Package 'ICDS'

January 17, 2019

Type Package

Title Identification of Cancer Dysfunctional Sub-pathway by integrating DNA methylation, copy number variation, and gene expression data

Version 0.1.0

Author Junwei Han,Baotong Zheng,Siyao Liu

Maintainer Junwei Han <hanjunwei1981@163.com>

Description Identify Cancer Dysfunctional Sub-pathway by integrating gene expression, DNA methylation and copy number variation, and pathway topological information. 1)We firstly calculate the gene risk scores by integrating three kinds of data: DNA methylation, copy number variation, and gene expression. 2)Secondly, we perform a greedy search algorithm to identify the key dysfunctional sub-pathways within the pathways for which the discriminative scores were locally maximal. 3)Finally, the permutation test was used to calculate statistical significance level for these key dysfunctional sub-pathways.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Suggests knitr,
rmarkdown,
prettydoc

VignetteBuilder knitr

Imports igraph,
graphite,
metap,
methods,
org.Hs.eg.db

Depends R (>= 2.10)

Roxygen list(markdown = TRUE)

RoxygenNote 6.1.1

R topics documented:

`ICDS-package` *Identification of Cancer Dysfunctional Subpathway by integrating DNA methylation, copy number variation, and gene expression data*

Description

Identify Cancer Dysfunctional Subpathway by integrating gene expression, DNA methylation and copy number variation, and pathway topological information. 1)We firstly calculate the gene risk scores by integrating three kinds of data: DNA methylation, copy number variation, and gene expression. 2)Secondly, we perform a greedy search algorithm to identify the key dysfunctional subpathways within the pathways for which the discriminative scores were locally maximal. 3)Finally, the permutation test was used to calculate statistical significance level for these key dysfunctional subpathways.

`combinep_three` *combinep_three*

Description

`combinep_three` combine three kinds of p-values,then,calculate z-score for them.

Usage

```
combinep_three(p1, p2, p3)
```

Arguments

`p1`	the p-values or corrected p-values
`p2`	the p-values or corrected p-values
`p3`	the p-values or corrected p-values

Value

A numeric vector of z_scores

Examples

```
exp.p<-GetExampleData("exp.p")
meth.p<-GetExampleData("meth.p")
cnv.p<-GetExampleData("cnv.p")
## Not run: combinep_three(exp.p,meth.p,cnv.p)
```

combinep_two *combinep_two*

Description

combinep_two combine two kinds of p-values,then,calculate z-score for them.

Usage

```
combinep_two(p1, p2)
```

Arguments

p1	A numeric vector of p-values or corrected p-values
p2	A numeric vector of p-values or corrected p-values

Value

A numeric vector of z_scores

Examples

```
exp.p<-GetExampleData("exp.p")
meth.p<-GetExampleData("meth.p")
## Not run: combinep_two(exp.p,meth.p)
```

coverp2zscore *coverp2zscore*

Description

coverp2zscore calculate z-scores for p-values

Usage

```
coverp2zscore(pdata)
```

Arguments

pdata	A numeric vector of p-values or corrected p-values

Value

A numeric vector of z_scores

Examples

```
exp.p<-GetExampleData("exp.p")
meth.p<-GetExampleData("meth.p")
cnv.p<-GetExampleData("cnv.p")
## Not run: coverp2zscore(exp.p)
## Not run: coverp2zscore(meth.p)
## Not run: coverp2zscore(cnv.p)
```

envData *The variables in the environment include an example expression profile,an methylation profile,an copy number variation data,amplified genes,deleted genes,A numeric vector of z_scores,p-values,A vector of 0/1s, indicating the class of samples,interested subpathways,Optimized subpathway,and the statistical significance p value and FDR for these optimal subpathways*

Description

Identify Cancer Dysfunctional Subpathway by integrating gene expression, DNA methylation and copy number variation, and pathway topological information. 1)We firstly calculate the gene risk scores by integrating three kinds of data: DNA methylation, copy number variation, and gene expression. 2)Secondly, we perform a greedy search algorithm to identify the key dysfunctional subpathways within the pathways for which the discriminative scores were locally maximal. 3)Finally, the permutation test was used to calculate statistical significance level for these key dysfunctional subpathways.

Format

An environment variable

Details

The environment variable includes the variable `exp_data`, `meth_data`,`cnv_data`,`amp_gene`,`del_gene`,`zzz`,`exp.p`,`meth.p`

Author(s)

Junwei Han<hanjunwei1981@163.com>,Baotong Zheng<btzheng1116@163.com>,Siyao Liu <liusiyao29@163.com>

FindSubPath *FindSubPath*

Description

`FindSubPath` uses a greedy search algorithm to search for key subpathways in each entire pathway.

Usage

```
FindSubPath(zz, Pathway = "kegg", delta = 0.05, seed_p = 0.05,
  min.size = 5, out.F = FALSE, out.file = "Subpath.txt")
```

Arguments

`zz`	A numeric vector of z_scores.
`Pathway`	The name of the pathway database.
`delta`	Diffusion coefficient in each step of searching subpath.
`seed_p`	Define gene whose p-value smaller than seed_p as seed gene.
`min.size`	The smallest size of subpathways.
`out.F`	Logical,tell if output subpathways.
`out.file`	file name of subpathways.

Value

Key dysfunctional subpathways in each pathway, in which the risk score of the genes were significantly higher.

Examples

```
require(graphite)
zz<-GetExampleData("zzz")
## Not run: k<-FindSubPath(zz)
```

---

getCnvp *getCnvp*

Description

getCnvp perform t-test on copy number variation data

Usage

```
getCnvp(exp_data, cnv_data, amp_gene, del_gene, p.adjust = TRUE,
  method = "fdr")
```

Arguments

exp_data	A data frame
cnv_data	Copy number variation data
amp_gene	A vector of strings, the IDs of amplified genes.
del_gene	A vector of strings, the IDs of deleted genes.
p.adjust	Logical,tell if returns corrected p-values
method	Correction method,which can be one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY",

Details

cnv_data is TCGA level4 data.if p.adjust=TRUE,return corrected p-values,if p.adjust=FALSE,return p-values

Value

A numeric vector of p-values or corrected p-values

Examples

```
exp_data<-GetExampleData("exp_data")
meth_data<-GetExampleData("meth_data")
cnv_data<-GetExampleData("cnv_data")
amp_gene<-GetExampleData("amp_gene")
del_gene<-GetExampleData(("del_gene"))
## Not run: getCnvp(exp_data,cnv_data,amp_gene,del_gene,p.adjust=FALSE,method="fdr")
```

GetExampleData *Get the example data*

Description

Get the example data of test package for litte trials.

Usage

```
GetExampleData(exampleData)
```

Arguments

exampleData	A character, should be one of "exp_data", "meth_data", "cnv_data", "amp_gene", "del_gene" ,"label1","label2","zz","exp.p","meth.p","cnv.p"and "pathdata".

Details

The function getExampleData(ExampleData = "exp.p)") obtains a vector of lncRNAs confirmed to be related with breast cancer. The function getExampleData(ExampleData = "Profile") obtains the expression pr

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

getExpp *getExpp*

Description

getExpp perform t-test on Expression profile data

Usage

```
getExpp(exp_data, label, p.adjust = TRUE, method = "fdr")
```

Arguments

exp_data	A data frame, the expression profile to calculate p-value for each gene, the rownames should be the symbol of genes.
label	A vector of 0/1s, indicating the class of samples in the expression profile, 0 represents case, 1 represents control.
p.adjust	Logical,tell if returns corrected p-values
method	Correction method,which can be one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY",

Details

For a given expression profile of two conditions, ICDS package provide t-test method to calculate p-values or corrected p-values(if p.adjust=TRUE,return corrected p-values,if p.adjust=FALSE,return p-values.) for each genes. The row of the expression profile should be gene symbols and the column of the expression profile should be names of samples. Samples should be under two conditions and the label should be given as 0 and 1.

Value

A numeric vector of p-values or corrected p-values

Examples

```
profile<-GetExampleData("exp_data")
label<-GetExampleData("label1")
## Not run: getExpp(profile,label,p.adjust=FALSE)
```

`getMethp` *getMethp*

Description

`getMethp` perform t-test on Methylation profile data

Usage

```
getMethp(meth_data, label, p.adjust = TRUE, method = "fdr")
```

Arguments

`meth_data`	A data frame, the Methylation profile to calculate p-value for each gene, the rownames should be the symbol of genes.
`label`	label A vector of 0/1s, indicating the class of samples in the Methylation profile, 0 represents case, 1 represents control.
`p.adjust`	Logical,tell if returns corrected p-values
`method`	Correction method,which can be one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY",

Details

For a given Methylation profile of two conditions, ICDS package provide t-test method to calculate p-values or corrected p-values(if p.adjust=TRUE,return corrected p-values,if p.adjust=FALSE,return p-values.) for each genes. The row of the Methylation profile should be gene symbols and the column of the Methylation profile should be names of samples. Samples should be under two conditions and the label should be given as 0 and 1.

Value

A numeric vector of p-values or corrected p-values

Examples

```
profile<-GetExampleData("meth_data")
label<-GetExampleData("label2")
## Not run: getMethp(profile,label,p.adjust=FALSE)
```

`opt_subpath` *opt_subpath*

Description

`opt_subpath` Optimize interested subpathways.If the number of genes shared by the two pathways accounted for more than the Overlap ratio of each pathway genes,then combine two pathways.

Usage

```
opt_subpath(subpathdata, zz, overlap = 0.6)
```

Arguments

`subpathdata`	interested subpathways
`zz`	a vector of z-scores
`overlap`	Overlap ratio of each two pathway genes

Value

Optimized subpathway:the number of genes shared by any two pathways accounted for less than the Overlap ratio of each pathway genes.

Examples

```
zz<-GetExampleData("zzz")
subpathdata<-GetExampleData("subpathdata")
## Not run: optsubpath<-opt_subpath(subpathdata,zz,overlap=0.6)
```

`Permutation` *Permutation*

Description

the permutation test method 1 and method 2 were used to calculate the statistical significance level for these optimal subpathways.

Usage

```
Permutation(subpathwayz, zz, nperm1 = 1000, method1 = TRUE,
  nperm2 = 1000, method2 = FALSE)
```

Arguments

subpathwayz	Optimize intersted subpathways
zz	a vector of z-scores
nperm1	times of permutation to perform use method1
method1	permutation analysis method1
nperm2	times of permutation to perform use method2
method2	permutation analysis method2

Value

the statistical significance p value and FDR for these optimal subpathways

Examples

```
require(graphite)
keysubpathways<-GetExampleData("keysubpathways")
zzz<-GetExampleData("zzz")
## Not run: Permutation(keysubpathways,zzz,nperm1=10,method1=TRUE,nperm2=10,method2=FALSE)
```

PlotSubpathway *PlotSubpathway*

Description

PlotSubpathway:plot a network graph when user input a list of gene

Usage

```
PlotSubpathway(subpID, pathway.name, zz, Pathway = "kegg",
  layout = layout.fruchterman.reingold)
```

Arguments

subpID	gene list of a interested subpathway
pathway.name	name of the interested subpathway
zz	z-score of each gene
Pathway	the name of the pathway database
layout	The layout specification(layout_). It must be a call to a layout specification function.

Value

Network graph

Examples

```
require(graphite)

subpID<-unlist(strsplit("G6PC/HK3/GPI/FBP1/ALDOA/G6PC2","/"))
pathway.name="Glycolysis / Gluconeogenesis"
zzz<- GetExampleData("zzz")
## Not run: PlotSubpathway(subpID=subpID,pathway.name=pathway.name,zz=zzz)
```

Index