Package ‘metaMA’

February 20, 2015

Type Package
Title Meta-analysis for MicroArrays
Version 3.1.2
Date 2015-01-28
Author Guillemette Marot
Maintainer Guillemette Marot <guillemette.marot@inria.fr>
Depends R (>= 3.1.2),
Imports limma, SMVar
Suggests GEOquery, org.Hs.eg.db, VennDiagram, annaffy, hgu133plus2.db,
hgu133a.db, hgu95av2.db
Description Combines either p-values or modified effect sizes from different
studies to find differentially expressed genes
License GPL
LazyLoad yes
NeedsCompilation no
Repository CRAN
Date/Publication 2015-01-28 23:20:03

R topics documented:

metaMA-package .................................................. 2
calcfit2Diffrep .................................................... 3
directEScombi ..................................................... 3
directpvalcombi .................................................. 4
effectsize .......................................................... 5
EScombination ..................................................... 6
EScombination.paired ............................................. 7
IDDIRR .............................................................. 9
pvalcombination .................................................. 10
pvalcombination.paired .......................................... 11
row.ttest.stat .................................................... 12
# Description

Combines either p-values or moderated effect sizes from different studies to find differentially expressed genes.

# Details

- **Package:** metaMA
- **Type:** Package
- **Version:** 3.1.2
- **Date:** 2015-01-28
- **License:** GPL
- **LazyLoad:** yes

pvalcombination and EScombination are the most important functions to combine unpaired data. pvalcombination combines p-values from individual studies. EScombination combines effect sizes from individual studies. pvalcombination.paired and EScombination.paired are to be used for paired data. IDDIDR can help in the interpretation of gain and loss of information due to meta-analysis.

## Author(s)

Guillemette Marot <guillemette.marot@inria.fr>

## References


## Examples

```r
library(metaMA)
data(Singhdata)
EScombination(esets=Singhdata$esets,classes=Singhdata$classes)
pvalcombination(esets=Singhdata$esets,classes=Singhdata$classes)
#more details are provided in the vignette; only open it in interactive R sessions
if(interactive()){
```

calcfit2Diffrep

Empirical Bayes statistics from limma analysis with unpaired data

Description

Computes empirical Bayes statistics from limma analysis with only one group effect.

Usage

calcfit2Diffrep(C1, C2)

Arguments

C1 Gene expression data of the arrays in the first condition. Each row of C1 corresponds to one spot, each column to one replicate.

C2 Gene expression data of the arrays in the second condition. Each row of C2 corresponds to one spot, each column to one replicate.

Details

Returns fit2 described in limma vignette. To be used with unpaired data.

Value

fit2

Note

see Bioconductor limma vignette

directEScombi

Direct effect size combination

Description

Combines effect sizes already calculated.

Usage

directEScombi(ES, varES, BHth = 0.05, useREM = TRUE)
Arguments

ES  Matrix of effect sizes. Each column of ES corresponds to one study and each row to one gene.

varES  Matrix of effect size variances. Each column of varES corresponds to one study and each row to one gene.

BHth  Benjamini Hochberg threshold. By default, the False Discovery Rate is controlled at 5%.

useREM  A logical value indicating whether or not to include the between-study variance into the model.

Details

Combines effect sizes with the method presented in (Choi et al., 2003).

Value

List

DEindices  Indices of differentially expressed genes at the chosen Benjamini Hochberg threshold.

TestStatistic  Vector with test statistics for differential expression in the meta-analysis.

References


---

directpvalcombi  Direct p-value combination

Description

Combines one sided p-values with the inverse normal method.

Usage

directpvalcombi(pvalonesided, nrep, BHth = 0.05)

Arguments

pvalonesided  List of vectors of one sided p-values to be combined.

nrep  Vector of numbers of replicates used in each study to calculate the previous one-sided p-values.

BHth  Benjamini Hochberg threshold. By default, the False Discovery Rate is controlled at 5%.
**Value**

List

DEindices  Indices of differentially expressed genes at the chosen Benjamini Hochberg threshold.

TestStatistic  Vector with test statistics for differential expression in the meta-analysis.

**Note**

One-sided p-values are required to avoid directional conflicts. Then a two-sided test is performed to find differentially expressed genes.

**Author(s)**

Guillemette Marot

**References**


---

**effectsize**

*Calculates effect sizes from given t or moderated t statistics*

**Description**

Function not to be used separately.

**Usage**

`effectsize(tstat, ntilde, m)`

**Arguments**

`tstat`  Vector of test statistics and effect sizes.

`ntilde`  Proportion factor between a test statistic and its corresponding effect size.

`m`  Number of degrees of freedom.

**Value**

Matrix with one row per gene, and in column:

`d`  Commonly used effect size (which is biased)

`vard`  Variance of the commonly used effect size

`dprime`  Unbiased effect size

`vardprime`  Variance of the unbiased effect size
EScombination

Effect size combination for unpaired data

Description

Calculates effect sizes from unpaired data either from classical or moderated t-tests (Limma, SMVar) for each study and combines these effect sizes.

Usage

EScombination(esets, classes, moderated = c("limma", "SMVar", "t")[1], BHth = 0.05)

Arguments

- **esets**: List of matrices (or data frames), one matrix per study. Each matrix has one row per gene and one column per replicate and gives the expression data for both conditions with the order specified in the classes argument. All studies must have the same genes. If the data are already stored as ExpressionSets objects (cf. Bioconductor project), then exprs(yourdata) will give an appropriate element of the list esets used for this function.

- **classes**: List of class memberships, one per study. Each vector or factor of the list can only contain two levels which correspond to the two conditions studied.

- **moderated**: Method to calculate the test statistic inside each study from which the effect size is computed. moderated has to be chosen between "limma", "SMVar" and "t".

- **BHth**: Benjamini Hochberg threshold. By default, the False Discovery Rate is controlled at 5%.
Value

List

Study1 Vector of indices of differentially expressed genes in study 1. Similar names are given for the other individual studies.

AllIndStudies Vector of indices of differentially expressed genes found by at least one of the individual studies.

Meta Vector of indices of differentially expressed genes in the meta-analysis.

TestStatistic Vector with test statistics for differential expression in the meta-analysis.

Note

While the invisible object resulting from this function contains the values described previously, other quantities of interest are printed: DE,IDD,Loss,IDR,IRR. All these quantities are defined in function IDDIDR and in (Marot et al., 2009)

Author(s)

Guillemette Marot

References


Examples

data(Singhdata)
#Meta-analysis
res=EScombination(esets=Singhdata$esets,classes=Singhdata$classes)
#Number of differentially expressed genes in the meta-analysis
length(res$Meta)
#To plot an histogram of raw p-values
rawpval=2*(1-pnorm(abs(res$TestStatistic)))
hist(rawpval,nclass=100)

EScombination.paired Effect size combination for paired data

Description

Calculates effect sizes from paired data either from classical or moderated t-tests (Limma, SMVar) for each study and combines these effect sizes.

Usage

EScombination.paired(logratios, moderated = c("limma", "SMVar", "t")[1], BHth = 0.05)
Arguments

logratios List of matrices (or data frames). Each matrix has one row per gene and one column per replicate and gives the logratios of one study. All studies must have the same genes.

moderated Method to calculate the test statistic inside each study from which the effect size is computed. moderated has to be chosen between "limma", "SMVar" and "t".

BHth Benjamini Hochberg threshold. By default, the False Discovery Rate is controlled at 5%.

Value

List

Study1 Vector of indices of differentially expressed genes in study 1. Similar names are given for the other individual studies.

AllIndStudies Vector of indices of differentially expressed genes found by at least one of the individual studies.

Meta Vector of indices of differentially expressed genes in the meta-analysis.

TestStatistic Vector with test statistics for differential expression in the meta-analysis.

Note

While the invisible object resulting from this function contains the values described previously, other quantities of interest are printed: DE, IDD, Loss, IDR, IRR. All these quantities are defined in function `iddidr` and in (Marot et al., 2009)

Author(s)

Guillemette Marot

References


Examples

data(Singhdata)
# create artificially paired data:
artificialdata=lapply(Singhdata$esets,FUN=function(x) (x[,1:10]-x[,11:20]))
# Meta-analysis
res=ESc combination.paired(artificialdata)
# Number of differentially expressed genes in the meta-analysis
length(res$Meta)
# To plot an histogram of raw p-values
rawpval=2*(1-pnorm(abs(res$TestStatistic)))
hist(rawpval,nclass=100)
Description

Calculates the gain or the loss of differentially expressed genes due to meta-analysis compared to individual studies.

Usage

IDDIRR(finalde, deindst)

Arguments

- `finalde`: Vector of indices of differentially expressed genes after meta-analysis
- `deindst`: Vector of indices of differentially expressed genes found at least in one study

Value

- **DE**: Number of Differentially Expressed (DE) genes
- **IDD**: Integration Driven Discoveries: number of genes that are declared DE in the meta-analysis that were not identified in any of the individual studies alone.
- **Loss**: Number of genes that are declared DE in individual studies but not in meta-analysis.
- **IDR**: Integration-driven Discovery Rate: proportion of genes that are identified as DE in the meta-analysis that were not identified in any of the individual studies alone.
- **IRR**: Integration-driven Revision Rate: percentage of genes that are declared DE in individual studies but not in meta-analysis.

Author(s)

Guillemette Marot

References


Examples

```r
data(Singhdata)
out=EScombination(esets=Singhdata$esets,classes=Singhdata$classes)
IDDIRR(out$Meta,out$AllIndStudies)

## The function is currently defined as
#function(finalde,deindst)
```
pvalcombination

P-value combination for unpaired data

Description

Calculates differential expression p-values from unpaired data either from classical or moderated t-tests (Limma, SMVar) for each study and combines these p-values by the inverse normal method.

Usage

pvalcombination(esets, classes, moderated = c("limma", "SMVar", "t")[1], BHth = 0.05)

Arguments

- **esets**: List of matrices (or data frames), one matrix per study. Each matrix has one row per gene and one column per replicate and gives the expression data for both conditions with the order specified in the classes argument. All studies must have the same genes. If the data are already stored as ExpressionSets objects (cf. Bioconductor project), then exprs(yourdata) will give an appropriate element of the list esets used for this function.

- **classes**: List of class memberships, one per study. Each vector or factor of the list can only contain two levels which correspond to the two conditions studied.

- **moderated**: Method to calculate the test statistic inside each study from which the p-value is computed. moderated has to be chosen between "limma", "SMVar" and "t".

- **BHth**: Benjamini Hochberg threshold. By default, the False Discovery Rate is controlled at 5%.

Value

- **list**: Vector of indices of differentially expressed genes found by at least one of the individual studies.

- **study1**: Vector of indices of differentially expressed genes in study 1. Similar names are given for the other individual studies.
Meta Vector of indices of differentially expressed genes in the meta-analysis.
TestStatistic Vector with test statistics for differential expression in the meta-analysis.

Note
While the invisible object resulting from this function contains the values described previously, other quantities of interest are printed: DE, IDD, Loss, IDR, IRR. All these quantities are defined in function `iddidr` and in (Marot et al., 2009)

Author(s)
Guillemette Marot

References

Examples
```r
data(singhdata)
#Meta-analysis
res=pvalcombination(logratios=singhdata$logratios, moderated=c("limma", "SMVar", "t"))[1], BHth = 0.05)
```

Description
Calculates differential expression p-values from paired data either from classical or moderated t-tests (Limma, SMVar) for each study and combines these p-values by the inverse normal method.

Usage
`pvalcombination.paired(logratios, moderated = c("limma", "SMVar", "t")[1], BHth = 0.05)`

Arguments
- `logratios` List of matrices. Each matrix has one row per gene and one column per replicate and gives the logratios of one study. All studies must have the same genes.
- `moderated` Method to calculate the test statistic inside each study from which the effect size is computed. `moderated` has to be chosen between "limma", "SMVar" and "t".
- `BHth` Benjamin-Hochberg threshold. By default, the False Discovery Rate is controlled at 5%.
Value

List

Study1 Vector of indices of differentially expressed genes in study 1. Similar names are given for the other individual studies.

AllIndStudies Vector of indices of differentially expressed genes found by at least one of the individual studies.

Meta Vector of indices of differentially expressed genes in the meta-analysis.

TestStatistic Vector with test statistics for differential expression in the meta-analysis.

Note

While the invisible object resulting from this function contains the values described previously, other quantities of interest are printed: DE, IDD, Loss, IDR, IRR. All these quantities are defined in function IDDIDR and in (Marot et al., 2009)

Author(s)

Guillemette Marot

References


Examples

data(Singhdata)
# create artificially paired data:
artificialdata=lapply(Singhdata$esets,FUN=function(x) (x[,1:10]-x[,11:20]))
# Meta-analysis
res=pvalcombination.paired(artificialdata)
# Number of differentially expressed genes in the meta-analysis length(res$Meta)
# To plot an histogram of raw p-values
drawpval=2*(1-pnorm(abs(res$TestStatistic)))
hist(rawpval,nclass=100)

row.ttest.stat Row t-tests

Description

Performs t-tests for unpaired data row by row.

Usage

row.ttest.stat(mat1, mat2)
Description

Performs t-tests for paired data row by row.

Usage

row.ttest.statp(mat)

Arguments

mat Matrix with data to be tested (for example, log-ratios in microarray experiments).

Details

This function is much faster than employing apply with FUN=t.test.
Value

Vector with t-test statistics.

Examples

## The function is currently defined as
```
function(mat){
m<-rowMeans(mat,na.rm=TRUE)
sd<-rowSds(mat,na.rm=TRUE)
tstat<-m/(sd*sqrt(1/dim(mat)[2]))
return(tstat)}```

Description

Calculates variances of each row of an array

Usage

```
rowVars(x, na.rm = TRUE)
```

Arguments

- **x**  
  Array of two or more dimensions, containing numeric, complex, integer or logical values, or a numeric data frame.

- **na.rm**  
  Logical. Should missing values (including NA) be omitted from the calculations?

Details

This function is the same as applying apply with FUN=var but is a lot faster.

Value

A numeric or complex array of suitable size, or a vector if the result is one-dimensional. The dimnames (or names for a vector result) are taken from the original array.

Examples

## The function is currently defined as
```
function (x, na.rm = TRUE)
{
sqr = function(x) x * x
n = rowSums(!is.na(x))
n[n <= 1] = NA
return(rowSums(sqr(x - rowMeans(x, na.rm = na.rm)), na.rm = na.rm)/(n - 1))
}
```
**Description**

Publicly available microarray dataset artificially split in 5 studies

**Usage**

data(Singhdata)

**Format**

List of 3 elements:

- **esets**  List of 5 data frames corresponding to 5 artificial studies, each with 12625 genes and 20 replicates (10 normal samples and 10 tumoral samples)
- **classes** List of 5 numeric vectors with class memberships, one per study
- **geneNames** Factor with 12625 levels corresponding to gene names

**Source**

These data are available on the website [http://www.bioinf.ucd.ie/people/ian/](http://www.bioinf.ucd.ie/people/ian/). We considered 50 normal samples and 50 tumoral samples, leaving out the 2 last tumoral samples. Data are already normalized.

**References**


**Examples**

data(Singhdata)
Index

*Topic datasets
Singhdata, 15

*Topic methods
  calcfit2Diffrep, 3
directEScombi, 3
directpvalcombi, 4
effects, 5
EScombination, 6
EScombination.paired, 7
IDDIRR, 9
pvalcombination, 10
pvalcombination.paired, 11
row.ttest.stat, 12
row.ttest.statp, 13
rowVars, 14

*Topic models
  calcfit2Diffrep, 3
directEScombi, 3
directpvalcombi, 4
effects, 5
EScombination, 6
EScombination.paired, 7
IDDIRR, 9
pvalcombination, 10
pvalcombination.paired, 11
row.ttest.stat, 12
row.ttest.statp, 13
rowVars, 14

*Topic package
  metaMA (metaMA-package), 2
  metaMA-package, 2
  pvalcombination, 10
  pvalcombination.paired, 11
  row.ttest.stat, 12
  row.ttest.statp, 13
  rowVars, 14
  Singhdata, 15