Package 'DSviaDRM'

May 12, 2015

Type Package

Title Exploring Disease Similarity in Terms of Dysfunctional Regulatory Mechanisms

Version 1.0

Date 2015-05-12

Author Jing Yang

Maintainer Jing Yang <yangjing@scbit.org>

Description Elucidation of human disease similarities has emerged as an active research area, which is highly relevant to etiology, disease classification, and drug repositioning. This package was designed and implemented for identifying disease similarities. It contains five functions which are 'DCEA', 'DCpathway', 'DS', 'comDCGL' and 'comDCGLplot'. In 'DCEA' function, differentially co-expressed genes and differentially co-expressed links are extracted from disease vs. health samples. Then 'DCpathway' function assigns differential co-expression values of pathways to be the average differential co-expression value of their component genes. Then 'DS' employs partial correlation coefficient of pathways as the disease similarity for each disease pairs. And 'DS' contains a permutation process for evaluating the statistical significant of observed disease partial correlation coefficients. At last, 'comDCGL' and 'comDCGLplot' sort out shared differentially co-expressed genes and differentially co-expressed links with regulation information and visualize them.

Depends R (>= 2.10)

Imports igraph, ppcor

License GPL (> 2)

LazyLoad yes

LazyData true

Repository CRAN

NeedsCompilation no

Date/Publication 2015-05-12 11:45:13

R topics documented:

DSviaDRM-package . . . 2
comDCGL . . . 4
comDCGLplot . . . 5
DCEA . . . 7
DCpathway . . . 9
DS . . . 10
expressionBasedfilter . . . 12
exprs1 . . . 13
exprs2 . . . 14
exprs3 . . . 14
pathways . . . 15
tf2target . . . 16
varianceBasedfilter . . . 17

Index **18**

`DSviaDRM-package` *Exploring disease similarity in terms of dysfunctional regulation mechanisms*

Description

DSviaDRM package contains five functions which are DCEA, DCpathway, DS, comDCGL and comDCGLplot. In DCEA function, differentially co-expressed genes (DCGs) and differentially co-expressed links (DCLs) are extracted from disease vs. health samples. Then DCpathway function assigns differential co-expression values (dCs) of pathways to be the average dC of their component genes. Then DS employs partial correlation coefficient as the disease similarity for each disease pairs. And DS contains a permutation process for evaluating the statistical significant of observed disease partial correlation coefficients. At last, comDCGL and comDCGLplot sort out shared DCGs and DCLs with regulation information and visualize them.

Details

Package:	DSviaDRM
Type:	Package
Version:	1.0
Date:	2015-05-07
License:	GPL (>2)

~~ An overview of how to use the package, including the most important ~~ ~~ functions ~~

Author(s)

Jing Yang

Maintainer: Jing Yang

References

Yang J, Wu S-J, Li Y-Y, Li Y-X. The human disease network in terms of dysfunctional regulatory mechanisms. (2015)

Examples

```
######################################################################
## computate DCEA results for three disease, Allergic asthma (AA),
## Chronic kidney disease (CKD) and Type 2 Diabetes (T2D).
######################################################################
#data(exprs1)
#data(exprs2)
#data(exprs3)

##########################################################################
## the default value of cutoff in DCEA is 0.25,
## here cutoff is set to 1 for saving time when demonstrating the examples.
#DCEA.AA.res<-DCEA(exprs1[1:200,1:5],exprs1[1:200,6:10],link.method="percent",
# cutoff=1,N=0,nbins=20,p=0.1)
#DCEA.CKD.res<-DCEA(exprs2[1:300,1:25],exprs2[1:300,26:31],link.method="percent",
# cutoff=1,N=0,nbins=20,p=0.1)
#DCEA.T2D.res<-DCEA(exprs3[1:200,1:12],exprs3[1:200,13:35],link.method="percent",
# cutoff=1,N=0,nbins=20,p=0.1)
##########################################################################

##########################################################################
## computate dCs of pathways for each disease
#data(pathways)
#DCpathway.AA.res<-DCpathway(DCEA.res=DCEA.AA.res, DisName="AA",pathways)
#DCpathway.CKD.res<-DCpathway(DCEA.res=DCEA.CKD.res, DisName="CKD",pathways)
#DCpathway.T2D.res<-DCpathway(DCEA.res=DCEA.T2D.res, DisName="T2D",pathways)
##########################################################################

##########################################################################
## computate disease similarities
#DCpathway.disn = cbind(DCpathway.AA.res, DCpathway.CKD.res, DCpathway.T2D.res)
#DCEA.disn = list(dis1 = DCEA.AA.res, dis2 = DCEA.CKD.res, dis3 = DCEA.T2D.res)
#DS.res<-DS(DCpathway.disn, Ndis = 3, DCEA.disn,
# DisNames = c("AA", "CKD", "T2D"), pathways, cutoff = 0.05,
# Nper = 0,
# FigName = "DisNetwork.pdf", vsize = 5, lcex = 0.3, ewidth = 1.5)
#DS.res[1:3,]
##########################################################################

##########################################################################
## sort out common DCGs and DCLs in AA, CKD and T2D.
#data(tf2target)
#comDCGL.res<-comDCGL(Ndis = 3, DCEA.disn ,
# DisNames = c("AA", "CKD", "T2D"),
# cutoff = 0.25, tf2target)
```

```
#comDCGL.res$comDCGs[1:3,]
#comDCGL.res$comDCLs[1:3,]
##########################################################################

##########################################################################
## plot common DCGs and common DCLs with regulation information.
#comDCGLplot.res<-comDCGLplot(comDCGL.res,FigName="comDCGL.pdf",tf2target,
# vsize=5,asize=0.25,lcex=0.3,ewidth=1.5)
##########################################################################
```

---

comDCGL *Sort out common DCGs and DCLs in multi-diseases*

Description

common DCGs and DCLs maybe imply the similar regulation mechanisms which are the underlying reasons to disease relationships. Therefore comDCGL are implemented to sort out shared DCGs and DCLs in multi-disease.

Usage

```
comDCGL(Ndis = 3, DCEA.disn, DisNames = c("AA", "CKD", "T2D"), cutoff = 0.05, tf2target)
```

Arguments

Ndis	Disease number.
DCEA.disn	DCEA results for multi-diseases.
DisNames	Disease names.
cutoff	the cutoff of 'q.vaule' in DCEA results; must be within [0,1]; If there is no 'q.value' value (when N=0 in DCEA function), 'dC' will be sorted in decreasing order and retained the highest by 'cutoff' percent.
tf2target	a data frame or matrix for TF-to-target interaction pairs.

Value

A list with two components:

comDCGs	Displaying common DCGs with annotated regulator information.
comDCLs	Displaying common DCLs with annotated regulation relationship information.

Author(s)

Jing Yang

Examples

```
#####################################################################
## computate DCEA results for three disease, Allergic asthma (AA),
## Chronic kidney disease (CKD) and Type 2 Diabetes (T2D).
#####################################################################
data(exprs1)
data(exprs2)
data(exprs3)

#############################################################################
## the default value of cutoff in DCEA is 0.25,
## here cutoff is set to 1 for saving time when demonstrating the examples.
DCEA.AA.res<-DCEA(exprs1[1:200,1:5],exprs1[1:200,6:10],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
DCEA.CKD.res<-DCEA(exprs2[1:300,1:25],exprs2[1:300,26:31],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
DCEA.T2D.res<-DCEA(exprs3[1:200,1:12],exprs3[1:200,13:35],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
#############################################################################

#############################################################################
## sort out common DCGs and DCLs in AA, CKD and T2D.
data(tf2target)
DCEA.disn <- list(dis1 = DCEA.AA.res, dis2 = DCEA.CKD.res, dis3 = DCEA.T2D.res)
comDCGL.res<-comDCGL(Ndis = 3, DCEA.disn, DisNames = c("AA", "CKD", "T2D"),
cutoff = 0.25, tf2target)
comDCGL.res$comDCGs[1:3,]
comDCGL.res$comDCLs[1:3,]
#############################################################################
```

`comDCGLplot` *Visualization of common DCGs and DCLs*

Description

Graphical Representation of common DCLs with regulation information.

Usage

```
comDCGLplot(comDCGL.res, FigName = "comDCGL.pdf", tf2target,
vsize = 5, asize = 0.3, lcex = 0.3, ewidth = 1.5)
```

Arguments

`comDCGL.res`	Result of comDCGL function.
`FigName`	A character string of graph name.
`tf2target`	a data frame or matrix for TF-to-target interaction pairs.
`vsize`	A numeric of node size.

asize	A numeric of arrow size.
lcex	a numeric of lable size.
ewidth	a numeric of edge width.

Details

We built a function comDCGLplot to display combined information of common DCGs and DCLs with regulation information. In this graph, we display DCLs networks, and we rely on different node shapes differentiate TFs and non-TFs (square for TFs, circle for non-TFs), different node colors to categorize genes (pink for DCGs, blue for non-DCGs, and different edge types to express different relations of DCLs (solid for DCLs, edges with arrow indicate TF-to-target relations).

Value

One graph as users' wish have been saved in currently working directory

Author(s)

Jing Yang

Examples

```
#####################################################################
## computate DCEA results for three disease, Allergic asthma (AA),
## Chronic kidney disease (CKD) and Type 2 Diabetes (T2D).
#####################################################################
data(exprs1)
data(exprs2)
data(exprs3)

#############################################################################
## the default value of cutoff in DCEA is 0.25,
## here cutoff is set to 1 for saving time when demonstrating the examples.
DCEA.AA.res<-DCEA(exprs1[1:200,1:5],exprs1[1:200,6:10],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
DCEA.CKD.res<-DCEA(exprs2[1:300,1:25],exprs2[1:300,26:31],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
DCEA.T2D.res<-DCEA(exprs3[1:200,1:12],exprs3[1:200,13:35],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
#############################################################################

#####################################################
## sort out common DCGs and DCLs in AA, CKD and T2D.
data(tf2target)
comDCGL.res<-comDCGL(Ndis = 3,
DCEA.disn = list(dis1 = DCEA.AA.res, dis2 = DCEA.CKD.res, dis3 = DCEA.T2D.res),
DisNames = c("AA", "CKD", "T2D"),
cutoff = 0.25, tf2target)
#####################################################

#####################################################################
```

```
## plot common DCGs and common DCLs with regulation information.
#comDCGLplot.res<-comDCGLplot(comDCGL.res,FigName="comDCGL.pdf",tf2target,
# vsize=5,asize=0.25,lcex=0.3,ewidth=1.5)
################################################################
```

DCEA *Identify DCGs (Differentially Co-expressed Genes) and DCLs (Differentially Co-expressed Links) based on the 'Differential Co-Expression Analysis'*

Description

A method to pick out DCGs and DCLs from gene expression data based on the 'Differential Co-Expression Analysis' (DCEA). DCEA method was published in our previous work, it cantained DCp and DCe method. (Yu et al. 2010, Liu et al. 2011, Yang et al. 2013). DCGs are identified by DCp, DCLs are identified by DCe.

Usage

```
DCEA(exprs.1,exprs.2,
r.method=c('pearson','spearman')[1],
link.method=c('qth','rth','percent')[1],
cutoff=0.25,
N=0, N.type=c('pooled','gene_by_gene')[1],
q.method=c("BH","holm", "hochberg", "hommel", "bonferroni", "BY","fdr")[1],
nbins=20,p=0.1)
```

Arguments

exprs.1	a data frame or matrix for condition A, with rows as variables (genes) and columns as samples.
exprs.2	a data frame or matrix for condition B, with rows as variables (genes) and columns as samples.
r.method	a character string indicating which correlation coefficient (or covariance) is to be computed. One of "pearson" (default) or "spearman", can be abbreviated.
link.method	a character string indicating link filtering method, default is 'qth'.
cutoff	cutoff used for link filtration, can be rth, qth, or percent depending on link.method. must be within [0,1].
N	permutation times. If N>0, the permutation step will be implemented. The default value for N is 0.
N.type	a character string indicating permutation type, default is 'pooled'.
q.method	a character string indicating which correction method to be utilized. the default is 'BH'.
nbins	number of x bins for fitting y=a+(b/x).
p	the cutoff of q-value; must be within [0,1].

Details

DCp starts with a set of gene coexpression value pairs, where each pair is made up with two co-expression values of a gene pair calculated under two different conditions. For a particular gene, a 'condition-specific coexpression profile' is defined as the vector of the coexpression values that relate to it in one condition, and the two condition-specific coexpression profiles of one gene become the two components of the gene's 'differential coexpression profile'. A differential coexpression measure (dC) is calculated from the differential coexpression profile as a length-normalized Euclidean Distance. Then the samples between the two conditions will be disturbed and the samples will be separated to two conditions. Calculate dC of this condition. Repeat the above process for N times. Pool all the dC together to form a null distribution of dC. The corresponding statistical significance (p-value) is estimated against null statistics (short for pooled). Or calculate p-value of a gene only in this gene's null distribution of dC (short for gene_by_gene).

DCe is based on the 'Limit Fold Change' or 'LFC' model, a robust statistical method originally proposed for selecting DEGs from microarray data (Mutch et al. 2002). In LFC, we categorize correlation coefficients into bins according to their maximum coexpression values, and within each bin, select a fraction p of links with highest log fold changes, and fit a curve y = a + (b/x) over the boundary links. Links lying above the fitted curve are considered as DCLs. In most experiments of this work, we set p = 0.1.

Value

A table with DCGs will be listed including 'dC' value, profile 'links','p.value' and 'q.vaule' value for each gene. A table with DCLs will be listed including 'cor.1' and 'cor.2' (correlation coefficients in two conditions), 'type' and 'diff.cor'.

Author(s)

Jing Yang

References

H. Yu, B.H. Liu, Z.Q. Ye, C. Li, Y.X. Li, Y.Y. Li, Link-based Quantitative Methods to Identify Differentially Coexpressed Genes and Gene Pairs, BMC Bioinformatics 12 (2011) 315.

B.H. Liu, H. Yu, K. Tu, C. Li, Y.X. Li, Y.Y. Li, DCGL: an R package for identifying differentially coexpressed genes and links from gene expression microarray data, Bioinformatics 26 (2010) 2637-2638.

J. Yang, H. Yu, B.H. Liu, Z. Zhao, L. Liu, L.X. Ma, Y.X. Li, Y.Y. Li, DCGL v2.0: An R Package for Unveiling Differential Regulation from Differential Co-expression, PLoS One 8 (2013) e79729.

Mutch, D.M., Berger, A., Mansourian, R., Rytz, A. and Roberts, M.A, The limit fold change model: a practical approach for selecting differentially expressed genes from microarray data, BMC Bioinformatics, 3 17 (2002).

Examples

```
data(exprs1)
exprs.1<-exprs1[1:200,1:5]
exprs.2<-exprs1[1:200,6:10]
```

```
############################################################
## calculate differentially co-expressed genes (DCGs) and
## differentially co-expressed links (DCLs) by DCEA.
DCEA.res<-DCEA(exprs.1,exprs.2,
r.method=c('pearson','spearman')[1],
link.method=c('qth','rth','percent')[3],
cutoff=0.05,
N=0, N.type=c('pooled','gene_by_gene')[1],
q.method=c("BH","holm", "hochberg", "hommel", "bonferroni", "BY","fdr")[1],
nbins=20,p=0.1)
DCEA.res$genes[1:3,]
DCEA.res$links[1:3,]
############################################################

############################################################
## calculate differentially co-expressed genes (DCGs) and
## differentially co-expressed links (DCLs) with 100 permutation times.
#DCEA.res.N<-DCEA(exprs.1,exprs.2,
# r.method=c('pearson','spearman')[1],
# link.method=c('qth','rth','percent')[3],
# cutoff=0.05,
# N=100, N.type=c('pooled','gene_by_gene')[1],
# q.method=c("BH","holm", "hochberg", "hommel", "bonferroni", "BY","fdr")[1],
# nbins=20,p=0.1)
#DCEA.res.N$genes[1:3,]
#DCEA.res.N$links[1:3,]
############################################################
```

`DCpathway` *Calculate differential co-expression values (dCs) for each pathways*

Description

dCs of pathways are defined as average dC of their component genes.

Usage

```
DCpathway(DCEA.res = DCEA.res, DisName = "COPD", pathways)
```

Arguments

`DCEA.res`	DCEA results for multi-diseases.
`DisName`	Disease names.
`pathways`	A data frame or matix for pathways which generated from MSigDB.

Details

We computed the differential co-expression value (dC) of each gene between the disease and the control samples for all disease status via DCEA function. Then we assigned dCs of pathways to be the average dC of their component genes. Therefore a vector of pathways dC for each disease was obtained.

Value

A data frame of dCs for all pathways with

Author(s)

Jing Yang

Examples

```
data(exprs1)
data(exprs2)
data(exprs3)

#############################################################################
## the default value of cutoff in DCEA is 0.25,
## here cutoff is set to 1 for saving time when demonstrating the examples.
DCEA.AA.res<-DCEA(exprs1[1:200,1:5],exprs1[1:200,6:10],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
#############################################################################

#############################################################################
## computate dCs of pathways for disease
data(pathways)
DCpathway.AA.res<-DCpathway(DCEA.res=DCEA.AA.res, DisName="AA",pathways)
DCpathway.AA.res[1:3,]
#############################################################################
```

DS *Identification of disease similarity*

Description

Identify disease similarity through computating partial correlation coefficients of differenital co-expression values of pathways in different diseases.

Usage

```
DS(DCpathway.disn , Ndis = 3, DCEA.disn,
DisNames = c("AA", "IgA", "T2D"), pathways, cutoff = 0.05,
Nper = 0,
FigName = "DisNetwork.pdf", vsize = 5, lcex = 0.5, ewidth = 1.5)
```

Arguments

DCpathway.disn	Results of DCpathway function for each disease.
Ndis	Disease number.
DCEA.disn	DCEA results for multi-diseases.

DisNames	Disease names.
pathways	A data frame or matix for pathways which generated from MSigDB.
cutoff	the cutoff of 'p.vaule' for sorting out a list of significant disease pairs; must be within [0,1].
Nper	Permutation times
FigName	A character string of Figure name for plotting disease network.
vsize	a numeric of node size.
lcex	a numeric of lable size.
ewidth	a numeric of edge width.

Details

We computed the differential co-expression value (dC) of each gene between the disease and the control samples for all disease status via DCEA function. Then we assigned dCs of pathways to be the average dC of their component genes. So a vector of pathways dC for each disease was obtained. We emploied the R package, ppcor, to account the partial spearman correlation between two diseases finally.

In order to evaluate the statistical significance of observed disease partial correlation coefficients, we performed a permutation test, in which we randomly re-assigned the affiliation of gene to pathway with the number of pathways, the number of pathways component genes and the number of pathways a given gene belongs to unchanged, accounted the pathways dCs, and calculated the partial correlation coefficients using permuted data. This procedure can be set with 'Nper', when 'Nper' is larger than 0, a large number of partial correlation coefficient statistics form an empirical null distribution. The p-value for each disease pairs can then be estimated and FDR value can be obtained accordingly.

Value

A graph have been saved in currently working directory. And a data frame for each disease pairs with its partial correlation coefficient (pcor) and 'p.value' and 'q.value'.

Author(s)

Jing Yang

Examples

```
######################################################################
## computate DCEA results for three disease, Allergic asthma (AA),
## Chronic kidney disease (CKD) and Type 2 Diabetes (T2D).
######################################################################
data(exprs1)
data(exprs2)
data(exprs3)

#############################################################################
## the default value of cutoff in DCEA is 0.25,
## here cutoff is set to 1 for saving time when demonstrating the examples.
```

```
DCEA.AA.res<-DCEA(exprs1[1:200,1:5],exprs1[1:200,6:10],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
DCEA.CKD.res<-DCEA(exprs2[1:300,1:25],exprs2[1:300,26:31],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
DCEA.T2D.res<-DCEA(exprs3[1:200,1:12],exprs3[1:200,13:35],link.method="percent",
cutoff=1,N=0,nbins=20,p=0.1)
##############################################################################

##############################################################################
### computate dCs of pathways for each disease.
#data(pathways)
#DCpathway.AA.res<-DCpathway(DCEA.res=DCEA.AA.res, DisName="AA",pathways)
#DCpathway.CKD.res<-DCpathway(DCEA.res=DCEA.CKD.res, DisName="CKD",pathways)
#DCpathway.T2D.res<-DCpathway(DCEA.res=DCEA.T2D.res, DisName="T2D",pathways)
##############################################################################

##############################################################################
### computate disease similarities.
#DCpathway.disn = cbind(DCpathway.AA.res, DCpathway.CKD.res, DCpathway.T2D.res)
#DCEA.disn = list(dis1 = DCEA.AA.res, dis2 = DCEA.CKD.res, dis3 = DCEA.T2D.res)
#DS.res<-DS(DCpathway.disn , Ndis = 3, DCEA.disn ,
# DisNames = c("AA", "CKD", "T2D"), pathways, cutoff=0.05,
# Nper = 0,
# FigName = "DisNetwork.pdf", vsize = 5, lcex = 0.3, ewidth = 1)
#DS.res[1:3,]
##############################################################################
```

`expressionBasedfilter` *Filter genes according to expression level*

Description

Genes that have a higher between-sample mean expression signal are retained, while other genes are discarded.

Usage

```
expressionBasedfilter(exprs)
```

Arguments

`exprs`	a data frame or matrix with rows as variables (genes) and columns as samples.

Details

Genes which have a Between-Experiment Mean Expression Signal (BEMES) lower than the median of BEMES's of all genes will be filtered out.

Value

A data frame or matrix with a reduced number of rows.

Author(s)

Jing Yang

Examples

```
data(exprs1)
expressionBasedfilter(exprs1)
```

---

exprs1 *Real dataset pulled down from GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22528)*

Description

Gene expression dataset, containing 1000 rows and 10 columns.

Usage

```
data(exprs1)
```

Format

A data frame with 1000 observations 10 variables.

- exprs1 A data frame with 1000 observations 10 columns. The expression values.

Details

In the sample gene expression data matrix exprs, it was designed to study gene expression in Allergic Asthma samples and health samples.

Author(s)

Jing Yang

Examples

```
data(exprs1)
exprs1[,1:5] # exprssion data for Allergic Asthma samples
exprs1[,6:10] # exprssion data for health samples
row.names(exprs1) # gene identifiers
```

exprs2 *Real dataset pulled down from GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35487)*

Description

Gene expression dataset, containing 1000 rows and 31 columns.

Usage

```
data(exprs2)
```

Format

A data frame with 1000 observations 31 variables.

- exprs2 A data frame with 1000 observations 31 columns. The expression values.

Details

In the sample gene expression data matrix exprs, it was designed to study gene expression in Chronic kidney disease and health samples.

Author(s)

Jing Yang

Examples

```
data(exprs2)
exprs2[,1:25] # exprssion data for Chronic kidney disease samples
exprs2[,26:31] # exprssion data for health samples
row.names(exprs2) # gene identifiers
```

exprs3 *Real dataset pulled down from GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9006)*

Description

Gene expression dataset, containing 1000 rows and 35 columns.

Usage

```
data(exprs3)
```

Format

A data frame with 1000 observations 35 variables.

- exprs3 A data frame with 1000 observations 35 columns. The expression values.

Details

In the sample gene expression data matrix exprs, it was designed to study gene expression in Type 2 Diabetes and health samples.

Author(s)

Jing Yang

Examples

```
data(exprs3)
exprs3[,1:12] # exprssion data for Type 2 Diabetes samples
exprs3[,13:35] # exprssion data for health samples
row.names(exprs3) # gene identifiers
```

---

pathways *The dataset of human pathways downloaded from two collections of MSigDB*

Description

There are 572,639 pathways and their component genes downloaded from collection 2 and collection 5 of MSigDB (http://www.broadinstitute.org/gsea/msigdb/index.jsp).

Usage

```
data(pathways)
```

Format

A data frame with 572,639 observations 2 variables.

- pathways A data frame with 572,639 observations 2 columns. TF-to-target interaction pairs.

Details

We downloaded collection 2 and collection 5 files from MSigDB (http://www.broadinstitute.org/gsea/msigdb/index.jsp). There are 6,176 pathways including 21,075 genes.

Author(s)

Jing Yang

Examples

```
data(pathways)
pathways[1:3,] # pathways (column 1) and their component genes (column 2)
```

`tf2target` *The dataset of human Transcription Factors regulate potential target genes*

Description

There are 19,9950 TF-to-target interaction pairs downloaded from UCSC (http://genome.ucsc.edu/).

Usage

```
data(tf2target)
```

Format

A data frame with 19,9950 observations 2 variables.

- tf2target A data frame with 19,9950 observations 2 columns. TF-to-target interaction pairs.

Details

We downloaded tfbsConsSites and tfbsConsFactors files from UCSC (http://genome.ucsc.edu). tfbsConsSites gives the coordinate information of all TFs acted on, tfbsConsFactors gives TFs identifier information. We can predict all the TFs to potential targets relationship based on refGene file which also came from UCSC and contained genes' position information of hg18. There are 199 TFs and 19,9950 TF-to-target interactions.

Author(s)

Jing Yang, Hui Yu

Examples

```
data(tf2target)
tf2target[1:3,] # TFs (column 1) act on potential target genes (column 2)
```

varianceBasedfilter *To filter genes according to expression variability*

Description

Those genes not significantly more variable than the median gene are filtered out.

Usage

```
varianceBasedfilter(exprs,p)
```

Arguments

exprs	a data frame or matrix with rows for variables (genes) and columns for samples.
p	the probability cut-off of the chi-squared model of the gene-specific variance-like statistics.

Details

This is an approximate test of the hypothesis that gene has the same variance as the median variance. A statistical significance criterion based on the variance can be used. If the significance criterion is chosen, then the variance of the log-values for each gene is compared to the median of all the variances. The quantity for each gene compared to a percentile of the chi-square distribution with n-1 degrees of freedom. Those genes not significantly more variable than the median gene are filtered out [BRB-ArrayTools Version 3.7].

Value

A data frame or matrix with a reduced number of rows.

Author(s)

Jing Yang

References

Dr. Richard Simon & Amy Peng Lam, BRB-ArrayTools (v3.7) User's Manual: 'Log expression variation filter'.

Examples

```
data(exprs1)
varianceBasedfilter(exprs1,0.05)
```

Index

∗Topic **DCEA**
DCEA, 7
∗Topic **DCGL**
DCEA, 7
∗Topic **DCGs**
comDCGL, 4
comDCGLplot, 5
DCEA, 7
∗Topic **DCLs**
comDCGL, 4
comDCGLplot, 5
DCEA, 7
∗Topic **MSigDB**
pathways, 15
∗Topic **Target Genes**
tf2target, 16
∗Topic **Transcription Factors**
tf2target, 16
∗Topic **datasets**
exprs1, 13
exprs2, 14
exprs3, 14
∗Topic **differential co-expression**
DCpathway, 9
∗Topic **disease similarity**
DS, 10
∗Topic **gene filtering**
expressionBasedfilter, 12
varianceBasedfilter, 17
∗Topic **package**
DSviaDRM-package, 2
∗Topic **pathways**
pathways, 15
∗Topic **pathway**
DCpathway, 9
DS, 10
∗Topic **visualization**
comDCGLplot, 5

comDCGL, 4
comDCGLplot, 5

DCEA, 7
DCpathway, 9
DS, 10
DSviaDRM *(*DSviaDRM-package*)*, 2
DSviaDRM-package, 2

expressionBasedfilter, 12
exprs1, 13
exprs2, 14
exprs3, 14

pathways, 15

tf2target, 16

varianceBasedfilter, 17