

# Package ‘dexus’

April 14, 2017

**Type** Package

**Title** DEXUS - Identifying Differential Expression in RNA-Seq Studies with Unknown Conditions or without Replicates

**Description** DEXUS identifies differentially expressed genes in RNA-Seq data under all possible study designs such as studies without replicates, without sample groups, and with unknown conditions. DEXUS works also for known conditions, for example for RNA-Seq data with two or multiple conditions. RNA-Seq read count data can be provided both by the S4 class Count Data Set and by read count matrices. Differentially expressed transcripts can be visualized by heatmaps, in which unknown conditions, replicates, and samples groups are also indicated. This software is fast since the core algorithm is written in C. For very large data sets, a parallel version of DEXUS is provided in this package. DEXUS is a statistical model that is selected in a Bayesian framework by an EM algorithm. DEXUS does not need replicates to detect differentially expressed transcripts, since the replicates (or conditions) are estimated by the EM method for each transcript. The method provides an informative/non-informative value to extract differentially expressed transcripts at a desired significance level or power.

**Version** 1.14.0

**Date** 2016-08-22

**Maintainer** Guenter Klambauer <klambauer@bioinf.jku.at>

**Author** Guenter Klambauer

**License** LGPL (>= 2.0)

**Depends** R (>= 2.15), methods, BiocGenerics

**Suggests** parallel, statmod, stats, DESeq, RColorBrewer

**Collate** 'AllClasses.R' 'AllGenerics.R' 'binomTest.R' 'normalization.R'  
'dexus.R' 'getSizeNB.R' 'functions.R' 'plot-methods.R'  
'show-methods.R' 'methodsAccess.R' 'dexss.R'

**biocViews** Sequencing, RNASeq, GeneExpression, DifferentialExpression, CellBiology, Classification, QualityControl

**NeedsCompilation** yes

## R topics documented:

accessors	2
countsBottomly	3
countsGilad	3
countsLi	4
countsMontgomery	4
countsPickrell	5
dexss	6
dexus	8
dexus.parallel	10
DEXUSResult-class	11
getSizeNB	13
INI	14
INIThreshold<-	15
normalizeData	15
plot	16
sort	17
'['	17
<b>Index</b>	<b>19</b>

---

accessors	<i>Accessors for a "DEXUSResult".</i>
-----------	---------------------------------------

---

### Description

These generic functions return the slots of an RNA-Seq analysis performed by DEXUS. The results of DEXUS are stored as an instance of [DEXUSResult-class](#).

### Arguments

object            An instance of "DEXUSResult".

### Value

The accessor functions return a the matrices or vectors contained in the corresponding slot of the "DEXUSResult".

### Author(s)

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

### Examples

```
data(dexus)
result <- dexus(countsBottomly[1:20,1:10])
transcriptNames(result)
sampleNames(result)
inputData(result)
normalizedData(result)
sizeFactors(result)
INIValues(result)
```

```
INIThreshold(result)
INICalls(result)
pvals(result)
responsibilities(result)
posteriorProbs(result)
logFC(result)
conditionSizes(result)
sizeParameters(result)
means(result)
dispersions(result)
params(result)
```

---

countsBottomly

*RNA-Seq data of two mice strains.*

---

### Description

The two common mice strains C57BL/6J (B6) and DBA/2J (D2) were used for comparing gene expression measures of RNA-Seq and microarrays.

### Usage

```
countsBottomly
```

### Format

A data matrix of 36229 rows (genes) and 21 columns (samples).

### Source

<http://bowtie-bio.sourceforge.net/recount/>

### References

Bottomly, D., Walter, N. A. R., Hunter, J. E., Darakjian, P., Kawane, S., Buck, K. J., Searles, R. P., Mooney, M., McWeeney, S. K., and Hitzemann, R. (2011). *Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays*. Plos One, 6(3), e17820.

---

countsGilad

*RNA-Seq data of humans, chimpanzees and rhesus macaques.*

---

### Description

Liver RNA samples of three males and three females from each of the species human, chimpanzee and rhesus macaques were sequenced.

### Usage

```
countsGilad
```

**Format**

A data matrix of 20689 rows (genes) and 18 columns (samples).

**Source**

[ftp://ftp.ncbi.nlm.nih.gov/pub/geo/DATA/supplementary/series/GSE17274/GSE17274\\_ReadCountPerLane.txt.gz](ftp://ftp.ncbi.nlm.nih.gov/pub/geo/DATA/supplementary/series/GSE17274/GSE17274_ReadCountPerLane.txt.gz)

**References**

Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M., and Gilad, Y. (2010). *Sex-specific and lineage-specific alternative splicing in primates*. *Genome Res*, 20(2), 180-189.

---

countsLi

*RNA-Seq data of the developmental zones of maize leaves.*

---

**Description**

RNA-Sequencing was performed on different locations of the maize plant leaf.

**Usage**

countsLi

**Format**

A data matrix of 110185 rows (genes) and 12 columns (samples).

**Source**

<http://www.ncbi.nlm.nih.gov/sra/> accession number:SRP002265

**References**

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., Kebrom, T. H., Provart, N., Patel, R., Myers, C. R., Reidel, E. J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., and Brutnell, T. P. (2010). *The developmental dynamics of the maize leaf transcriptome*. *Nat Genet*, 42(12), 1060-1067.

---

countsMontgomery

*RNA-Seq data of 60 European HapMap individuals.*

---

**Description**

The RNA of lymphoblastoid cell lines of 60 HapMap individuals was sequenced in order to study eQTLs.

**Usage**

countsMontgomery

**Format**

A data matrix of 12984 rows (genes) and 60 columns (samples).

**Source**

<http://bowtie-bio.sourceforge.net/recount/>

**References**

Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbett, J., Guigo, R., and Dermitzakis, E. T. (2010). *Transcriptome genetics using second generation sequencing in a caucasian population*. *Nature*, 464(7289), 773-777.

---

countsPickrell

*RNA-Seq data of 69 Nigerian HapMap individuals.*

---

**Description**

The RNA of lymphoblastoid cell lines of 69 HapMap individuals was sequenced in order to study eQTLs.

**Usage**

countsPickrell

**Format**

A data matrix of 12984 rows (genes) and 69 columns (samples).

**Source**

<http://bowtie-bio.sourceforge.net/recount/>

**References**

Pickrell, J. K., Marioni, J. C., Pai, A. A., Degner, J. F., Engelhardt, B. E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., and Pritchard, J. K. (2010). *Understanding mechanisms underlying human gene expression variation with RNA sequencing*. *Nature*, 464(7289), 768-772.

**Description**

Performs the DEXSS algorithm for detection of differentially expressed genes in RNA-seq data for a semi-supervised setting, i.e. that the condition of some samples is known, and for some samples the condition is unknown.

**Usage**

```
dexss(X, nclasses = 2, G = 1, alphaInit, cyc = 20,
      labels, normalization = "RLE", kmeansIter = 10,
      ignoreIfAllCountsSmaller = 1, theta = 2.5, minMu = 0.5,
      rmax = 13, initialization = "kmeans",
      multiclassPhiPoolingFunction = NULL, quiet = FALSE,
      resultObject = "S4")
```

**Arguments**

X	either a vector of counts or a raw data matrix, where columns are interpreted as samples and rows as genomic regions. An instance of "countDataSet" is also accepted.
nclasses	The number of conditions, i.e. mixture components. (Default = 2)
G	The weight of the prior distribution of the mixture weights. Not used in the supervised case. (Default = 1).
cyc	Positive integer that sets the number of cycles of the EM algorithm. (Default = 20).
alphaInit	The initial estimates of the condition sizes, i.e., mixture weights. Not used in the supervised case. (Default = c(0.5,0.5)) .
labels	The labels for the classes, will be coerced into an integer. For this semi-supervised version the known labels/conditions must be coded as integers starting with 1. The samples with the label 1 will be considered as being in the "major condition". For the samples with unknown labels/conditions an "NA" must be set.
normalization	method used for normalizing the reads. "RLE" is the method used by (Anders and Huber, 2010), "upperquartile" is the Upper-Quartile method by (Bullard et al., 2010), and none deactivates normalization. (Default = "RLE").
kmeansIter	number of times the K-Means algorithm is run. (Default = 10).
ignoreIfAllCountsSmaller	Ignores transcript for which all read counts are smaller than this value. These transcripts are considered as "not expressed" (Default = 1).
theta	The weight of the prior on the size parameter or inverse dispersion parameter. Theta is adjusted to each transcript by dividing by the mean read count of the transcript. The higher theta, the lower r and the higher the overdispersion will be. (Default = 2.5).
minMu	Minimal mean for all negative binomial distributions. (Default = 0.5).

<code>rmax</code>	Maximal value for the size parameter. The inverse of this parameter is the lower bound on the dispersion. In analogy to (Anders and Huber, 2010) we use 13 as default. (Default = 13).
<code>initialization</code>	Method used to find the initial clusters. Dexus can either use the quantiles of the readcounts of each gene or run k-means on the counts. (Default = "kmeans").
<code>multiclassPhiPoolingFunction</code>	In "multiClass" mode the dispersion is either estimated across all classes at once (NULL), or separately for each condition, i.e., class. The size parameters or dispersion per class are then joined to one estimate by the mean ("mean"), minimum ("min") or maximum ("max"). In our investigations estimation across all classes at once performed best. (Default = NULL).
<code>quiet</code>	Logical that indicates whether dexus should report the steps of the algorithm. Suppresses messages from the program if set to TRUE. (Default = FALSE).
<code>resultObject</code>	Type of the result object; can either be a list ("list") or an instance of "DEXUS-Result" ("S4"). (Default="S4").

## Details

The read count  $x$  is explained by a finite mixture of negative binomials:

$$p(x) = \sum_{i=1}^n \alpha_i \text{NB}(x; \mu_i, r_i),$$

where  $\alpha_i$  is the weight of the mixture component, NB is the negative binomial with mean parameter  $\mu_i$  and size parameter  $r_i$ . The parameters are selected by an EM algorithm in a Bayesian framework.

Each component in the mixture model corresponds to one condition.

- If the groups, conditions, replicate status, or labels are unknown, DEXUS tries to estimate these conditions. For each transcript DEXUS tries to explain the read counts by one negative binomial distribution. If this is possible, the transcript is explained by one condition and therefore it is not differentially expressed. If more than one negative binomial distribution is needed to explain the read counts of a transcript, this transcript indicates that it is differentially expressed. Evidence for differential expression is strong if a large amount of samples participate in each condition and the mean expression values are well separated. Both of these criteria are measured by the informative/non-informative (I/NI) call.
- If there are more than two groups given by the vector `labels`, DEXUS uses a generalized linear model to explain the data in analogy to (McCarthy, 2012).
- If there are two groups given by the vector `labels`, DEXUS uses the exact test for count data to test between the sample groups, as implemented by (Anders and Huber, 2010) in the package "DESeq".

## Value

"list" or "DEXUSResult". A list containing the results and the parameters of the algorithm or an instance of "DEXUSResult".

## Author(s)

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

## References

- Anders, S. and Huber, W. (2010). *Differential expression analysis for sequence count data*. *Genome Biol*, 11(10), R106.
- Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). *Evaluation of statistical methods for normalization and differential expression in mRNA-seq experiments*. *BMC Bioinformatics*, 11, 94.
- McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). *Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation*. *Nucleic Acids Res*, 40(10), 4288-4297.

## Examples

```
data(dexus)
labels1 <- substr(colnames(countsBottomly),1,2)
labels2 <- c()
labels2[which(labels1=="D2")] <- 1
labels2[which(labels1=="B6")] <- 2
labels2[c(3,7,8,10,12,15)] <- NA
res <- dexus(countsBottomly[1:100, ],labels=labels2,nclasses=2,G=0)
```

---

dexus

*Detection of Differential Expression in an Unsupervised Setting*

---

## Description

Performs the DEXUS algorithm for detection of differentially expressed genes in RNA-seq data for a) unknown conditions, b) multiple known conditions, and c) two known conditions.

## Usage

```
dexus(X, nclasses = 2, alphaInit, G = 1, cyc = 20,
      labels = NULL, normalization = "RLE", kmeansIter = 10,
      ignoreIfAllCountsSmaller = 1, theta = 2.5, minMu = 0.5,
      rmax = 13, initialization = "kmeans",
      multiclassPhiPoolingFunction = NULL, quiet = FALSE,
      resultObject = "S4")
```

## Arguments

- |           |   |
|-----------|---|
| X         | either a vector of counts or a raw data matrix, where columns are interpreted as samples and rows as genomic regions. An instance of "countDataSet" is also accepted. |
| nclasses  | The number of conditions, i.e. mixture components. (Default = 2)  |
| alphaInit | The initial estimates of the condition sizes, i.e., mixture weights. Not used in the supervised case. (Default = c(0.5,0.5)) .  |
| G         | The weight of the prior distribution of the mixture weights. Not used in the supervised case. (Default = 1).  |
| cyc       | Positive integer that sets the number of cycles of the EM algorithm. (Default = 20).  |



labels	labels for the classes, will be coerced into a factor by <code>as.factor</code> . Can either be a factor, character or integer. If this vector is given, supervised detection is used. If this vector is set to NULL the unsupervised detection is performed. (Default=NULL).
normalization	method used for normalizing the reads. "RLE" is the method used by (Anders and Huber, 2010), "upperquartile" is the Upper-Quartile method by (Bullard et al., 2010), and none deactivates normalization. (Default = "RLE").
kmeansIter	number of times the K-Means algorithm is run. (Default = 10).
ignoreIfAllCountsSmaller	Ignores transcript for which all read counts are smaller than this value. These transcripts are considered as "not expressed" (Default = 1).
theta	The weight of the prior on the size parameter or inverse dispersion parameter. Theta is adjusted to each transcript by dividing by the mean read count of the transcript. The higher theta, the lower $r$ and the higher the overdispersion will be. (Default = 2.5).
minMu	Minimal mean for all negative binomial distributions. (Default = 0.5).
rmax	Maximal value for the size parameter. The inverse of this parameter is the lower bound on the dispersion. In analogy to (Anders and Huber, 2010) we use 13 as default. (Default = 13).
initialization	Method used to find the initial clusters. Dexus can either use the quantiles of the readcounts of each gene or run k-means on the counts. (Default = "kmeans").
multiclassPhiPoolingFunction	In "multiClass" mode the dispersion is either estimated across all classes at once (NULL), or separately for each condition, i.e., class. The size parameters or dispersion per class are then joined to one estimate by the mean ("mean"), minimum ("min") or maximum ("max"). In our investigations estimation across all classes at once performed best. (Default = NULL).
quiet	Logical that indicates whether dexus should report the steps of the algorithm. Suppresses messages from the program if set to TRUE. (Default = FALSE).
resultObject	Type of the result object; can either be a list ("list") or an instance of "DEXUS-Result" ("S4"). (Default="S4").

## Details

The read count  $x$  is explained by a finite mixture of negative binomials:

$$p(x) = \sum_{i=1}^n \alpha_i \text{NB}(x; \mu_i, r_i),$$

where  $\alpha_i$  is the weight of the mixture component, NB is the negative binomial with mean parameter  $\mu_i$  and size parameter  $r_i$ . The parameters are selected by an EM algorithm in a Bayesian framework.

Each component in the mixture model corresponds to one condition.

- If the groups, conditions, replicate status, or labels are unknown, DEXUS tries to estimate these conditions. For each transcript DEXUS tries to explain the read counts by one negative binomial distribution. If this is possible, the transcript is explained by one condition and therefore it is not differentially expressed. If more than one negative binomial distribution is needed to explain the read counts of a transcript, this transcript indicates that it is differentially expressed. Evidence for differential expression is strong if a large amount of samples participate in each condition and the mean expression values are well separated. Both of these criteria are measured by the informative/non-informative (I/NI) call.

- If there are more than two groups given by the vector labels, DEXUS uses a generalized linear model to explain the data in analogy to (McCarthy, 2012).
- If there are two groups given by the vector labels, DEXUS uses the exact test for count data to test between the sample groups, as implemented by (Anders and Huber, 2010) in the package "DESeq".

### Value

"list" or "DEXUSResult". A list containing the results and the parameters of the algorithm or an instance of "DEXUSResult".

### Author(s)

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

### References

Anders, S. and Huber, W. (2010). *Differential expression analysis for sequence count data*. Genome Biol, 11(10), R106.

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). *Evaluation of statistical methods for normalization and differential expression in mRNA-seq experiments*. BMC Bioinformatics, 11, 94.

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). *Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation*. Nucleic Acids Res, 40(10), 4288-4297.

### Examples

```
data(dexus)
result <- dexus(countsMontgomery[1:10, ])
```

---

dexus.parallel      *A parallel version of DEXUS.*

---

### Description

Speeds up DEXUS by using multiple processors. Uses the parallel package to parallelize a DEXUS call.

### Usage

```
dexus.parallel(X, ncores = 2, normalization = "RLE",
  ignoreIfAllCountsSmaller = 1, resultObject = "S4", ...)
```

### Arguments

X	Either a vector of counts or a raw data matrix, where columns are interpreted as samples and rows as genomic regions.
ncores	The number of cores (CPUs) that will be used by the parallelization.
normalization	Normalization method to be used. (Default="RLE")

ignoreIfAllCountsSmaller A transcript is considered as not expressed if all counts are smaller than the given value. (Default=1)

resultObject Type of the result object; can either be a list ("list") or an instance of "DEXUS-Result" ("S4"). (Default="S4").

... Other options to be passed to dexus().

**Value**

"list"

**Author(s)**

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

**Examples**

```
data(dexus)
result <- dexus.parallel(countsPickrell[1:10, ], ncores=1)
```

---

DEXUSResult-class      *Class* "DEXUSResult"

---

**Description**

This class contains the result of an RNA-Seq data analysis. The class contains the transcript names together with the parameters per condition, i.e., overdispersion and mean. Further it contains informative/non-informative values or p-values.

**Objects from the Class**

Objects can be created by calls of the form `new("DEXUSResult", ...)`.

**Slots**

transcriptNames The names of the transcripts, genes, exons, or regions of interest

sampleNames The sample names as they were given in the input matrix.

inputData The original read count matrix.

normalizedData The normalized read count matrix.

sizeFactors The size factors that were calculated for the normalization. This is that factor that scales each column or sample.

INIValues An informative/non-informative value for each sample that measures the evidence for differential expression.

INIThreshold The threshold for the I/NI values. Transcript with I/NI values above the threshold will be considered as differentially expressed.

INICalls A binary value for each transcript indicating whether it is differentially expressed.

pvals In case of two known conditions or multiple known conditions it is possible to calculate a *p*-value for each transcript. This value is given in this slot.

**responsibilities** A matrix of the size of the input matrix. It indicates the condition for each sample and transcript. The condition named "1" is the major condition. All other conditions are minor conditions. In case of supervised (two known conditions or multiple known conditions) analyses this clustering matrix will be the same for all transcripts.

**posteriorProbs** An array of the dimension of transcripts times samples times conditions. It gives the probability that a certain read count  $x$  was generated under a condition.

**logFC** The log foldchanges between the conditions. The reference is always condition "1".

**conditionSizes** The ratio of samples belonging to that condition. These are the  $\alpha_i$  values of the model.

**sizeParameters** The size parameter estimates for each condition. These are the  $r_i$  values of the model.

**means** The mean of each condition. The  $\mu_i$  values of the model.

**dispersions** The dispersion estimates for each condition. The inverse size parameters.

**params** The input parameters of the DEXUS algorithm.

## Methods

[ Subsetting of a DEXUSResult.

**as.data.frame** Converts the result object into a data frame.

**conditionSizes** Returns the condition sizes or  $\alpha_i$  parameters of the model.

**dispersions** Returns the dispersion, i.e. the inverse size parameters, of the model.

**INI** I/NI filtering of the result object.

**INICalls** Returns a logical value indication whether this transcript is differentially expressed or not.

**INIThreshold** Returns the thresholds for the I/NI values.

**INIThreshold<-** Sets the I/NI threshold. I/NI calls will be changed accordingly.

**INIValues** Returns the I/NI values.

**inputData** Returns the input read counts.

**logFC** Returns the log foldchange with respect to the first condition.

**means** Returns the mean per condition.

**normalizedData** Returns the normalized data.

**params** Returns a list of input parameters of DEXUS.

**plot** Plots a heatmap of the read counts of the top genes.

**posteriorProbs** Returns an array of posterior probabilities.

**pvals** Returns the  $p$ -values per transcript in supervised mode.

**responsibilities** Returns the clustering vector.

**sampleNames** Returns the sample names.

**show** Displays a data frame of results.

**sizeFactors** Returns the size factors used for normalization.

**sizeParameters** Returns the size parameters, i.e. the  $r_i$  values of the model.

**sort** Sorts the result object by I/NI values or  $p$ -values.

**transcriptNames** Returns the transcript names.

**Author(s)**

Guenter Klambauer

**Examples**

```
showClass("DEXUSResult")
```

getSizeNB	<i>Maximum-likelihood and maximum-a-posteriori estimators for the negative binomial distribution.</i>
-----------	---

**Description**

Estimates the size parameter of a a negative binomial distribution from given data.

**Usage**

```
getSizeNB(x, maxCyc = 1000, eta = 0, rmax = Inf,
          method = "bisection")
```

**Arguments**

- x                    The input data. Must be a numeric vector.
- maxCyc             The maximum number of cycles of the numeric procedure to find the estimator. (Default = 1000).
- eta                 The weight of the exponential prior. The higher eta, the lower the estimate for the size parameter. Setting eta = 0 means that the prior is not used and, therefore, the maximum-likelihood estimator is calculated. (Default = 0).
- rmax                Upper bound on the size parameter. This corresponds to a truncated exponential prior. If not used there is a non-zero probability that the estimator for the size parameter is  $\infty$ . (Default = Inf).
- method             The procedure used to solve the equation

$$\sum_{k=1}^N \psi(x_i + r) - N\psi(r) + N \log \left( \frac{r}{r + 1/N \sum_{i=1}^N x_i} \right) - \eta = 0$$

for  $r$ .

This can either be "bisection" or "regula falsi". (Default="bisection").

**Details**

Depending on the parameters you can either obtain the *Maximum-likelihood estimator* or the *maximum-a-posteriori estimator* using an exponential prior.

maximum-likelihood estimator	eta = 0
maximum-a-posteriori estimator	eta > 0

By setting the variable rmax to a positive value one can enforce an upper bound on the parameter.

The inverse of the size parameter is the overdispersion parameter.

**Value**

"numeric" An estimate of the size parameter of the negative binomial distribution. The overdispersion parameter is the inverse of the size parameter of a negative binomial distribution

**Author(s)**

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

**Examples**

```
x <- rnbinom(mu=50, size=5, n=10)
getSizeNB(x)
```

---

INI

*INI filtering of a DEXUS result.*

---

**Description**

This function filters the result object for informative transcripts. Transcripts with an I/NI value below the given threshold are filtered out.

**Arguments**

object	An instance of "DEXUSResult".
threshold	A numeric determining the threshold for the I/NI values.

**Value**

An instance of "DEXUSResult".

**Author(s)**

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

**Examples**

```
data(dexus)
res <- dexus(countsBottomly[1:100, ])
INI(res)
```

---

INIThreshold<-                    *Set the I/NI threshold.*


---

### Description

This generic function sets the threshold of the I/NI value. Transcripts with I/NI values above the I/NI threshold are considered as differentially expressed. The results of DEXUS are stored as an instance of `DEXUSResult-class`.

### Arguments

object                    An instance of "DEXUSResult".  
value                     A numeric to be used for thresholding the I/NI values.

### Value

INIThreshold<- returns an instance of "DEXUSResult".

### Author(s)

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

### Examples

```
data(dexus)
result <- dexus(countsBottomly[1:20,1:10])
INIThreshold(result) <- 0.1
```

---

normalizeData                    *Normalization of RNA-Seq count data.*


---

### Description

Normalizes RNA-seq count data using previously published approaches. Each samples' read counts are corrected by a normalizing factor. The options are "RLE" by (Anders and Huber, 2010), and "upperquartile" by (Bullard et al., 2010).

### Usage

```
normalizeData(X, normalization)
```

### Arguments

X                         data a raw data matrix, where' columns are interpreted as samples and rows as genomic regions.  
normalization            method used for normalizing the reads. RLE is the method used by (Anders and Huber, 2010), upperquartile is the Upper-Quartile method from (Bullard et al., 2010), and none deactivates normalization. (Default = "RLE").

**Value**

"list" A list containing the normalized data (in its "X" component) as well as the size-factors used for the normalization ("sizeFactors").

**Author(s)**

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

**References**

Anders, S. and Huber, W. (2010). *Differential expression analysis for sequence count data*. *Genome Biol*, 11(10), R106.

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). *Evaluation of statistical methods for normalization and differential expression in mRNA-seq experiments*. *BMC Bioinformatics*, 11, 94.

**Examples**

```
data(dexus)
norm <- normalizeData(countsBottomly, "RLE")
```

---

plot

*Visualization of a result of the DEXUS algorithm.*

---

**Description**

Plots a heatmap of the log read counts of the top ranked genes or of selected genes.

**Arguments**

x	An instance of "CNVDetectionResult"
idx	The indices or the transcript names of the transcripts that should be visualized as heatmap.
cexSamples	Size of the column labels, i.e. the samples.
cexGenes	Size of the row labels, i.e. the transcripts.
newColNames	renames the samples.
type	Mark the samples, that do not belong to the major class by crosses ("crosses"), or boxes ("boxes").

**Value**

Generates a heatmap of the expression values of the top-ranked transcripts.

**Author(s)**

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

**Examples**

```
data(dexus)
r <- dexus(countsBottomly[1:100, ])
plot(r)
```



---

sort	<i>Sorting a DEXUS result.</i>
------	--------------------------------

---

**Description**

This function sorts the result object by I/NI values or p-values such that the transcripts with the highest I/NI value or the lowest p-value are ranked first.

**Arguments**

object            An instance of "DEXUSResult".

**Value**

An instance of "DEXUSResult".

**Author(s)**

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

**Examples**

```
data(dexus)
res <- dexus(countsBottomly[1:100, ])
sort(res)
```

---

'['	<i>Subsetting a "DEXUSResult".</i>
-----	------------------------------------

---

**Description**

Information about specific transcripts can be accessed in the "DEXUSResult" object by using the standard brackets "[idx]" for subsetting. Either transcript names or transcript indices can be used.

**Arguments**

x                "DEXUSResult"  
i                Either a numeric vector of indices or a character vector containing the transcript names.

**Value**

An instance of "DEXUSResult".

**Author(s)**

Guenter Klambauer <klambauer@bioinf.jku.at> and Thomas Unterthiner <unterthiner@bioinf.jku.at>

**Examples**

```
data(dexus)
res <- dexus(countsBottomly[1:100, ])
res["ENSMUSG00000000486"]
res[50:55]
```

# Index

- \*Topic **classes**
  - DEXUSResult-class, 11
- \*Topic **datasets**
  - countsBottomly, 3
  - countsGilad, 3
  - countsLi, 4
  - countsMontgomery, 4
  - countsPickrell, 5
- [,DEXUSResult,character-method (DEXUSResult-class), 11
- [,DEXUSResult,logical-method (DEXUSResult-class), 11
- [,DEXUSResult,numeric-method (DEXUSResult-class), 11
- ‘[‘, 17
- ‘[‘,DEXUSResult,character-method (‘[‘), 17
- ‘[‘,DEXUSResult,logical-method (‘[‘), 17
- ‘[‘,DEXUSResult,numeric-method (‘[‘), 17
  
- accessors, 2
- as.data.frame,DEXUSResult-method (DEXUSResult-class), 11
  
- conditionSizes (accessors), 2
- conditionSizes,DEXUSResult-method (DEXUSResult-class), 11
- countsBottomly, 3
- countsGilad, 3
- countsLi, 4
- countsMontgomery, 4
- countsPickrell, 5
  
- dexss, 6
- DEXSS, (dexss), 6
- dexus, 8
- DEXUS, (dexus), 8
- dexus.parallel, 10
- DEXUSResult-class, 11
- dispersions (accessors), 2
- dispersions,DEXUSResult-method (DEXUSResult-class), 11
  
- getSizeNB, 13
  
- INI, 14
- INI,DEXUSResult-method (DEXUSResult-class), 11
- INICalls (accessors), 2
- INICalls,DEXUSResult-method (DEXUSResult-class), 11
- INIThreshold (accessors), 2
- INIThreshold,DEXUSResult-method (DEXUSResult-class), 11
- INIThreshold-set (INIThreshold<-), 15
- INIThreshold<-, 15
- INIThreshold<- ,DEXUSResult-method (DEXUSResult-class), 11
- INIValues (accessors), 2
- INIValues,DEXUSResult-method (DEXUSResult-class), 11
- inputData (accessors), 2
- inputData,DEXUSResult-method (DEXUSResult-class), 11
  
- logFC (accessors), 2
- logFC,DEXUSResult-method (DEXUSResult-class), 11
  
- means (accessors), 2
- means,DEXUSResult-method (DEXUSResult-class), 11
  
- normalizeData, 15
- normalizedData (accessors), 2
- normalizedData,DEXUSResult-method (DEXUSResult-class), 11
  
- params (accessors), 2
- params,DEXUSResult-method (DEXUSResult-class), 11
- plot, 16
- plot,DEXUSResult,missing-method (DEXUSResult-class), 11
- posteriorProbs (accessors), 2
- posteriorProbs,DEXUSResult-method (DEXUSResult-class), 11
- pvals (accessors), 2
- pvals,DEXUSResult-method (DEXUSResult-class), 11

responsibilities (accessors), [2](#)  
responsibilities, DEXUSResult-method  
(DEXUSResult-class), [11](#)

sampleNames (accessors), [2](#)  
sampleNames, DEXUSResult-method  
(DEXUSResult-class), [11](#)  
show, DEXUSResult-method  
(DEXUSResult-class), [11](#)  
sizeFactors (accessors), [2](#)  
sizeFactors, DEXUSResult-method  
(DEXUSResult-class), [11](#)  
sizeParameters (accessors), [2](#)  
sizeParameters, DEXUSResult-method  
(DEXUSResult-class), [11](#)  
sort, [17](#)  
sort, DEXUSResult-method  
(DEXUSResult-class), [11](#)

transcriptNames (accessors), [2](#)  
transcriptNames, DEXUSResult-method  
(DEXUSResult-class), [11](#)