Package 'DEDS'

September 24, 2012

Title Differential Expression via Distance Summary for Microarray Data

Date 2007-09-24

Version 1.30.0

Author Yuanyuan Xiao <yxiao@itsa.ucsf.edu>, Jean Yee Hwa Yang <jeany@maths.usyd.edu.au>.

Description This library contains functions that calculate various statistics of differential expression for microarray data,including t statistics, fold change, F statistics, SAM,moderated t and F statistics and B statistics. It also implements a new methodology called DEDS (Differential Expression via Distance Summary), which selects differentially expressed genes by integrating and summarizing a set of statistics using a weighted distance approach.

Maintainer Yuanyuan Xiao <yxiao@itsa.ucsf.edu>

Depends R (>= 1.7.0)

License LGPL

biocViews Bioinformatics, Microarray, DifferentialExpression

R topics documented:

`affySpikeIn` *Gene expression dataset from Affymetrix Spike-in Experiments*

Description

The spike-in experiment represents a portion of the data used by Affymetrix to develop their MAS 5.0 preprocessing algorithm. Here we utilize the RMA (Irizarry et. al) probe level summaries. The data features 14 human genes spiked-in at a series of 14 known concentrations ($0, 2^{-2}, 2^{-1}, \ldots, 2^{10}$ pM) according to a Latin square design among 12612 null genes. The data matrix `affySpikeIn` represents the two array groups among the 14 array groups that contain 12 replicates. Further details are available at http://www.affymetrix.com/analysis/download_center2.affx.

Usage

```
data(affySpikeIn)
```

Value

`affySpikeIn`	matrix of gene expression levels measurements, rows correspond to genes (12626 genes) and columns to 24 samples.
`affySpikeIn.L`	numeric vector indicating the sample class, 12 (code 0) vs. 12 (code 1).
`affySpikeIn.gnames`	character vector containing the names of the 12626 genes.
`spikedgene`	numeric vector given the location of the 14 spiked genes.

References

R. A. Irizarry, B. M. Bolstad, F. Collin, L. Cope, B. Hobbs and T. P. Speed (2003) Summeries of affymetrix genechip probe level data. *Nucleic Acide Research*, 31:e15.

aggregateFun *Aggregate Statistical Functions for DEDS*

Description

This function takes statistic functions and creates a function that takes a matrix as a single argument. The statistic functions are bound in the environment of the returned function and are applied sequentially to the argument of the returned function.

Usage

```
aggregateFun(...)
```

Arguments

...	Functions of various statistics, could be in a list.

Details

The function takes several statistics functions or a list of these functions and returns a function (F) with bindings to the input statistics functions. F takes a data matrix as its single argument, and apply the bound statistical functions sequentially to the data matrix.

Value

It returns a function that takes a matrix as its single argument. The function returns a matrix of statistics, with m rows corresponding to variables (hypotheses) and n columns corresponding to specified statistics.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

comp.t, comp.FC, comp.F, comp.SAM, comp.modt, comp.modF, comp.B

Examples

```
X <- matrix(rnorm(100, 0, 1), nc=10)
L <- rep(0:1,c(5,5))
t.fun <- comp.t(L)
fc.fun <- comp.FC(L)
sam.fun <- comp.SAM(L)
ffun <- aggregateFun(list(t.fun, fc.fun, sam.fun))
stats <- ffun(X)
```

ApoA1 *Gene expression dataset from the ApoA1 Experiment*

Description

Gene expression data (6384 genes and 16 samples) from a study of a mouse model with very low HDL cholesterol levels described in Dudoit et al. (2002). Pre-processing was done as described in Dudoit et al. (2002).

Usage

```
data(ApoA1)
```

Value

ApoA1	matrix of gene expression levels measurements, rows correspond to genes (6384 genes) and columns to 16 samples.
ApoA1.L	numeric vector indicating the sample class, 8 (code 0) vs. 8 (code 1).

References

S. Dudoit, Y. H. Yang, T. P. Speed and M. J. Callow (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, Vol. 12, No. 1, pp. 111-139.

comp.adjp *Computing permutation based step-down maxT adjusted p values for each row of a matrix*

Description

This function computes permutation based step-down maxT adjusted p values for a selected test statistic, e.g., one- or two-sample t-statistics, F-statistics, SAM, Fold change, moderated t-statistics and moderated F-statistics, for each row of a matrix. The procedure is based on codes from `mt.maxT` and described in Westfall & Young (1993).

Usage

```
comp.adjp(X, L, B = 1000, test = c("t", "fc", "sam", "f", "modt", "modf"), tail = c("abs", "lowe
```

Arguments

X	A matrix, with m rows corresponding to variables (hypotheses) and$n$ columns corresponding to observations. In the case of gene expression data, rows correspond to genes and columns to mRNA samples. The data can be read using `read.table`.
L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k - 1$.

B	The number of permutations. For a complete enumeration, B should be 0 (zero) or any number not less than the total number of permutations.
test	A character string specifying the statistic to be used to test the null hypothesis of no association between the variables and the class labels. If test="t", for one-class, the tests are based on one-sample t-statistics; for two-class, the tests are based on two-sample t-statistics (unequal variances). If test="f", the tests are based on F-statistics. If test="fc", the tests are based on fold changes among classes. If test="sam", the tests are based on SAM-statistics. If test="modt", the tests are based on moderated t-statistics. If test="modf", the tests are based on moderated F-statistics.
tail	A character string specifying the type of rejection region. If side="abs", two-tailed tests, the null hypothesis is rejected for large absolute values of the test statistic. If side="higher", one-tailed tests, the null hypothesis is rejected for large values of the test statistic. If side="lower", one-tailed tests, the null hypothesis is rejected for small values of the test statistic.
extra	Extra parameter need for the test specified; see deds.genExtra.

Details

see mt.maxT.

Value

A matrix of the following columns:

order	order of rows (genes) based on statistics.
stat	a vector of statistics.
unadj.p	a vector of unadjusted p values.
adj.p	a vector of adjusted p values.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

comp.unadjp, comp.fdr, comp.stat

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# t statistics
unadjp.t <- comp.adjp(X, L, test="t")
```

comp.B *Computing B-statistics for Differential Expression*

Description

comp.B returns a function of one argument with bindings for L and proportion. This function accepts a microarray data matrix as its single argument, when evaluated, computes lod-odds of differential expression by empirical Bayes shrinkage of the standard error toward a common value. The lod-odds are sometimes called B statistics.

Usage

```
comp.B(L = NULL, proportion = 0.01)
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
proportion	A numeric variable specifying the proportion of differential expression.

Details

The function returned by comp.B calculates B statistics for each row of the microarray data matrix, with bindings for L and proportion. It interfaces to a C function. comp.stat is another function that wraps around the same C function that could be used for computing B statistics (see examples below).

Value

comp.B returns a function (F) with the bindings for L and proportion . The function F when supplied with a microarray data matrix and evaluated will return a numeric vector of B statistics for each row of the matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jeany@maths.usyd.edu.au>.

References

Lonnstedt, I. and Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica* 12, 31-46.

Smyth, G. K. (2003). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. http://www.statsci.org/smyth/pubs/ebayes.pdf

See Also

comp.modt,comp.stat.

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# compute B statistics, proportion set as 0.01
B.fun <- comp.B(L)
B.X <- B.fun(X)

# compute B statistics, proportion set as 0.1
B.fun <- comp.B(L, proportion=0.1)
B.X <- B.fun(X)

# Another way of computing B statistics
B.X<- comp.stat(X, L, "B")
```

---

comp.ebayes *Computing Empirical Bayes Statistics for Differential Expression*

Description

comp.ebayes returns a function of one argument with bindings for L and proportion. This function accepts a microarray data matrix as its single argument, when evaluated, computes lod-odds (B statistics) and moderated t statistics of differential expression by empirical Bayes shrinkage of the standard error toward a common value.

Usage

```
comp.ebayes(L = NULL, proportion = 0.01)
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
proportion	A numeric variable specifying the proportion of differential expression.

Details

The function returned by comp.ebayes calculates B statistics and moderated t statistics for each row of the microarray data matrix, with bindings for L and proportion. It interfaces to a C function.

Value

comp.ebayes returns a function (F) with the bindings for L and proportion . The function F when supplied with a microarray data matrix and evaluated will return a matrix of two columns:

t	Moderated t statistics
B	B statistics (log-odds) of differential expression

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Lonnstedt, I. and Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica* **12**, 31-46.

Smyth, G. K. (2003). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. http://www.statsci.org/smyth/pubs/ebayes.pdf

See Also

comp.modt,comp.B.

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# compute B and moderated t statistics, proportion set as 0.01
ebayes.fun <- comp.ebayes(L)
ebayes.X <- ebayes.fun(X)

# compute B and moderated t statistics, proportion set as 0.1
ebayes.fun <- comp.ebayes(L, proportion=0.1)
ebayes.X <- ebayes.fun(X)
```

---

comp.F *Computing F-statistic for Differential Expression*

Description

comp.F returns a function of one argument with bindings for L. This function accepts a microarray data matrix as its single argument, when evaluated, computes F statistics for each row of the matrix.

Usage

```
comp.F(L = NULL)
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.

Value

comp.F returns a function with bindings for L, which calculates and returns of vector of F statistics for each row in the data matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jeany@maths.usyd.edu.au>.

See Also

comp.FC, comp.t

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# two sample test, unequal variance
F <- comp.F(L)
f.X <- F(X)
```

---

comp.FC *Computing Fold Change for Differential Expression*

Description

comp.FC returns a function of one argument with bindings for L, is.log and FUN. This function accepts a microarray data matrix as its single argument, when evaluated, computes fold change for each row of the matrix.

Usage

```
comp.FC(L = NULL, is.log = TRUE, FUN = mean)
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
is.log	A logical variable indicating whether the data has been logged.
FUN	The summary statistics function used to calculate fold change, the default is set as mean, the user can also use median.

Details

The function returned by comp.FC calculates fold change for each row of the matrix, given specific class labels. If is.log=TRUE, fold change is calculated by subtraction; if is.log=FALSE, fold change is calculated by division.

Value

comp.FC returns a function with bindings for L, is.log and FUN, which calculates and returns a vector of fold changes for each row in the data matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

comp.t,comp.F

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

f <- comp.FC(L=L)
f.X <- f(X)
```

---

comp.fdr *Computing permutation based q values controlling false discovery rate for each row of a matrix*

Description

This function computes permutation based q values for a selected test statistic, e.g., one- or two-sample t-statistics, F-statistics, SAM, Fold change, moderated t-statistics and moderated F-statistics, for each row of a matrix.

Usage

```
comp.fdr(X, L, B = 1000, test = c("t", "fc", "sam", "f", "modt", "modf"), tail = c("abs", "lower
```

Arguments

X	A matrix, with m rows corresponding to variables (hypotheses) and$n$ columns corresponding to observations. In the case of gene expression data, rows correspond to genes and columns to mRNA samples. The data can be read using read.table.
L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
B	The number of permutations. For a complete enumeration, B should be 0 (zero) or any number not less than the total number of permutations.
test	A character string specifying the statistic to be used to test the null hypothesis of no association between the variables and the class labels. If test="t", for one-class, the tests are based on one-sample t-statistics; for two-class, the tests are based on two-sample t-statistics (unequal variances). If test="f", the tests are based on F-statistics. If test="fc", the tests are based on fold changes among classes. If test="sam", the tests are based on SAM-statistics. If test="modt", the tests are based on moderated t-statistics. If test="modf", the tests are based on moderated F-statistics.

tail	A character string specifying the type of rejection region. If side="abs", two-tailed tests, the null hypothesis is rejected for large absolute values of the test statistic. If side="higher", one-tailed tests, the null hypothesis is rejected for large values of the test statistic. If side="lower", one-tailed tests, the null hypothesis is rejected for small values of the test statistic.
extra	Extra parameter need for the test specified; see deds.genExtra.

Value

A matrix of the following columns:

order	order of rows (genes) based on statistics.
stat	a vector of statistics.
unadj.p	a vector of unadjusted p values.
qvalues	a vector of q values.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Storey, J.D. (2003) The positive False Discovery Rate: A Bayesian Interpretation and the q-value. *Annals of Statistics*, 31:2013-2035.

See Also

comp.unadjp, comp.adjp, comp.stat

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# t statistics
unadjp.t <- comp.fdr(X, L, test="t")
```

comp.modF *Computing Moderated t-statistics for Differential Expression*

Description

comp.modF returns a function of one argument with bindings for L. The function accepts a microarray data matrix as its single argument, when evaluated, computes moderated F-statistics by empirical Bayes shrinkage of the standard error toward a common value.

Usage

```
comp.modF(L = NULL)
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k - 1$.

Details

The function returned by comp.modF computes moderated F statistics for the assessment of differential expression. It interfaces to a C function. comp.stat is another function that wraps around the C function that could be used for computing moderated F statistics. For details of moderated statistics, see Smyth (2003).

Value

comp.modF returns a function (F) with the bindings for L. The function F when supplied with a microarray data matrix and evaluated will return a numeric vector of moderated F statistics for each row of the matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Lonnstedt, I. and Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica* **12**, 31-46.

Smyth, G. K. (2003). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. http://www.statsci.org/smyth/pubs/ebayes.pdf

See Also

comp.FC, comp.modt, comp.stat

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1
fmod <- comp.modF(L)
fmod.X <- fmod(X)

# Another way of computing moderated F statistics
fmod.X <- comp.stat(X, L, "modf")
```

`comp.modt` *Computing Moderated t-statistics for Differential Expression*

Description

`comp.modt` returns a function of one argument with bindings for `L`. This function accepts a microarray data matrix as its single argument, when evaluated, computes moderated t-statistics by empirical Bayes shrinkage of the standard error toward a common value.

Usage

```
comp.modt(L = NULL)
```

Arguments

`L`	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.

Details

The function returned by `comp.modt` computes moderated t statistics for the assessment of differential expression. It interfaces to a C function. `comp.stat` is another function that wraps around the same C function that could be used for computing moderated t statistics. For details of moderated statistics, see Smyth (2003).

Value

`comp.modt` returns a function (F) with the bindings for `L`. The function F when supplied with a microarray data matrix and evaluated will return a numeric vector of moderated t statistics for each row of the matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Lonnstedt, I. and Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica* 12, 31-46.

Smyth, G. K. (2003). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. http://www.statsci.org/smyth/pubs/ebayes.pdf

See Also

comp.FC, comp.t.

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

tmod <- comp.modt(L)
tmod.X <- tmod(X)

# Another way of computing moderated t statistics
tmod.X <- comp.stat(X, L, "modt")
```

---

comp.SAM *Computing SAM Statistics for Differential Expression*

Description

comp.SAM returns a function of one argument. This function has a environment with bindings for a series of arguments (see below). It accepts a microarray data matrix as its single argument, when evaluated, computes SAM statistics for each row of the matrix.

Usage

```
comp.SAM(L = NULL, prob = 0.5, B = 200, stat.only = TRUE, verbose = FALSE,
deltas, s.step=0.01, alpha.step=0.01, plot.it=FALSE)
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
prob	A numeric variable used to set the fudge factor s_0 in terms of the percentile of the standard deviations of the genes. If set as NULL, s_0 is calculated using the algorithm by Tusher et al. (see reference).
B	The number of permutations. For a complete enumeration, B should be 0 (zero) or any number not less than the total number of permutations.
stat.only	A logical variable, if TRUE, only statistics are calculated and returned; if FALSE, false discovery rates (FDRs) for a set of δ(deltas) are calculated and returned.
verbose	A logical variable, if TRUE, informative messages are printed during the computation process.
deltas	A vector of values for the threshold δ; see Tusher et al.
s.step	A numeric variable specifying the size of the moving window across the gene-wise standard deviations for the selection of the fudge factor s_0.
alpha.step	A numeric variable specifying the increment of a percentile sequence between 0 and 1, from which the fudge factor will be chosen to minimize the coefficient of variation of statistics.
plot.it	A logical variable, if TRUE, a plot between the coefficient of variation and the percentile sequence will be made.

Details

The function returned by `comp.SAM` calculates SAM statistics for each row of the microarray data matrix, with bindings for `L`, `prob`, `B`, `stat.only`, `verbose`, `deltas`, `s.step`, `alpha.step` and `plot.it`. If `quantile=NULL`, the fudge factor s_0 is calculated as the percentile of the gene-wise standard deviations that minimizes the coefficient of variation of the statistics; otherwise s_0 is set as the specified percentile of standard deviations. If `stat.only=T`, only SAM statistics are returned; otherwise, permutation will be carried out to calculate the FDRs for a set of `deltas` specified and a FDR table will be returned in addition to the SAM statistics.

Value

`SAM` returns a function (F) with bindings for a series of arguments. When `stat.only=T`, the function F when evaluated returns a numeric vector of SAM statistics; When `stat.only=F`, the function F when evaluated returns a list of the following components:

`geneOrder`	Order of genes in terms of differential expression;
`sam`	Sorted SAM statistics;
`fdr.table`	A matrix with columns: delta, no.significance, no.positive, no.negatvie, FDR(50%), FDR(90%).

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response, *PNAS*, 98, 5116-5121.

See Also

`comp.t`

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# two sample test, statistics only
sam.fun <- comp.SAM(L)
sam.X <- sam.fun(X)

# two sample test, FDR
sam.fun <- comp.SAM(L, stat.only=FALSE, delta=c(0.1, 0.2, 0.5))
sam.X <- sam.fun(X)
```

`comp.stat` *Computing Test Statistics for Differential Expression*

Description

This function computes test statistics, e.g., t-statistics, F-statistics, SAM, fold changes, moderated t or F statistics, B statistics, for each row of a microarray data matrix.

Usage

```
comp.stat(X, L, test = c("t", "fc", "sam", "f", "modt", "modf", "B"), extra = NULL)
```

Arguments

`X`	A matrix, with m rows corresponding to variables (hypotheses) and n columns to observations. In the case of gene expression data, rows correspond to genes and columns to mRNA samples. The data can be read using `read.table`.
`L`	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
`test`	A character string specifying the statistic to be used to test the null hypothesis of no association between the variables and the class labels.

`test="t"`:	t-statistics;
`test="f"`:	F-statistics;
`test="fc"`:	fold changes;
`test="sam"`:	SAM-statistics;
`test="modt"`:	moderated t-statistics;
`test="modf"`:	moderated F-statistics;
`test="B"`:	B-statistics.

`extra`	Extra parameter needed for the test specified; see `deds.genExtra`.

Details

The function `comp.stat` interfaces to a C function and computes various statistics for differential expression in the C environment and therefore faster than functions in R. However, functions in R that are implemented in the DEDS packages may have more flexibility in terms of specifications of arguments. Below is a table the details `comp.stat` and its equivalent R functions in the DEDS package. Note that all the R functions listed in the 2nd column of the table below return a function with bindings for a series of arguments which accept the microarray data matrix as its single argument and compute accordingly statistics.

Interface to C	R functions
deds.stat(X, L, test="t")	tTest(L=NULL, mu=0, var.equal=FALSE)
deds.stat(X, L, test="fc")	FC(L=NULL, is.log=TRUE, FUN=mean)
deds.stat(X, L, test="sam")	Sam(L=NULL, prob=0.5, B=200, stat.only=TRUE, verbose=FALSE, deltas, s.step=0.01, a
deds.stat(X, L, test="f")	fTest(L=NULL)
deds.stat(X, L, test="modt")	tmodTest(L=NULL)
deds.stat(X, L, test="modf")	fmodTest(L=NULL)

deds.stat(X, L, test="B") BTest(L=NULL, proportion=0.01)

Value

A vector of test statistics for each row of the matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

For references on B-statistics and moderated t and F statistics:

Lonnstedt, I. and Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica* **12**, 31-46.

Smyth, G. K. (2003). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. http://www.statsci.org/smyth/pubs/ebayes.pdf

See Also

deds.genExtra, for B statistics: lm.series and ebayes

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# t statistics
tstat <- comp.stat(X, L, test="t")

# SAM, fudge factor set as the median of pooled genewise standard deviations
samstat <- comp.stat(X, L, test="sam")
# SAM, fudge factor set as the 90% of pooled genewise standard deviations
samstat <- comp.stat(X, L, test="sam", extra=c(0.9))

# moderated t
modtstat <- comp.stat(X, L, test="modt")

# B, proportion of differentially expressed genes is set at default, 1%
Bstat <- comp.stat(X, L, test="B")
# B, proportion of differentially expressed genes is set at  10%
Bstat <- comp.stat(X, L, test="B", extra=c(0.1))
```

comp.t *Computing One and Two Sample t-statistic for Differential Expression*

Description

comp.t returns a function of one argument with bindings for L, mu, var.equal. This function accepts a microarray data matrix as its single argument, when evaluated, computes t statistics for each row of the matrix.

Usage

```
comp.t(L = NULL, mu = 0, var.equal = FALSE)
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k - 1$.
mu	A number indicating the true value of the mean (or difference in means if you are performing a two sample test).
var.equal	a logical variable indicating whether to treat the two variances as being equal. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch statistic will be calculated.

Details

The function returned by comp.t calculates t statistics for each row of the microarary data matrix, given specific class labels.

Value

comp.t returns a function with bindings for L, mu, var.equal, which calculates and returns of vector of t statistics for each row in the data matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

comp.FC, comp.F

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# two sample test, unequal variance
t.fun <- comp.t(L)
```

```
t.X <- t.fun(X)

# two sample test, equal variance
t.fun <- comp.t(L, var.equal=TRUE)
t.X <- t.fun(X)
```

`comp.unadjp` *Computing permutation based unadjusted p values for each row of a matrix*

Description

This function computes permutation based unadjusted p values for a selected test statistic, e.g., one- or two-sample t-statistics, F-statistics, SAM, Fold change, for each row of a matrix.

Usage

```
comp.unadjp(X, L, B = 1000, test = c("t", "fc", "sam", "f"), tail = c("abs", "lower", "higher"),
```

Arguments

`X`	A matrix, with m rows corresponding to variables (hypotheses) and$n$ columns corresponding to observations. In the case of gene expression data, rows correspond to genes and columns to mRNA samples. The data can be read using `read.table`.
`L`	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
`B`	The number of permutations. For a complete enumeration, `B` should be 0 (zero) or any number not less than the total number of permutations.
`test`	A character string specifying the statistic to be used to test the null hypothesis of no association between the variables and the class labels. If `test="t"`, for one-class, the tests are based on one-sample t-statistics; for two-class, the tests are based on two-sample t-statistics (unequal variances). If `test="f"`, the tests are based on F-statistics. If `test="fc"`, the tests are based on fold changes among classes. If `test="sam"`, the tests are based on SAM-statistics.
`tail`	A character string specifying the type of rejection region. If `side="abs"`, two-tailed tests, the null hypothesis is rejected for large absolute values of the test statistic. If `side="higher"`, one-tailed tests, the null hypothesis is rejected for large values of the test statistic. If `side="lower"`, one-tailed tests, the null hypothesis is rejected for small values of the test statistic.
`extra`	Extra parameter need for the test specified; see `deds.genExtra`.

Details

The function `comp.unadjp` computes unadjusted p values using a permutation scheme.

Value

A vector of unadjusted p values for each row of the matrix.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

deds.genExtra, comp.stat

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# t statistics
unadjp.t <- comp.unadjp(X, L, test="t")
```

DEDS-class *DEDS Result List - class*

Description

A simple list-based class to store DEDS results.DEDS objects are usually created by functions deds.pval, deds.stat and deds.stat.linkC.

Slots/List Components

DEDS objects can be created by new("DEDS",deds) where deds is a list. This class contains no slots, but objects should contain the following list components:

E:	A numeric vector of the most extreme point in the direction of differential expression.
p:	A numeric vector of q- or adjusted p-values.
geneOrder:	An integer vector giving the index of the top genes in terms of differential expression.
stats:	A matrix of p values or statistics.
options:	A character vector of options used in the test.

Methods

This class inherits directly from class list, so any operation appropriate for lists will work on objects of this class. In addition, Other functions which operate on DEDS objects include pairs and hist.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

deds.pval, deds.stat, deds.stat.linkC

---

deds.chooseTest *Selection of the Most Common Statistics for Differential Expression*

Description

This function selects a set of functions of common statistics for differential expression in microarray data analysis, given specific observation class labels. As a default, t-statistics, fold change and SAM are selected.

Usage

```
deds.chooseTest(L = NULL, tests = c("t", "sam", "fc"))
```

Arguments

L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
tests	A character vector specifying the statistics to be used to test the null hypothesis of no association between the variables and the class labels. For DEDS, there should be more than one statistic chosen from the following:

"t":	t-statistics;
"f":	F-statistics;
"fc":	fold changes;
"sam":	SAM-statistics;
"modt":	moderated t-statistics;
"modf":	moderated F-statistics;
"B":	B-statistics.

Details

deds.chooseTest can be used together with the function deds.stat. The user specifies the types of statistics needed for subsequent DEDS analysis by the argument tests and the function returns accordingly a list the statistics function, which could be used for input testfun in the function deds.stat.

Value

A list of statistics functions specified by the user which could be used for input in the function deds.stat.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

comp.t, comp.FC, comp.SAM

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# as a default, chooses t, fc and sam
funcs <- deds.chooseTest(L)
deds.X <- deds.stat(X, L, testfun=funcs)

# chooses F statistic, SAM statistic, and moderated F statistic
L <- rep(0:2, c(3,3,4))
funcs <- deds.chooseTest(L, tests=c("f", "sam", "modf"))
```

---

deds.genExtra — *Generating Extra Parameters for Test Statistics Functions for Differential Expression*

Description

deds.genExtra is used to pass in extra arguments for comp.stat and deds.stat.linkC, which computes various test statistics for differential expression in microarray data.

Usage

```
deds.genExtra(classlabel, tests)
```

Arguments

classlabel	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
tests	A vector of character string specifying the statistics to be used to test differential expression. The character string could be any of the followings:

test="t":	one- or two-sample t-statistics;
test="f":	F-statistics;
test="fc":	fold changes among classes;
test="sam":	SAM-statistics;
test="modt":	moderated t-statistics;
test="modf":	moderated F-statistics;
test="B":	B-statistics.

Details

Given the names of the test statistics, `deds.genExtra` generates extra parameters needed to be passed in the functions `comp.stat` and `deds.stat.linkC` for the assessment of differential expression. Both functions are interfaces to C functions. `deds.genExtra` generates default parameters as follows:
If `test="t" or "f", "fc", "modt", "modf"`, the extra parameter needed is the number of classes;
If `test="sam"`, the extra parameter needed is the percentile of within-gene standard deviations that the fudge factor s_0 will be set at and the default is 0.5;
If `test="B"`, the extra parameter needed is the percentage of alternative hypotheses (differential expression) and the default is set at 0.01.

Value

A numeric vector, the length of which is determined by the length of the names of the test statistics for the argument `test`.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

`comp.stat`, `deds.stat.linkC`

Examples

```
## two sample test
L <- rep(0:1, c(5,5))
extras <- deds.genExtra(L, c("t","sam", "B"))
## extras will be c(2, 0.5, 0.01)
```

`deds.pval` *Differential Expression via Distance Summary of p Values from Multiple Models*

Description

`deds.pval` integrates different p values of differential expression (DE) to rank and select a set of DE genes.

Usage

```
deds.pval(X, E = rep(0, ncol(X)), adj = c("fdr", "adjp"), B = 200, nsig = nrow(X))
```

Arguments

`X`	A matrix, with m rows corresponding to variables (hypotheses) and n columns corresponding to p values from different statistical models.
`E`	A numeric vector indicating the location of the most extreme p values in the direction of differential expression.

adj	A character string specifying the type of multiple testing adjustment. If adj="fdr", False Discovery Rate is controlled and q values are returned. If adj="adjp", adjusted p values that controls family wise type I error rate is returned.
B	The number of permutations. For a complete enumeration, B should be 0 (zero) or any number not less than the total number of permutations.
nsig	A numeric variable specifying the number of top genes that will be returned.

Details

deds.pval summarizes p values from multiple statistical models for the evidence of DE. The DEDS methodology treats each gene as a point corresponding to a gene's vector of DE measures. An "extreme origin" is defined as the point that indicate DE, typically a vector of zero p values. The distance from all points to the extreme is computed and the ranking of a gene for DE is determined by the closeness of the gene to the extreme. To determine a cutoff for declaration of DE, null referent distributions are generated using an approach similar to the gap statistic (see Reference below). DEDS can also summarize different statistics, see deds.stat and deds.stat.linkC.

Value

An object of class DEDS. See DEDS-class.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Tibshirani, R., Walther G., and Hastie T. (2000). Estimating the number of clusters in a dataset via the gap statistic. Department of Statistics, Stanford University, http://www-stat.stanford.edu/~tibs/ftp/gap.ps

Yang, Y.H., Xiao, Y. and Segal M.R.: Selecting differentially expressed genes from microarray experiment by sets of statistics. *Bioinformatics* 2005 21:1084-1093.

See Also

deds.stat, deds.stat.linkC.

---

deds.stat *Differential Expression via Distance Summary of Multiple Statistics*

Description

deds.stat integrates different statistics of differential expression (DE) to rank and select a set of DE genes.

Usage

```
deds.stat(X, L, B = 1000, testfun = list(t = comp.t(L), fc = comp.FC(L),
sam = comp.SAM(L)), tail = c("abs", "lower", "higher"), distance =
c("weuclid", "euclid"), adj = c("fdr", "adjp"), nsig = nrow(X))
```

Arguments

X	A matrix, with m rows corresponding to variables (hypotheses) and n columns corresponding to observations. In the case of gene expression data, rows correspond to genes and columns to mRNA samples. The data can be read using read.table.
L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
B	The number of permutations. For a complete enumeration, B should be 0 (zero) or any number not less than the total number of permutations.
testfun	A list of functions specifying the statistics to be used to test the null hypothesis of no association between the variables and the class labels. The default uses t, fold change and SAM. The input can also be generated using the function deds.chooseTest.
tail	A character string specifying the type of rejection region. If side="abs", two-tailed tests, the null hypothesis is rejected for large absolute values of the test statistic. If side="higher", one-tailed tests, the null hypothesis is rejected for large values of the test statistic. If side="lower", one-tailed tests, the null hypothesis is rejected for small values of the test statistic.
distance	A character string specifying the type of distance measure used for the calculation of the distance to the extreme point (E). If distance="weuclid", weighted euclidean distance, the weight for statistic t is $\frac{1}{MAD(t)}$; If distance="euclid", euclidean distance.
adj	A character string specifying the type of multiple testing adjustment. If adj="fdr", False Discovery Rate is controled and q values are returned. If adj="adjp", ajusted p values that controls family wise type I error rate is returned.
nsig	If adj = "fdr", nsig specifies the number of top differentially expressed genes whose q values will be calculated; we recommend setting nsig < m, as the computation of q values will be extensive. q values for the rest of genes will be approximated to 1. If adj = "adjp", the calculation of the adjusted p values will be for the whole dataset.

Details

deds.stat summarizes multiple statistical measures for the evidence of DE. The DEDS methodology treats each gene as a point corresponding to a gene's vector of DE measures. An "extreme origin" is defined as the maxima of all statistics and the distance from all points to the extreme is computed and ranking of a gene for DE is determined by the closeness of the gene to the extreme. To determine a cutoff for declaration of DE, null referent distributions are generated by permuting the data matrix.

Statistical measures currently in the DEDS package include t statistics (comp.t), fold changes(comp.FC), F statistics (comp.F), SAM ((comp.SAM), moderated t (comp.modt), moderated F statistics (comp.modF), and B statistics (comp.B). The user can also supply their own function for a statistic other than the above, provided the function is written in a similar format as the above ones.

The function deds.stat could be slow if the size of the data matrix and the number of permutations are big. We hence recommend the user to use deds.stat.linkC as the default function.

deds.stat.linkC interfaces to a C function, which handles a 10,000 by 10 matrix and 1000 permutations in minutes.

DEDS can also summarize p values from different statistical models, see deds.pval.

Value

An object of class DEDS. See DEDS-class.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Yang, Y. H., Xiao, Y. and Segal MR: Selecting differentially expressed genes from microarray experiment by sets of statistics. *Bioinformatics*, 2004, accepted. http://www.biostat.ucsf.edu/jean/Papers/DEDS.pdf.

See Also

deds.pval, deds.stat.linkC

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1

# DEDS summarizing t, sam and fc
deds.X <- deds.stat(X, L, B=200)

# DEDS summarizing t, tmod and fc
## Not run: deds.X <- deds.stat(X, L, testfun=list(t=comp.t(L),
tmod=comp.modt(L), sam=comp.SAM(L)))
## End(Not run)

# one can also use:
## Not run: deds.X <- deds.stat(X, L, testfun=deds.chooseTest(L,
tests=c("t","modt","fc")))

## End(Not run)
```

---

deds.stat.linkC *Differentail Expression via Distance Summary of Multiple Statistics*

Description

deds.stat.linkC integrates different statistics of differential expression (DE) to rank and select a set of DE genes.

Usage

```
deds.stat.linkC(X, L, B = 1000, tests = c("t", "fc", "sam"), tail =
c("abs", "lower", "higher"), extras = NULL, distance = c("weuclid",
"euclid"), adj = c("fdr", "adjp"), nsig = nrow(X), quick = TRUE)
```

Arguments

X	A matrix, with m rows corresponding to variables (hypotheses) and n columns corresponding to observations. In the case of gene expression data, rows correspond to genes and columns to mRNA samples. The data can be read using `read.table`.
L	A vector of integers corresponding to observation (column) class labels. For k classes, the labels must be integers between 0 and $k-1$.
B	The number of permutations. For a complete enumeration, B should be 0 (zero) or any number not less than the total number of permutations.
tests	A character vector specifying the statistics to be used to test the null hypothesis of no association between the variables and the class labels, test could be any of the following:

"t":	one or two sample t-statistics;
"f":	F-statistics;
"fc":	fold changes among classes;
"sam":	SAM-statistics;
"modt":	moderated t-statistics;
"modt":	moderated F-statistics;
"B":	B-statistics.

tail	A character string specifying the type of rejection region. If side="abs", two-tailed tests, the null hypothesis is rejected for large absolute values of the test statistic. If side="higher", one-tailed tests, the null hypothesis is rejected for large values of the test statistic. If side="lower", one-tailed tests, the null hypothesis is rejected for small values of the test statistic.
extras	Extra parameter needed for the test specified; see `deds.genExtra`.
distance	A character string specifying the type of distance measure used for the calculation of the distance to the extreme point (E). If distance="weuclid", weighted euclidean distance, the weight for statistic t is $\frac{1}{MAD(t)}$; If distance="euclid", euclidean distance.
adj	A character string specifying the type of multiple testing adjustment. If adj="fdr", False Discovery Rate is controled and q values are returned. If adj="adjp", ajusted p values that controls family wise type I error rate are returned.
nsig	If adj = "fdr", nsig specifies the number of top differentially expressed genes whose q values will be calculated; we recommend setting nsig < m, as the computation of q values will be extensive. q values for the rest of genes will be approximated to 1. If adj = "adjp", the calculation of the adjusted p values will be for the whole dataset.

quick — A logical variable specifying if a quick but memory requiring procedure will be selected. If quick=TRUE, permutation will be carried out once and stored in memory; If quick=FALSE a fixed seeded sampling procedure will be employed, which requires more computation time as the permutation will be carried out twice, but will not use extra memory for storage.

Details

deds.stat.linkC summarizes multiple statistical measures for the evidence of DE. The DEDS methodology treats each gene as a point corresponding to a gene's vector of DE measures. An "extreme origin" is defined as the maxima of all statistics and the distance from all points to the extreme is computed and ranking of a gene for DE is determined by the closeness of the gene to the extreme. To determine a cutoff for declaration of DE, null referent distributions are generated by permuting the data matrix.

Statistical measures currently in the DEDS package include t statistics (tests="t"), fold changes (tests="fc"), F statistics (tests="f"), SAM (tests="sam"), moderated t (tests="modt"), moderated F statistics (tests="modf"), and B statistics (tests="B"). The function deds.stat.linkC interfaces to C functions for the tests and the computation of DEDS. For more flexibility, the user can also use deds.stat which has the same functionality as deds.stat.linkC but is written completely in R (therefore slower) and the user can supply their own function for a statistic not covered in the DEDS package.

DEDS can also summarize p values from different statistical models, see deds.pval.

Value

An object of class DEDS. See DEDS-class.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

References

Yang, Y.H., Xiao, Y. and Segal M.R.: Selecting differentially expressed genes from microarray experiment by sets of statistics. *Bioinformatics* 2005 21:1084-1093.

See Also

deds.pval, deds.stat.

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1
# DEDS summarizing t, fc and sam
d <- deds.stat.linkC(X, L, B=200)
```

hist-methods *Histogram for DEDS Objects*

Description

The function hist.DEDS produces histograms of unadjusted p-values for DEDS-class objects.

Usage

```
## S3 method for class 'DEDS'
hist(x, subset=c(1:nrow(x$stats)), ...)
```

Arguments

x	An object of DEDS, produced by deds.pval.
subset	A numeric vector indicating the subset of points to be plotted.
...	Further graphical parameters, for example, "col", "border", "main", "nclass".

Details

The function hist.DEDS implements a S3 method of hist for DEDS. The DEDS class is a simple list-based class to store DEDS results and hist.DEDS is used for a DEDS object that is created by the function deds.pval. The list contains a "stat" component, which stores unadjusted p-values from various statistical models. The function hist.DEDS extracts the "stat" component and produces a histogram of the unadjusted p-values for each model.

For DEDS objects that are created by functions deds.stat and deds.stat.linkC, the "stat" matrix consists of different types of statistics. For graphical display of these statistics, the user can use qqnorm.DEDS and pairs.DEDS.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

deds.stat, deds.pval, deds.stat.linkC, pairs.DEDS, qqnorm.DEDS

pairs-methods *Pairs Plot for DEDS Objects*

Description

The function pairs-DEDS produces pairs plots of statistics or p values for DEDS-class objects.

Usage

```
## S3 method for class 'DEDS'
pairs(x, subset=c(1:nrow(x$stats)), labels =
colnames(x$stats[,-1]), logit = FALSE,
diagonal = c("qqnorm", "boxplot", "density", "histogram", "none"),
lower = c("cor", "none"), groups.by.deds = TRUE, thresh = 0.05, reg.line
= NULL, smooth = FALSE, line.by.group = FALSE, diag.by.group = TRUE, lower.by.group =
FALSE, col = palette(), pch = 1:n.groups, lwd = 1, legend.plot =
length(levels(groups)) > 1, ...)
```

Arguments

x	An object of DEDS.
subset	A numeric vector indicating the subset of points to be plotted.
labels	A character vector specifying the names of the variables.
logit	A logical variable, if TRUE the variables are logged, useful when plotting p values.
diagonal	A character string specifying the type of plot to be applied in the diagonal panels.

diagonal="qqnorm":	qqnorm on the diagonal
diagonal="boxplot":	boxplot on the diagonal
diagonal="density":	density on the diagonal
diagonal="histogram":	hist on the diagonal
diagonal="none":	no special plot will be applied on the diagonal

lower	A character string specifying the function to be applied in the lower panels.

lower="cor":	absolute correlation will be put on the lower panel
none="cor";	no special function will be applied

groups.by.deds	A logcial variable, if TRUE, points will be separated into groups according to their magnitude of q- or p-values by DEDS.
thresh	A numeric variable, if thresh<1, it specifies the threshold of significance in differential expression (DE) for q- or p-values of the DEDS object; default is set at 0.05. If thresh>1, it specifies the number of top DE genes to be highlighted.
reg.line	A function name specifying the type of regression line to be plotted in the scatter plots. If reg.line=lm, linear regression line will be plotted; If reg.line=NULL, no regression line will be plotted in the scatter plot.
smooth	A logical variable specifying if smooth regression lines will be plotted in the scatter plots. If smooth=TRUE, a lowess line will be applied.
line.by.group	A logical variable specifying if the regression lines should be applied within groups.
diag.by.group	A logical variable specifying if the plot in the diagonal panels would be applied groupwise.
lower.by.group	A logical variable, if lower.by.group=TRUE and lower="cor", correlation coefficients will be calculated and printed separated according to groups in the lower panels.

col	A specification for the colors to be used for plotting different groups, see par.
pch	A specification for the type of points to be used for plotting different groups, see par.
lwd	A specification for the width of lines to be used if lines are plotted; see par.
legend.plot	A logical variable specifying if the legend will be plotted.
...	Extra parameters for plotting.

Details

The function pairs.DEDS implements a S3 method of pairs for DEDS. The DEDS class is a simple list-based class to store DEDS results and it is usually created by functions deds.pval, deds.stat, deds.stat.linkC. The list contains a "stat" component, which stores statistics or p values from various statistical tests. The function pairs.DEDS extracts the "stat" component and produces a matrix of scatterplot.

pairs.DEDS as a default highlights points (corresponding to genes) with adjusted p- or q-values less than a user defined threshold. The user can select among a series of options a plot for the diagonal panel; as a default, it produces a qqnorm for each column in the "stat" matrix. Both the diagonal and lower panels can be stratified by specifying the diag.by.group or lower.by.group arguments.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

deds.stat, deds.pval, deds.stat.linkC, hist.DEDS, qqnorm.DEDS

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1
# DEDS
d <- deds.stat.linkC(X, L, B=200)

# pairs plot
pairs(d)
# plot regression line
pairs(d, reg.line=lm, lwd=2)
# histogram in the diagonal panel
pairs(d, diagonal="hist")
# boxplot on the diagonal panel and stratified
pairs(d, diagonal="boxplot", diag.by.group=TRUE)
```

qqnorm-methods *Normal Q-Q Plot for DEDS Objects*

Description

The function qqnorm.DEDS produces normal Quantile-Quantile plots of statistics for DEDS-class objects. The points corresponding to genes with DEDS q- or adjusted p-values less than a user defined threshold are highlighted.

Usage

```
## S3 method for class 'DEDS'
qqnorm(y, subset=c(1:nrow(y$stats)),
xlab = "Quantiles of standard normal", thresh = 0.05, col = palette(), pch, ...)
```

Arguments

y	An object of DEDS, produced by deds.stat.linkC or deds.stat.
subset	A numeric vector indicating the subset of points to be plotted.
xlab	A title for the x axis
thresh	A numeric variable specifying the threshold of significance in differential expression (DE) for q- or p-values of the DEDS object.
col	A specification for the colors to be used for plotting. It should have a length bigger than two. The first is used for points with q- or adjusted p-values smaller than the specified threshold (group I) and the second for points with q- or adjusted p-values bigger than the threshold (group II).
pch	A specification for the type of points to be used for plotting. It should have a length bigger than two. The first parameter is used for group I genes, and the second for group II genes.
...	Extra parameters for plotting.

Details

The function qqnorm.DEDS implements a S3 method of qqnorm for DEDS. The DEDS class is a simple list-based class to store DEDS results and qqnorm.DEDS is used for a DEDS object that is created by functions deds.stat, deds.stat.linkC. The list contains a "stat" component, which stores statistics from various statistical tests. The function qqnorm.DEDS extracts the "stat" component and produces a normal QQ plot for each type of statistics. qqnorm.DEDS as a default highlights points (corresponding to genes) with DEDS adjusted p- or q-values less than a user defined threshold.

For DEDS objects that are created by the function deds.pval, the "stat" matrix consists of unadjusted p-values from different statistical models. For graphical display of these p values, the user can use hist.DEDS and pairs.DEDS.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

deds.stat, deds.pval, deds.stat.linkC, hist.DEDS, qqnorm.DEDS

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1
# DEDS summarizing t, fc and sam
d <- deds.stat.linkC(X, L, B=200)

# qqnorm for t, fc and sam
qqnorm(d)
# change points color
qqnorm(d, col=c(2,3))
# change points type
qqnorm(d, pch=c(1,2))
```

---

topgenes *Table of Top Genes from DEDS*

Description

topgenes prints a table of top-ranked genes by DEDS.

Usage

```
topgenes(obj, number = 10, genelist = NULL, sort.by = c("deds", colnames(obj$stats[,-1])), tail
```

Arguments

obj	An object of DEDS
number	A numeric variable specifying the number of top genes to be printed out.
genelist	A data.frame or a vector containing gene names.
sort.by	A character string specifying the name of the statistic to sort genes by. The default uses the DEDS result, the user can also choose from the names of the statistics (or unadjusted p values) that DEDS is used to summarize.
tail	A character string specifying the type of rejection region. If side="abs", two-tailed tests, genes are ranked by their absolute values. If side="higher", one-tailed tests, genes are ranked decreasingly. If side="lower", one-tailed tests, genes are ranked increasingly.

Details

The function topgenes accepts a DEDS object as the first argument. The DEDS class is a simple list-based class to store DEDS results. The list contains a "stat" component, which stores statistics or unadjusted p-values from various statistical tests. The function topgenes.DEDS extracts the "stat" component and prints out the top genes according to the user defined criterion – usually by DEDS or by a single statistical measure that DEDS summarizes.

Value

A data.frame with rows for selected genes, and columns for the "stat" matrix and q- or adjusted p-values from DEDS.

Author(s)

Yuanyuan Xiao, <yxiao@itsa.ucsf.edu>,
Jean Yee Hwa Yang, <jean@biostat.ucsf.edu>.

See Also

deds.stat.linkC, deds.pval

Examples

```
X <- matrix(rnorm(1000,0,0.5), nc=10)
L <- rep(0:1,c(5,5))

# genes 1-10 are differentially expressed
X[1:10,6:10]<-X[1:10,6:10]+1
# DEDS summarizing t, fc and sam
d <- deds.stat.linkC(X, L, B=200)

# top table, ranked by DEDS
topgenes(d)
# top table, ranked by t
topgenes(d, sort.by="t")
```

Index