Package 'DCGL'

February 19, 2015

Type Package

Title Differential Co-expression Analysis and Differential Regulation Analysis of Gene Expression Microarray Data

Version 2.1.2

Date 2014-12-18

Author Jing Yang, Hui Yu, Bao-Hong Liu, Zhongming Zhao, Lei Liu, Liang-Xiao Ma, Yi-Xue Li, Yuan-Yuan Li

Maintainer Bao-Hong Liu <bhliu@scbit.org>

Description Functions for 1) gene filtration; 2) link filtration; 3) differential co-expression analysis: DCG (Differential Coexpressed Gene) identification and DCL (Differentially Coexpressed Link) identification; and 4) differential regulation analysis: DRG (Differential Regulated Gene) identification, DRL (Differential Regulated Link) identification, DRL visualization and regulator ranking.

Depends R (>= 2.10)

Imports igraph, limma

License GPL (> 2)

LazyLoad yes

Repository CRAN

Reference Yang J, Yu H, Liu B-H, Zhao Z, Liu L, Ma L-X, Li Y-X and Li Y-Y. (2013) DCGL v2.0: An R Package for Unveiling Differential Regulation from Differential Co-expression. PLoS ONE 8(11): e79729. doi:10.1371/journal.pone.0079729

NeedsCompilation no

Date/Publication 2014-12-18 07:06:56

R topics documented:

DCGL-package 2
ASC 4
DCe 5

DCp . . . 7
DCsum . . . 8
DRplot . . . 10
DRrank . . . 12
DRsort . . . 13
expressionBasedfilter . . . 15
exprs . . . 16
exprs_design . . . 17
intgenelist . . . 17
LFC . . . 18
LRC . . . 19
percentLinkfilter . . . 20
qLinkfilter . . . 21
RIF . . . 22
rLinkfilter . . . 24
tf2target . . . 25
varianceBasedfilter . . . 25
WGCNA . . . 26

Index **28**

`DCGL-package` *Differential Co-expression Analysis and Differential Regulation Analysis of Gene Expression Microarray Data*

Description

DCGL package contains four modules which are Gene filtration module, Link filtration module, Differential Co-expression Analysis (DCEA) module and Differential Regulation Analysis (DRA) module. In Gene filtration module, there are expressionBasedfilter and varianceBasedfilter functions to filter genes on expression microarray data. In Link filtration module, there are rLinkfilter, percentLinkfilter and qLinkfilter functions to filter gene coexpression links in coexpression networks. DCp, DCe, WGCNA, LRC and ASC functions were implemented in DCEA module for extracting differentially coexpressed genes (DCGs) and differentially coexpressed links (DCLs). The final step of DCEA module is DCsum to determine DCGs and DCLs which come from DCe and DCp or only DCe method(s). In DRA module, there are DRsort, DRplot and DRrank functions to identify differentially regulated genes (DRGs) and differentially regulated links (DRLs), to visualize DRLs and DRL-related TF-to-target networks and to rank regulators in terms of their potential relevance to the biological phenotype, respectively.

Details

Package:	DCGL
Type:	Package
Version:	2.1.2
Date:	2014-12-18
License:	GPL (>2)
LazyLoad:	yes

Author(s)

Jing Yang, Hui Yu, Bao-Hong Liu, Zhongming Zhao, Lei Liu, Liang-Xiao Ma, Yi-Xue Li, Yuan-Yuan Li

Maintainer: Bao-Hong Liu <bhliu@scbit.org>

References

Yang J, Yu H, Liu B-H, Zhao Z, Liu L, Ma L-X, Li Y-X and Li Y-Y. (2013) DCGL v2.0: An R Package for Unveiling Differential Regulation from Differential Co-expression. PLoS ONE 8(11): e79729. doi:10.1371/journal.pone.0079729 Friedrich Leisch. (2008) Creating R Packages: A Tutorial

Examples

```
data(exprs)

## divide exprs into two parts corresponding to condition 1
##(exprs.1) and condition 2 (exprs.2) respectively
expGenes<-rownames(exprs)
exprs<-exprs[1:100,]
exprs.1<-exprs[1:100,1:16]
exprs.2<-exprs[1:100,17:63]

DCp.res<-DCp(exprs.1,exprs.2,
link.method='qth',cutoff=0.25,N=0)

DCe.res<-DCe(exprs.1,exprs.2,
link.method='qth',
cutoff=0.25,
nbins=10,p=0.1)


## combine two Differential Co-expression Analysis results
DCsum.res<-DCsum(DCp.res,DCe.res,
DCpcutoff=0.25,DCecutoff=0.4)
DCsum.res$DCGs[1:3,]
DCsum.res$DCLs[1:3,]


## sort out differentially regulated genes and differentially regulated links
data(tf2target) ## TF-to-target relationships
DRsort.res<-DRsort(DCsum.res$DCGs,DCsum.res$DCLs,tf2target,expGenes)
## or
DRsort.res<-DRsort(DCe.res$DCGs,DCe.res$DCLs,tf2target,expGenes)

## plot differentially regulated links
DRplot.res<-DRplot(DCsum.res$DCGs,DCsum.res$DCLs,
tf2target,
expGenes,
```

```
type='TF_bridged_DCL',
vsize=5,asize=0.25,lcex=0.3,ewidth=1,
figname=c('TF2target_DCL.pdf','TF_bridged_DCL.pdf'))


## rank regulators by TED or TDD
DRrank.res<-DRrank(DCsum.res$DCGs,DCsum.res$DCLs,
tf2target,
expGenes,
rank.method=c('TED','TDD')[1],
Nperm=0)


## rank regulators by RIF\
data(exprs_design)
RIF.res<-RIF(exprs,exprs.1,exprs.2,
tf2target,
exprs_design,
p.value=0.05)
```

ASC *Identify DCGs (Differential Coexpressed Genes) based on 'Average Specific Connection'*

Description

A method to pick out DCGs from microarray data based on 'Average Specific Connection' (ASC) (Choi et al. 2005).

Usage

```
ASC(exprs.1, exprs.2, link.method = c("qth", "rth", "percent")[1], cutoff)
```

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.
link.method	a character string indicating link filtration method, default is 'qth'.
cutoff	the cutoff of correlation-value, q-value or percent of links after link filtering. must be within [0,1].

Details

ASC is the average value of the specific degree of the two conditions.

Value

ASC	the Average Specific Connections of genes. This measure can be used to rank gene in terms of differential coexpression.

Author(s)

Bao-Hong Liu, Hui Yu

References

Choi, J.K., Yu, U., Yoo, O.J. and Kim, S. (2005) Differential coexpression analysis using microarray data and its application to human cancer, Bioinformatics, 21, 4348-4355.

Examples

```
data(exprs)
ASC(exprs[1:100,1:16],exprs[1:100,17:63],
link.method = 'qth', cutoff=0.25)
```

DCe *Identify DCGs (Differentially Coexpressed Genes) and DCLs (Differentially Coexpressed Links)*

Description

The algorithm first determines DCLs using a LFC model originally proposed for differential expression analysis, and then determines DCGs with their surrounding links enriched for DCLs.

Usage

```
DCe(exprs.1, exprs.2,
link.method=c('qth','rth','percent')[1],
cutoff,
r.method,
q.method,
nbins=20, p=0.1,
figname=c('LFC.s.jpeg','LFC.d.jpeg'))
```

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.
link.method	a character string indicating link filtration 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].

r.method	a character string indicating which correlation coefficient (or covariance) is to be computed. One of "pearson" (default) or "spearman", can be abbreviated.
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].
figname	names of figures of the LFC fitting results.

Details

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). With the analysis units changed from expression values to coexpression values, the LFC method with moderate adaption can be applied to screen for putative DCLs. DCGs with their surrounding links enriched for DCLs are determined through a binomial probability model.

Value

A list with two components:

DCGs	Differential Coexpression Genes
DCLs	Differential Coexpression Links

Author(s)

Bao-Hong Liu, Hui Yu, Jing Yang

References

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

Yu H., Liu B-H., et al., Link-specific Quantitative Methods to Identify Differentially Coexpressed Genes and Gene Pairs. Submitted. 2010

Examples

```
data(exprs)
## Identifying DCGs and DCLs by DCe method
DCe.res<-DCe(exprs[1:100,1:16],exprs[1:100,17:63],
link.method = 'qth',
cutoff=0.25,
nbins=20,p=0.1)
DCe.res$DCGs[1:3,]
DCe.res$DCLs[1:3,]
```

DCp *Identify DCGs (Differential Coexpressed Genes) based on the 'Differential Coexpression Profile'*

Description

A method to pick out DCGs from microarray data based on a novel concept of 'Differential Coexpression Profiles' (DCp) (Yu et al. 2010).

Usage

```
DCp(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])
```

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'.

Details

DCp starts with a set of gene coexpression value pairs, where each pair is made up with two coexpression 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).

Value

A table with DCGs will be listed including 'dC' value and profile 'links' or 'dC' value, profile 'links','p.value' and 'q.vaule' value.

Author(s)

Bao-Hong Liu, Hui Yu, Jing Yang

References

Yu H., Liu B-H., et al., Link-specific Quantitative Methods to Identify Differentially Coexpressed Genes and Gene Pairs. Submitted. 2010

Examples

```
data(exprs)

## calculate differential coexpressed genes by DCp without permutation
DCp(exprs[1:50,1:16],exprs[1:50,17:63],
link.method='qth',cutoff=0.25,
N=0)

## calculate differential coexpressed genes by DCp with 100 times permutation
DCp(exprs[1:50,1:16],exprs[1:50,17:63],
link.method='qth',cutoff=0.25,
N=100)
```

DCsum *Summarize DCGs and DCLs*

Description

A algorithm to select synthetical DCGs and DCLs from the results of DCp and DCe.

Usage

```
DCsum(DCp.res, DCe.res, DCpcutoff = 0.25, DCecutoff = 0.25)
```

Arguments

DCp.res	a data frame generated by DCp, with rows as genes and columns as 'dC' score, 'links', 'p.value' and 'q.value' value.
DCe.res	a list generated by DCe, with two components which are DCGs and DCLs.
DCpcutoff	the cutoff of 'q.vaule' in DCp results; must be within [0,1]; If there is no 'q.value' value (when N=0), 'dC' will be sorted in decreasing order and retained the highest by 'DCpcutoff' percent.
DCecutoff	the cutoff of 'q' in DCGs components of DCe results; must be within [0,1].

Details

DCsum, short for Differentially Coexpression Summarization, summarizes 1) a set of DCGs, which is an intersection of DCp and DCe results; and 2) a set of DCLs which by definition must be connected with the DCGs. As a result, DCsum combines results from different coexpression analysis methods.

Value

A list with two components:

DCGs	Differentially Coexpressed Genes, combined two differential coexpression analysis
DCLs	Differentially Coexpressed Links, combined two differential coexpression analysis

Author(s)

Jing Yang, Hui Yu

Examples

```
data(exprs)
DCp.res<-DCp(exprs[1:100,1:16],exprs[1:100,17:63],
link.method = 'qth',cutoff=0.25)
DCe.res<-DCe(exprs[1:100,1:16],exprs[1:100,17:63],
link.method = 'qth',cutoff=0.25,nbins=20,p=0.1)
DCsum.res<-DCsum(DCp.res,DCe.res,DCpcutoff=0.25,DCecutoff=0.25)

## Differentially Coexpressed Genes after combination
DCsum.res$DCGs[1:3,]

## Differentially Coexpressed Links after combination
DCsum.res$DCLs[1:3,]
```

DRplot *Visualization of DRLs*

Description

Graphical Representation of TF2target_DCL-centered and TF_bridged_DCL-centered networks

Usage

```
DRplot(DCGs, DCLs, tf2target, expGenes,
type = c("both", "TF2target_DCL", "TF_bridged_DCL")[1],
intgenelist = NULL,
vsize=5,asize=0.25,lcex=0.3,ewidth=1,
figname = c("TF2target_DCL.pdf", "TF_bridged_DCL.pdf"))
```

Arguments

DCGs	a data frame or matrix for DCGs list.
DCLs	a data frame or matrix for DCLs list.
tf2target	a data frame or matrix for TF-to-target interaction pairs.
expGenes	a list for measured genes by array
type	a character string to determine which type of DRLs ('TF2target_DCL' or 'TF_bridged_DCL' or 'both') will be plotted; default is 'both'.
intgenelist	a list of gene symbols, which contains only one column to display your interesting genes symbol; default is NULL
vsize	a numeric of node size
asize	a numeric of arrow size
lcex	a numeric of lable size
ewidth	a numeric of edge width
figname	two character strings of graph names.

Details

We built a function DRplot to display combined information of DCGs/DCLs and DRGs/DRLs. DRpolt generates two figures which are 1): TF2target_DCL-centered network and 2): TF_bridged_DCL-centered network. In both networks, 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, gray for TFs which are not tested in expression microarray data), and different edge types to express different relations of gene pairs (solid for DCLs, dashed for non-DCLs; edges with arrow indicate TF-to-target relations).

Value

One or Two graphs as users' wish have been saved in currently working directory. And a list of two components:

TF2target_DCL	One kind of DRLs termed TF2target_DCL.
TF_bridged_DCL	Another kind of DRLs termed TF_bridged_DCL.

Author(s)

Jing Yang, Hui Yu

Examples

```
data(exprs)
## divide exprs into two parts corresponding to condition 1
## (exprs.1) and condition 2 (exprs.2) respectively
exprs.1<-exprs[1:100,1:16]
exprs.2<-exprs[1:100,17:63]
expGenes<-rownames(exprs)

data(tf2target)
DCp.res<-DCp(exprs.1,exprs.2,link.method = 'qth',cutoff=0.25)
DCe.res<-DCe(exprs.1,exprs.2,link.method = 'qth',cutoff=0.25,nbins=10,p=0.1)
DCsum.res<-DCsum(DCp.res,DCe.res,DCpcutoff=0.25,DCecutoff=0.4)
DRplot.res<-DRplot(DCsum.res$DCGs,DCsum.res$DCLs,
tf2target,
expGenes,
type='TF_bridged_DCL',
intgenelist=NULL,
vsize=5,asize=0.25,lcex=0.3,ewidth=1,
figname=c('TF2target_DCL.pdf','TF_bridged_DCL.pdf'))

## two types of Differentially Regulated Links which were plotted
DRplot.res$TF2target_DCL[1:3,]
DRplot.res$TF_bridged_DCL[1:3,]

## plot sub-network by predefined gene(s)
data(intgenelist)
DRplot.int.res<-DRplot(DCsum.res$DCGs,DCsum.res$DCLs,
tf2target,
expGenes,
type='TF_bridged_DCL',
intgenelist=intgenelist,
vsize=5,asize=0.25,lcex=0.3,ewidth=1,
figname=c('TF2target_DCL_int.pdf','TF_bridged_DCL_int.pdf'))
DRplot.int.res$TF2target_DCL[1:3,]
DRplot.int.res$TF_bridged_DCL[1:3,]
```

DRrank *Ranking Regulators by Target Enrichment Density (TED) and Targets' DCL Density (TDD)*

Description

The algorithm to rank candidate regulators

Usage

```
DRrank( DCGs, DCLs,
tf2target, expGenes,
rank.method=c('TED','TDD')[1],
Nperm=0 )
```

Arguments

DCGs	a data frame or matrix for DCGs list.
DCLs	a data frame or matrix for DCLs list.
tf2target	a data frame or matrix for TF-to-target interaction pairs.
expGenes	a list for measured genes by array
rank.method	a character string indicating which ranking method to be utilized. The default is 'TED'.
Nperm	permutation times. If Nperm>0, the permutation step will be implemented for TED and TDD methods. The default value for Nperm is 0.

Details

DRrank is implemented for ranking potential TFs in terms of their relevance to the phenotypic change or biophysical process of interest. It contains two methods: TED, and TDD.

TED, short for 'Target Enrichment Density', employs Binomial Probability model to quantify the enrichment of a TF's targets in the DCG set, and as such to evaluate which regulators are more likely to be subject-relevant or even causal. Suppose we sift K DCGs from expression profile which contains N genes. If TFi has Ti targets in regulation knowledge, there should be Ti * K / N DCGs appeared in TFi targets list randomly. Actually, it is found that TI DCGs are included in TFi's targets list. The larger TI than Ti * K / N is, the more targets of TFi enriched, the more likely TFi is a relevant or causative regulator.

TDD, short for 'Targets' DCL Density', uses Clustering Coefficient to quantify the density of DCLs among a regulator's targets, and so to judge the importance of a TF. Suppose that TFi has n targets, and that there are k DCLs among these targets. A larger k means more DCLs are bridged by the common TFi. We intuitively assume that, if a TF bridged more TF_bridged_DCL it is of more importance (even if the regulator is not a DCG).

Value

A matrix to display TED or TDD scores and ranks.

Author(s)

Jing Yang, Hui Yu

Examples

```
data(exprs)
## divide exprs into two parts corresponding to condition 1
## (exprs.1) and condition 2 (exprs.2) respectively
expGenes<-rownames(exprs)
exprs<-exprs[1:100,]
exprs.1<-exprs[1:100,1:16]
exprs.2<-exprs[1:100,17:63]

data(tf2target)
DCp.res<-DCp(exprs.1,exprs.2,
link.method = 'qth',cutoff=0.25)
DCe.res<-DCe(exprs.1,exprs.2,
link.method = 'qth',cutoff=0.25,nbins=10,p=0.1)
DCsum.res<-DCsum(DCp.res,DCe.res,DCpcutoff=0.25,DCecutoff=0.4)

## rank all the potential TFs
data(tf2target)
DRrank.TED.res<-DRrank(DCsum.res$DCGs, DCsum.res$DCLs,
tf2target, expGenes,
rank.method=c('TED','TDD')[1],
Nperm=0)

DRrank.TED.res[1:3,]

DRrank.TDD.res<-DRrank(DCsum.res$DCGs, DCsum.res$DCLs,
tf2target, expGenes,
rank.method=c('TED','TDD')[2],
Nperm=0)

DRrank.TDD.res[1:3,]
```

`DRsort` *Identify DRGs (Differential regulated Genes) and DRLs (Differential regulated Links)*

Description

The algorithm is to determine DRGs and DRLs from DCGs and DCLs by TF-to-target interaction knowledge.

Usage

```
DRsort(DCGs, DCLs, tf2target, expGenes)
```

Arguments

DCGs	a data frame or matrix for DCGs list.
DCLs	a data frame or matrix for DCLs list.
tf2target	a data frame or matrix for TF-to-target interaction pairs.
expGenes	a list for measured genes by array

Details

DRsort, is aimed to sift DCGs and DCLs according to regulation knowledge.

If a DCG is a TF, it is intuitively speculated that its related differential coexpression may be attributed to the change of its regulation relationships with its targets. So this type of DCGs are termed Differential Regulation Genes (DRGs). Besides if the upstream TFs of a DCG is identified, that DCG is possibly a differentially regulated target of an implicated regulator, and so such DCGs are also kept in the set of DRGs.

If a DCL happens to be a TF-to-target relation, we highlight this DCL because it is the direct attribution to differential regulation. This type of DCLs are termed TF2target_DCL. On the other hand, if there are one or more common TFs regulating the two genes of a DCL, we also give priority to this DCL because the change in the expression correlation of the two genes could be attributed to the disruption of their co-regulation by the common TFs. This type of DCLs are termed TF_bridged_DCLs. TF2target_DCL and TF_bridged_DCL, therefore, together form the set of Differentially Regulated Links (DRLs).

Value

A list with four components:

DCGs	Displaying all of Differentially Coexpressed Genes with annotated regulator information whether it is available.
DCLs	Displaying all of Differentially Coexpressed Links with annotated regulator information whether it is available.
DRGs	Differentially Regulated Genes by annotating regulator information.
DRLs	Differentially Regulated Links by annotating regulation pairs information
DCG2TF	Displaying DCGs and upstream TF of DCGs by pairs. It is another format of DRGs.
TF_bridged_DCL	Displaying another format of TF_bridged_DCL for the ease of following investigation.

Author(s)

Jing Yang, Hui Yu

Examples

```
data(exprs)
data(tf2target)
expGenes<-rownames(exprs[1:100,])

## Two differential co-expression analysis methods
DCp.res<-DCp(exprs[1:100,1:16],exprs[1:100,17:63],
link.method = 'qth',cutoff=0.25)
DCe.res<-DCe(exprs[1:100,1:16],exprs[1:100,17:63],
link.method = 'qth',cutoff=0.25,nbins=20,p=0.1)

## Summarizing DCGs and DCLs from DCp and DCe derived results
DCsum.res<-DCsum(DCp.res,DCe.res,DCpcutoff=0.25,DCecutoff=0.25)

## Sorting out DRGs and DRLs from DCsum-outputted results
DRsort.res<-DRsort(DCsum.res$DCGs,DCsum.res$DCLs,tf2target,expGenes)
## or only sorting out DRGs and DRLs from DCe-outputted results
DRsort.res<-DRsort(DCe.res$DCGs,DCe.res$DCLs,tf2target,expGenes)


## DRGs list
DRsort.res$DRGs[1:3,]

## DRLs list
DRsort.res$DRLs[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)

Bao-hong Liu, Hui Yu, Jing Yang

Examples

```
data(exprs)
expressionBasedfilter(exprs)
```

---

exprs *Real dataset pulled down from GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17967)*

Description

Gene expression dataset, containing 1000 rows and 63 columns.

Usage

```
data(exprs)
```

Format

A data frame with 1000 observations 63 variables.

- exprs A data frame with 1000 observations 63 columns. The expression values.

Details

In the sample gene expression data matrix exprs, it was designed to study gene expression in cirrhotic tissues with (N=16) and without (N=47) HCC.

Author(s)

Jing Yang, Hui Yu

Examples

```
data(exprs)
exprs[,1:16] # exprssion data for condition A
exprs[,17:63] # exprssion data for condition B
row.names(exprs) # gene identifier
```

exprs_design *Experiment design of microarray matrix data*

Description

A matrix display microarray experiment, which contains rows corresponding to arrays and columns to coefficients to be estimated.

Usage

```
data(exprs_design)
```

Format

A data frame with array samples.

Details

Using numbers to describe microarray experiment design. Usually defined normal samples to be 0, unnormal samples to be 1.

Author(s)

Jing Yang, Hui Yu

Examples

```
data(exprs_design)
exprs_design #define 1 to 16 rows as normal condition, 17 to 63 rows as unnormal condition.
```

intgenelist *Interesting genes list*

Description

Given a list of interesting genes

Usage

```
data(intgenelist)
```

Format

A data frame or matrix with 1 column and several rows of gene symbols.

- intgenelist A data frame with 1 column and several rows. Interesting Gene Symbols.

Details

In the intgenelist, there are 5 Gene symbols in 1 column with the column name 'GeneSymbol' needed.

Author(s)

Jing Yang, Hui Yu

Examples

```
data(intgenelist)
intgenelist # Gene symbol list
colnames(intgenelist) # needed 'GeneSymbol' as the column name
```

LFC *Select DCLs based on 'Limit Fold Change' model*

Description

The limit fold change (LFC) model is a robust statistical method modeling the relationship between maximum coexpression and log coexpression ratio of genes. (Mutch, et al., 2002). The algorithm starts with a set of gene coexpression value pairs each comprising two coexpression values of a gene pair calculated under two different conditions respectively.

Usage

```
LFC(exprs, nbins = 20, p = 0.1, sign, figname = "LFC.jpeg")
```

Arguments

exprs	a two-column data matrix, with column one the coexpression values for one condition and column two those for another.
nbins	number of x bins for fitting y=a+(b/x).
p	the fraction at y axis for determining boundary points; must be within [0,1].
sign	specifies the sign type of exprs. Exprs is either 'same-sign', with coexpression pairs of the same sign, or 'different-sign', with coexpression pairs of opposite signs.
figname	names of figures of the LFC fitting results.

Details

According to how the signs of coexpression values are paired, gene links are divided into two parts: the 'same-signed' set and the 'differently-signed' set. From the 'differently-signed' set, the 'correlation-switched' gene links that have two differently-signed coexpression values both surpassing a cutoff value are subtracted, who make the first part of DCLs. The remaining differently-signed gene links in aggregate inherit the title of 'differently-signed' set. For the 'same-signed' set, gene links are binned with respect to their maximum coexpression values, and those links ranked the top p of highest fold changes in each bin are fitted with a simple equation of the form y = a + (b/x); for the 'differently-signed set', the horizontal and vertical axes are exchanged and similar binning and fitting procedures are applied. Links lie above the fitted curves are considered DCLs.

Value

the identified DCLs will be returned, as a subset of coexpression pairs.

Author(s)

Bao-Hong Liu, Hui Yu

References

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

LRC *Identify DCGs (Differential Coexpressed genes) based on 'Log Ratio Connections'*

Description

A method to pick out DCGs from microarray data based on 'Log Ratio of Connections' (LRC) (Reverter et al. 2006).

Usage

```
LRC(exprs.1, exprs.2, link.method = c("qth", "rth", "percent")[1], cutoff)
```

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.
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].

Details

'Log Ratio of Connections' (LRC) calculates the logarithm of the ratio of the connectivities of a gene between two conditions (Reverter, et al., 2006). A connectivity of zero is changed to one.

Value

LRC	the log Ratio Connections of genes. This measure can be used to rank gene in terms of differential coexpression.

Author(s)

Bao-Hong Liu, Hui Yu, Jing Yang

References

Reverter, A., Ingham, A., Lehnert, S.A., Tan, S.H., Wang, Y., Ratnakumar, A. and Dalrymple, B.P. (2006) Simultaneous identification of differential gene expression and connectivity in inflammation, adipogenesis and cancer, Bioinformatics, 22, 2396-2404.

Examples

```
data(exprs)
LRC(exprs[1:100,1:16],exprs[1:100,17:63],link.method = 'qth', cutoff=0.25)
```

---

percentLinkfilter *Filter gene coexpression links according to the top percent of expression correlation coefficient value*

Description

Keep a fraction (specified with 'cutoff') of the links (gene pairs) with the highest max correlation values.

Usage

```
percentLinkfilter(exprs.1, exprs.2, cutoff = 0.25,
r.method = c("pearson", "spearman")[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.
cutoff	fraction of links to be retained. Default is 0.25.
r.method	a character string to indicating 'pearson' (default) or 'spearman' correlation coefficient will be computed.

Details

Each gene link is associated with two correlation values (one out of condition A and the other out of condition B) and thus a list of 'maximum absolute values' of the two values is decided. Then these 'maximum absolute values' are sorted in decreasing order. At last, a fraction of gene pairs with the highest max correlation values will be retained.

Value

A list with two components of data frames, one for filtered data of condition A, the other for the counterpart of condition B.

Author(s)

Bao-Hong Liu, Hui Yu, Jing Yang

Examples

```
data(exprs)
percentLinkfilter(exprs[1:100,1:16],exprs[1:100,17:63],cutoff=0.25,r.method='pearson')
```

---

qLinkfilter *Filter gene coexpression links according to the q-values of expression correlation values*

Description

Gene links with q-values of coexpression value pairs in either of two conditions higher than the cutoff are retained, while the coexpression values of other links are set to zero.

Usage

```
qLinkfilter(exprs.1, exprs.2,
cutoff = 0.25,
r.method = c("pearson", "spearman")[1],
q.method = c("BH", "holm", "hochberg", "hommel", "bonferroni", "BY", "fdr")[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.
cutoff	the cutoff of q-value; must be within [0,1].
r.method	a character string indicating 'pearson' (default) or 'spearman' correlation coefficient will be computed.
q.method	a character string indicating adjust method of p-value, default is 'BH'.

Details

For each of the two conditions, the coexpression values are associated with the corresponding p-values (student T-test of the zero nature of a PCC), and these p-values are sorted and transformed to q-values (false discovery rates). Gene links with q-values of coexpression values in either of two conditions lower than the cutoff are reserved.

Value

A list with two components of data frames, one for filtered data of condition A, the other for the counterpart of condition B.

Author(s)

Bao-hong Liu, Hui Yu, Jing Yang

Examples

```
data(exprs)
qLinkfilter(exprs[1:100,1:16],exprs[1:100,17:63],
cutoff=0.25,
r.method='pearson',
q.method='BH')
```

RIF *Ranking Regulators by Regulator Impact Factor (RIF) Method*

Description

The algorithm to rank candidate regulators

Usage

```
RIF(exprs, exprs.1, exprs.2,
tf2target,
exprs_design,
p.value)
```

Arguments

exprs	a data frame or matrix for expression dataset, with rows as variables (genes) and columns as samples.
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.
tf2target	a data frame or matrix for regulator-to-target interaction pairs.
exprs_design	a data frame or matrix for displaying microarray experiment design.
p.value	a p value threshold to determine differential expression genes (DEGs).

Details

RIF method, short for 'Regulator Impact Factor' (Reverter etal. 2010), assesses the change of regulation-accountable expression value of Differentially Expressed Genes (DEGs) and correlation coefficient between DEGs and TFs to rank TFs.

Value

A matrix to display RIF scores and ranks.

Author(s)

Jing Yang, Hui Yu

References

Reverter, A., Hudson, N. J., Nagaraj, S. H., Perez-Enciso, M., Dalrymple, B. P. (2010) Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data, 26, 896-904.

Examples

```
data(exprs)
## divide exprs into two parts corresponding to condition 1
## (exprs.1) and condition 2 (exprs.2) respectively
exprs<-exprs[1:100,]
exprs.1<-exprs[1:100,1:16]
exprs.2<-exprs[1:100,17:63]

DCp.res<-DCp(exprs.1,exprs.2,
link.method = 'qth',cutoff=0.25)
DCe.res<-DCe(exprs.1,exprs.2,
link.method = 'qth',cutoff=0.25,nbins=10,p=0.1)
DCsum.res<-DCsum(DCp.res,DCe.res,DCpcutoff=0.25,DCecutoff=0.4)

## rank all the potential TFs
data(tf2target)
data(exprs_design)
RIF.res<-RIF(exprs,exprs.1,exprs.2,
tf2target,
exprs_design,
p.value=0.05)

RIF.res[1:3,]
```

rLinkfilter *Filter gene coexpression links according to correlation coefficient value*

Description

Keep a fraction of the links (gene pairs) with the higher correlation values than user given threshold (specified with 'cutoff').

Usage

```
rLinkfilter(exprs.1, exprs.2, cutoff = 0.8,
r.method = c("pearson", "spearman")[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.
cutoff	the threshold of correlation coefficient value. Default is 0.8.
r.method	a character string to indicate which type of correlation coefficient, 'pearson' (default) or 'spearman', will be computed.

Details

Each gene link is associated with two correlation values (one out of condition A and the other out of condition B) and thus a list of 'maximum absolute values' of the two values is decided. Then these links have larger correlation coefficient than threshold are kept.

Author(s)

Bao-Hong Liu, Hui Yu, Jing Yang

Examples

```
data(exprs)
rLinkfilter(exprs[1:100,1:16],exprs[1:100,17:63],cutoff=0.8,r.method='pearson')
```

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)

Bao-hong Liu, Hui Yu, Jing Yang

References

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

Examples

```
data(exprs)
varianceBasedfilter(exprs,0.05)
```

WGCNA *Identify DCGs (Differential Coexpressed Genes) based on the 'Weighted Gene Coexpression Network Analysis'*

Description

A method to pick out DCGs from microarray data based on 'Weighted Gene Coexpression Network Analysis' (WGCNA) (Mason, MJ. Et al. 2009; van Nas et al. 2009).

Usage

```
WGCNA(exprs.1, exprs.2, power = 12, variant = "WGCNA")
```

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.
power	the thresholding parameter, an integer >1.
variant	if the variant is 'WGCNA' the original version is evoked; if it is 'DCp', the length-normalized Euclidean distance is adopted to replace the connectivity difference measure.

Details

The 'weighted gene coexpression network analysis' (WGCNA) weights links with correlation coefficients and compares the sums of the correlation coefficients of a gene (Mason, et al., 2009; van Nas, et al., 2009). Correlation coefficients are firstly softly thresholded by a 'power'.

Value

WGCNA	score of 'WGCNA' to identify DCGs

Author(s)

Bao-Hong Liu, Hui Yu

References

Mason, M.J., et al. (2009) Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells, BMC Genomics, 10, 327.

van Nas, A., Guhathakurta, D., Wang, S.S., Yehya, N., Horvath, S., Zhang, B., Ingram-Drake, L., Chaudhuri, G., Schadt, E.E., Drake, T.A., Arnold, A.P. and Lusis, A.J. (2009) Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks, Endocrinology, 150, 1235-1249.

Examples

```
data(exprs)
WGCNA(exprs[1:100,1:16],exprs[1:100,17:63],power=12,variant='WGCNA')
```

Index

∗Topic **DCGs**
ASC, 4
DCe, 5
DCp, 7
LRC, 19
WGCNA, 26
∗Topic **DCLs**
DCe, 5
DCp, 7
LFC, 18
∗Topic **DRGs**
DRsort, 13
∗Topic **DRLs**
DRsort, 13
∗Topic **DRL**
DRplot, 10
∗Topic **Interesting Gene Symbol**
intgenelist, 17
∗Topic **Regulator Rank**
DRrank, 12
RIF, 22
∗Topic **Summary of DCGs**
DCsum, 8
∗Topic **Summary of DCLs**
DCsum, 8
∗Topic **Target Genes**
tf2target, 25
∗Topic **Transcription Factors**
tf2target, 25
∗Topic **datasets**
exprs, 16
∗Topic **design**
exprs_design, 17
∗Topic **gene filtering**
expressionBasedfilter, 15
varianceBasedfilter, 25
∗Topic **link filtering**
percentLinkfilter, 20
qLinkfilter, 21
rLinkfilter, 24
∗Topic **package**
DCGL-package, 2
∗Topic **visualization**
DRplot, 10

ASC, 4

DCe, 5
DCGL *(*DCGL-package*)*, 2
DCGL-package, 2
DCp, 7
DCsum, 8
DRplot, 10
DRrank, 12
DRsort, 13

expressionBasedfilter, 15
exprs, 16
exprs_design, 17

intgenelist, 17

LFC, 18
LRC, 19

percentLinkfilter, 20

qLinkfilter, 21

RIF, 22
rLinkfilter, 24

tf2target, 25

varianceBasedfilter, 25

WGCNA, 26