plot subpathway activity change map*

Description

Plot a box diagram and heat map of subpathway activity in each phenotype.

Usage

```
plotSpwACmap(inputdata, spwid = "")
```

Arguments

inputdata   A list of result data generated by function 'SubSEA' or 'DCSA'.

spwid   The subpathway id which the user wants to plot.

Details

plotSpwACmap

Plot a box diagram of subpathway activity in each Phenotype and a heat map of the distribution of the phenotypic samples in the activity of the subpathways. The subpathway activity change map includes subpathway active change box plot and subpathway active change. Each row in the heat map is all samples of a phenotype. These samples are distributed in the subpathway high activity value area label is red, and the distribution in the low expression value area label is blue.

Value

a plot

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

Examples

```
# get the Subspwresult which is the result of SubSEA method.
Subspwresult<-get("Subspwresult")
# plot the subpathway 00120_9 in the SubSEA function result.
plotSpwACmap(Subspwresult,spwid="00120_9")
# get the DCspwresult which is the result of DCSA method.
DCspwresult<-get("DCspwresult")
# plot the subpathway 00982_2 in the DCSA function result.
plotSpwACmap(DCspwresult,spwid="00982_2")
```

---

plotSpwNetmap   *Polt a subpathway network map*

Description

Visualize a subpathway network map.

Usage

```
plotSpwNetmap(
  spwid,
  layout = NULL,
  margin = 0,
  vertex.label.cex = 0.6,
  vertex.label.font = 1,
  vertex.size = 8,
  vertex.size2 = 6,
  edge.arrow.size = 0.2,
  edge.arrow.width = 3,
  edge.label.cex = 0.6,
  vertex.label.color = "black",
  vertex.color = "#BFFFBF",
  vertex.frame.color = "dimgray",
  edge.color = "dimgray",
  edge.label.color = "dimgray",
  sub = NULL,
  main = NULL
)
```

Arguments

spwid	The subpathway id which the user wants to plot.
layout	A matrix of x-y coordinates with two dims. Determine the placement of the nodes for drawing a graph.

margin	A numeric. The value is usually between -0.5 and 0.5, which is able to zoom in or out a subpathway graph. The default is 0.
vertex.label.cex	
	A numeric vector of node label size.
vertex.label.font	
	A numeric vector of label font.
vertex.size	A numeric vector of Node size. See `plot.igraph`.
vertex.size2	A numeric vector of Node size.
edge.arrow.size	
	Edge arrow size.The default is 0.2.
edge.arrow.width	
	Edge arrow width. The default is 3.
edge.label.cex	Edge label size.
vertex.label.color	
	A vector of node label colors. The default is black.
vertex.color	A vector of node colors. The default is the KEGG node color.
vertex.frame.color	
	A vector of node frame color. The default is dimgray.
edge.color	A vector of edge color. The default is dimgray.
edge.label.color	
	A vector of edge label color. The default is dimgray.
sub	A character string of subtitle.
main	A character string of main title.

Details

plotSpwNetmap

The function plotSpwNetmap is able to display a subpathway graph. The argument layout is used to determine the placement of the nodes for drawing a graph.The layouts provided in igraph include 'layout_as_star', 'layout_as_tree', 'layout_in_circle', 'layout_nicely','layout_on_grid', 'layout_on_sphere', 'layout_randomly', 'layout_with_dh', 'layout_with_fr', 'layout_with_gem', 'layout_with_graphopt', 'layout_with_kk', 'layout_with_lgl', 'layout_with_mds'. The 'layout_as_tree' generates a tree-like layout, so it is mainly for trees. The 'layout_randomly' places the nodes randomly. The 'layout_in_circle' places the nodes on a unit circle. Detailed information on the parameters can be found in `layout_`

Value

a plot

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

Examples

```
# load depend package.
library(igraph)
# plot network graph of the subpathway 00982_2
plotSpwNetmap(spwid="00982_2",layout=layout_nicely)
```

---

plotSpwPSheatmap   *Plot subpathway phenotypic significant heat map*

Description

Visualize subpathway activity significant heat map between phenotypes

Usage

```
plotSpwPSheatmap(inputdata, spwid = "")
```

Arguments

inputdata   A list of result data generated by function 'SubSEA' or 'DCSA'.

spwid   The subpathway id which the user wants to plot.

Details

plotSpwPSheatmap

A heat map of the significance P-value of the activity of the subpathway between the phenotypes. The rows and columns of the heat map are sample phenotype labels. The values shown in the heat map are the T-test P values between the activity of the subpathway corresponding to the two phenotypes. The lower the number in the cells in the heat map, the greater the difference in the activity of the subpathways between the two phenotypes.

Value

a plot

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

Examples

```
# get the Subspwresult which is the result of SubSEA method.
Subspwresult<-get("Subspwresult")
# plot significant heat map between the activity of the subpathway in each subtype of breast cancer.
plotSpwPSheatmap(Subspwresult,spwid="00120_9")
# get the DCspwresult which is the result of DCSA method.
DCspwresult<-get("DCspwresult")
# plot significant heat map between the activity of the subpathway in each stage of breast cancer.
plotSpwPSheatmap(DCspwresult,spwid="00982_2")
```

---

plotSubSEScurve *Plot subtype set sample enrichment score curve graph*

Description

Draw a sample enrichment score curve graph of a single or all subtypes.

Usage

```
plotSubSEScurve(inputdata, spwid = "", phenotype = "all")
```

Arguments

inputdata	A list of result data generated by function 'SubSEA' or 'DCSA'.
spwid	The subpathway id which the user wants to plot.
phenotype	The 'phenotype" specifies which phenotypic phenotype set enrichment curve is drawn for subpathway. When 'phenotype="all"' (default), the phenotype set enrichment score curve graph of subpathway under all phenotypes is drawn.

Details

plotSubSEScurve

Plot a phenotype set enrichment score curve graph of a subpathway under all phenotypes or specified phenotype, including the location of the maximum enrichment score (ES) and the leading-edge subset. This function can only be used for the results of the 'SubSEA' function.

Value

a plot

an enrichment score curve graph

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

Examples

```
# get the results of the SubSEA function for breast cancer subtypes.
Subspwresult<-get("Subspwresult")
# plot enrichment score curve of the subpathway 00120_9 in all breast cancer subtypes.
plotSubSEScurve(Subspwresult,spwid="00120_9",phenotype="all")
# plot enrichment score curve of the subpathway 00120_9 in the basal breast cancer subtypes.
plotSubSEScurve(Subspwresult,spwid="00120_9",phenotype="Basal")
```

ReadClsFile   *psSubpathway internal functions*

Description

These are functions read sample label file (.cls format).

Usage

SubSEA

DCSA

Details

ReadClsFile

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

---

rndWalk   *psSubpathway internal functions*

Description

Calculating random walks.

Usage

SubSEA

DCSA

Details

rndWalk

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

SEAscore   *psSubpathway internal functions*

Description

Get subtype set enrichment score and sample locations, etc.

Usage

```
plotSubSEScurve
```

Details

SEAscore

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

---

ssgsea   *psSubpathway internal functions*

Description

Single sample GSEA (ssGSEA) calculates a gene set enrichment score per sample as the normalized difference in empirical cumulative distribution functions of gene expression ranks inside and outside the gene set.

Usage

```
SubSEA

DCSA
```

Details

ssgsea

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

SubSEA   *Subtype Set Enrichment Analysis (SubSEA)*

Description

The SubSEA (Subtype Set Enrichment Analysis) method to mine the specific subpathways of each sample Subtype.

Usage

```
SubSEA(
  expr,
  input.cls = "",
  subpathwaylist = "Symbol",
  kcdf = "Gaussian",
  method = "gsva",
  min.sz = 1,
  max.sz = Inf,
  nperm = 100,
  fdr.th = 1,
  mx.diff = TRUE,
  parallel.sz = 0
)
```

Arguments

expr	Matrix of gene expression values (rows are genes, columns are samples).
input.cls	Input sample class vector (phenotype) file in CLS format.
subpathwaylist	Character string denoting the gene label of the subpahtway list is 'Entrezid' (default) or 'Symbol'. Users can also enter their own subpathway list data. This list should be consistent with the gene label in the input gene expression profile.
kcdf	Character string denoting the kernel to use during the non-parametric estimation of the cumulative distribution function of expression levels across samples when method="gsva". By default, 'kcdf="Gaussian"' which is suitable when input expression values are continuous, such as microarray fluorescent units in logarithmic scale, RNA-seq log-CPMs, log-RPKMs or log-TPMs. When input expression values are integer counts, such as those derived from RNA-seq experiments, then this argument should be set to 'kcdf="Poisson"'.
method	Method to employ in the estimation of subpathway enrichment scores per sample. By default,this is set to 'gsva' (Hänzelmann et al, 2013) and other options are 'ssgsea' (Barbie et al, 2009).
min.sz	Minimum size of the resulting subpathway.
max.sz	Maximum size of the resulting subpathway.
nperm	Number of random permutations (default: 100). In practice, the users can set their own values as needed, and more than 1000 times may be fine in general.

fdr.th	Cutoff value for FDR. The only subpathway with lower fdr.th are listed (default: 1).
mx.diff	Offers two approaches to calculate the sample enrichment score (SES) from the KS random walk statistic. 'mx.diff=FALSE': SES is calculated as the maximum distance of the random walk from 0. 'mx.diff=TRUE' (default): SES is calculated as the magnitude difference between the largest positive and negative random walk deviations.
parallel.sz	Number of processors to use when doing the calculations in parallel. If this argument is left with its default value (parallel.sz=0) then it will use all available core processors unless we set this argument with a smaller number.

Details

SubSEA

This function calculates the subpathway activity profile based on the gene expression profile and subpathway list by 'gsva' or 'ssgssea'. Then we calculate the sample enrichment score (SES) of each subpathway by Subtype Set Enrichment Analysis (SubSEA).We permute the gene labels and recompute the SES for the permuted data, which generates a null distribution for the SES.The P-value and the FDR value are calculated according to the perturbation analysis.

Value

A list containing the results of the SubSEA and the subpathway activity profile.

Author(s)

Xudong Han, Junwei Han, Qingfei Kong

Examples

```
# load depend package.
require(GSVA)
require(parallel)
# get breast cancer disease subtype gene expression profile.
Bregenematrix<-get("Subgenematrix")
# get path of the sample disease subtype files.
Subtypelabels<- system.file("extdata", "Sublabels.cls", package = "psSubpathway")
# SubSEA(Bregenematrix,input.cls=Subtypelabels,nperm=50,fdr.th=0.01,parallel.sz=2)
# get the result of the SubSEA function
SubSEAresult<-get("Subspwresult")
str(SubSEAresult)
head(SubSEAresult$Basal)

# Simulated gene matrix
genematrix <- matrix(rnorm(500*40), nrow=500, dimnames=list(1:500, 1:40))
# Construct subpathway list data.
subpathwaylist <- as.list(sample(2:100, size=20, replace=TRUE))
subpathwaylist <- lapply(subpathwaylist, function(n) sample(1:500, size=n, replace=FALSE))
names(subpathwaylist)<-c(paste(rep("spw",20),c(1:20)))
# Construct sample labels data.
```

```
subtypelabel<-list(phen=c("subtype1","subtype2","subtype3","subtype4"),
                   class.labes=c(rep("subtype1",10),rep("subtype2",10),
                                 rep("subtype3",10),rep("subtype4",10)))
SubSEAcs<-SubSEA(genematrix,subtypelabel,subpathwaylist,nperm=0,parallel.sz=1)
```

Index