Package 'ABSSeq'

October 7, 2014

Type Package

Title ABSSeq: a new RNA-Seq analysis method based on absolute
expression differences and generalized Poisson model

Version 1.0.1

Author Wentao Yang

Maintainer Wentao Yang <wyang@zoologie.uni-kiel.de>

Description Inferring differential expression genes by absolute
expression differences between two groups, utilizing
generalized Poisson model to account for over-dispersion across
samples and heterogeneity of differential expression across genes.

License GPL (>= 3)

biocViews DifferentialExpression

Imports Rcpp

Depends R (>= 2.10), methods

Suggests

LinkingTo Rcpp

R topics documented:

ABSDataSet   *ABSDataSet object and constructors*

Description

ABSDataSet object and constructors

Usage

```
ABSDataSet(counts, groups, normMethod=c("User", "total","quartile", "DESeq"), sfactors=0)
```

Arguments

counts   a matrix or table with at least two columns and one row,

groups   a factor with two groups, whose length should be equal with sample size

normMethod   method for estimating the size factors, should be one of 'User', 'total', 'quartile'
                and 'DESeq'. See normalFactor for description.

sfactors   size factors for 'User' method, self-defined size factors by user.

Details

The function contructs an ABSDataSet object with counts table and groups. It also checks the structure of counts and groups.The ABSDataSet is a class, used to store the input values, intermediate calculations and results of an analysis of differential expression. It also contains information for the running time of an analysis.

Value

An ABSDataSet object.

Examples

```
counts <- matrix(1:4,ncol=2)
groups <- factor(c("a","b"))
obj <- ABSDataSet(counts, groups)
```

ABSSeq *Differential expression analysis based on the absolute difference.*

Description

This function performs a default analysis by calling, in order, the functions: [normalFactors](#), [calPara](#), [GPTest](#).

Usage

```
ABSSeq(object, perPrior=0.3, dcut=0, quiet=FALSE)
```

Arguments

object	an [ABSDataSet](#) object, contains the reads count matrix, groups and normalization method.
perPrior	default is 0.30, see [GPTest](#) for description.
dcut	default is 0, see [GPTest](#) for description.
quiet	default is FALSE, whether to print messages at each step

Details

The differential expression analysis uses a generalized generalized Poisson model of the form:

$$GP(D = k|\lambda, \theta) = (\lambda/(1 + \lambda * \theta)^k * (1 + \theta * k)^{(}k - 1)/k! * exp(-\lambda * (1 + \theta * k)/(1 + \lambda * \theta))$$

for $\lambda > 0, \theta >= 0, k = 0, 1, 2, \ldots$.

Value

a table with return a table with elements:

Amean and Bmean
          mean of reads count for group A and B

foldChange	log 2 of fold-change, B vs. A
pvalue	pvalue from GP model
adj.pvalue	adjuested p-value used BH method

Author(s)

Wentao Yang

References

Wentao Yang, Philip Rosenstiel & Hinrich Schulenburg: ABSSeq: a new RNA-Seq analysis method based on absolute expression differences and generalized Poisson model

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
res <- ABSSeq(obj)
head(res)
```

---

calPara   *Calculate parameters for generalized Poisson test (GPTest)*

Description

Calculate a set of parameters from normalized counts table before GPTest

Usage

```
calPara(object)
```

Arguments

object   an ABSDataSet object.

Value

An ABSDataSet object with absolute differences, basemean, mean of each group, variance, log2 of foldchange and mean of each group, named as 'absD', 'baseMean', 'Amean', 'Bmean', 'Variance' and 'foldChange', respectively. Use the results to get access it

Note

This function should run after normalFactors or providing size factors.

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
obj <- normalFactors(obj)
obj <- calPara(obj)
plotDifftoBase(obj)
```

---

counts   *Accessors for the 'counts' slot of a ABSDataSet object.*

Description

Accessors for the 'counts' slot of a ABSDataSet object, return a matrix

Usage

```
## S4 method for signature ABSDataSet
counts(object,norm=FALSE)

## S4 replacement method for signature ABSDataSet,matrix
counts(object)<-value
```

Arguments

object   a ABSDataSet object.

norm   logical indicating whether or not to normalize the counts before returning

value   an numeric matrix

Details

The counts slot holds the count data as a matrix of non-negative integer count values, rows and columns for genes and samples, respectively.

See Also

[sizeFactors](#), [normalFactors](#)

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
head(counts(obj))
counts(obj) <- matrix(1:50,nrow=5,ncol=10)
head(counts(obj))
```

---

estimateSizeFactorsForMatrix
                                Low-level function to estimate size factors with robust regression.

Description

This function is borrowed from DESeq.

Usage

```
estimateSizeFactorsForMatrix(counts, locfunc=median)
```

Arguments

counts   a matrix or data frame of counts, i.e., non-negative integer values

locfunc   a function to compute a location for a sample. By default, the median is used.

Details

Given a matrix or data frame of count data, this function estimates the size factors as follows:
Each column is divided by the geometric means of the rows. The median (or, if requested, another
location estimator) of these ratios (skipping the genes with a geometric mean of zero) is used as the
size factor for this column. Typically, you will not call this function directly.

Value

a vector with the estimates size factors, one element per column

Author(s)

Simon Anders

References

Simon Anders, Wolfgang Huber: Differential expression analysis for sequence count data. Genome
Biology 11 (2010) R106, <http://dx.doi.org/10.1186/gb-2010-11-10-r106>

Examples

```
data(simuN5)
dat=simuN5
estimateSizeFactorsForMatrix(dat$counts)
```

GPTest   *Testing the differential expression*

Description

Fitting generalized Poisson model and calculating p-value for each gene

Usage

```
GPTest(object, pPrior=0.3, dcut=0)
```

Arguments

object   an ABSDataSet object.

pPrior   the parameter for estimating rates of generalized Possion model for each gene,
                default is 0.30. The predefined parameter for estimating GP rate according to
                expression level of each gene. It is estimated by reducing the type I error rate
                in a acceptable level, i.e. 10% for perPrior=0.30 or 5% for perPrior=0.40. For
                details, please see the references,

dcut   the parameter for fitting GP model, default is 0.

Details

This function firstly fits the generalized Possion model used absolute differences between two
groups, then calculates the pvalue for each gene and finally adjusts the pvalues via BH method.

Value

an ABSDataSet object with additional elements: pvalue and adjusted p-value, denoted by pvalue
and adj-pvalue, respectively. Use the results to get access it.

Note

This function should run after calPara

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
obj <- normalFactors(obj)
obj <- calPara(obj)
obj <- GPTest(obj)
head(results(obj))
```

---

groups   *Accessors for the 'groups' slot of a ABSDataSet object.*

Description

Accessor functions for the 'groups' information in a ABSDataSet object.

Usage

```
## S4 method for signature ABSDataSet
groups(object)

## S4 replacement method for signature ABSDataSet,factor
groups(object)<-value
```

Arguments

object   an ABSDataSet object.

value   a factor object, includes two groups, equal with the number of samples

Details

The 'groups' is a factor object, contains the experiment design for differential expression analysis.
Its length should be equal with the sample size.

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
groups(obj)
groups(obj) <- factor(rep(c("A","B"),c(5,5)))
groups(obj)
```

---

normalFactors   *Estimating size factors from the reads count table*

Description

Function for esitmating size factors

Usage

```
normalFactors(object)
```

Arguments

object   an ABSDataSet object with element of 'counts' and 'normMethod', see the constructor functions `ABSDataSet`.

Details

Given a matrix of count data, this function esitmates the size factors by selected method. It provides four different methods for normalizing according to user-defined size factors, total reads, up quantile (75%) or DESeq (See `estimateSizeFactorsForMatrix`).

Value

an ABSDataSet object with the estimates size factors, one element per column. Use the `sizeFactors` to show it.

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
obj <- normalFactors(obj)
sizeFactors(obj)
```

---

normMethod   *Accessors for the 'normMethod' slot of a ABSDataSet object.*

Description

Accessor functions for the 'normMethod' information in a ABSDataSet object.

Usage

```
## S4 method for signature ABSDataSet
normMethod(object)

## S4 replacement method for signature ABSDataSet,character
normMethod(object)<-value
```

Arguments

object   an ABSDataSet object.

value   a character object, should be one of 'User', 'total', 'quartile' and 'DESeq'. See `normalFactors`

Details

The 'normMethod' is the method for calculating the size factors. Currently, Four methods: 'User', 'total', 'quartile' and 'DESeq' are available.

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
normMethod(obj)
normMethod(obj) <- "DESeq"
normMethod(obj)
```

---

plotDifftoBase *Plot absolute differencs*

Description

Plot absolute differencs against expression levels

Usage

```
plotDifftoBase(object, cols=c("black", "red"), fcut=1.5, pch=16,
    xlab="Expression level", ylab="ABS diffs", ...)
```

Arguments

object	an ABSDataSet object
fcut	the cutoff of fold-change
cols	the colors to mark the genes which greater or less than fcut
xlab	xlab
ylab	ylab
pch	pch
...	further arguments to plot

Details

Plot absolute differencs against expression levels and mark the gene with a color at a given cutoff of fold-change

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
obj <- normalFactors(obj)
obj <- calPara(obj)
plotDifftoBase(obj)
```

Description

Accessor functions for the result from a ABSDataSet by given names

Usage

```
## S4 method for signature ABSDataSet
results(object,cnames)
```

Arguments

object   an ABSDataSet

cnames   a vecotr of names for output, which are among:

'baseMean': average expression for each gene including all samples,

'Amean' and 'Bmean': average expression for each group:A and B,

'absD': absolute expression differences between two groups,

'foldChange': log2 of fold-change for each gene based on B vs. A,

'Variance': variance for each gene

'pvalue':pvalue from [GPTest](),

'adj.pvalue':ajusted pvalues by BH method. See [p.adjust]().

Details

This function returns the result of ABSSeq as a table or a vector depended on the given names.

Value

a table according to 'canmes'.

See Also

[ABSSeq]()

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
obj <- normalFactors(obj)
obj <- calPara(obj)
obj <- GPTest(obj)
head(results(obj))
head(results(obj,c("pvalue","adj.pvalue")))
```

---

simuN5 *Simulated study with random outliers*

Description

Simulated study with random outliers, include five samples for two groups. It contains counts table, groups and defined differential expression genes.

Usage

```
data(simuN5)
```

Format

The format is: List of 3

$ counts: integer, reads count matrix

$ groups: two groups

$ DEs : differential expression genes

Details

Multiple each gene with a value from 5-10 by chance at pvalue of 0.05.

Source

http://bcf.isb-sib.ch/data/compcodeR/

References

Soneson C, Delorenzi M: A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 2013, 14(1):91.

Examples

```
data(simuN5)
```

---

sizeFactors *Accessors for the 'sizeFactor' slot of a ABSDataSet object.*

Description

Accessor functions for the 'sizeFactor' slot of a ABSDataSet object.

Usage

```
## S4 method for signature ABSDataSet
sizeFactors(object)

## S4 replacement method for signature ABSDataSet,numeric
sizeFactors(object)<-value
```

Arguments

object   an ABSDataSet object.

value   a numeric object, one for each sample

Details

The sizeFactors vector assigns to each sample a value, used to normalize the counts in each sample according to selected normMethod.

See Also

normalFactors

Examples

```
data(simuN5)
obj <- ABSDataSet(counts=simuN5$counts, groups=factor(simuN5$groups))
obj <- normalFactors(obj)
sizeFactors(obj)
sizeFactors(obj) <- runif(10,1,2)
sizeFactors(obj)
```

SumInfo *An S4 class that stores a string.*

Description

An S4 class that stores a string.

Index