Package 'anota2seq'

November 11, 2026

Version 1.34.0

Title Generally applicable transcriptome-wide analysis of translational efficiency using anota2seq

Author
Christian Oertlin <christian.oertlin@ki.se>, Julie Lorent <julie.lorent@ki.se>, Ola Larsson <ola.larsson@ki.se>

Maintainer Christian Oertlin <christian.oertlin@ki.se>, Ola Larsson <ola.larsson@ki.se>

Description anota2seq provides analysis of translational efficiency and differential expression analysis for polysome-profiling and ribosome-profiling studies (two or more sample classes) quantified by RNA sequencing or DNA-microarray. Polysome-profiling and ribosome-profiling typically generate data for two RNA sources; translated mRNA and total mRNA. Analysis of differential expression is used to estimate changes within each RNA source (i.e. translated mRNA or total mRNA). Analysis of translational efficiency aims to identify changes in translation efficiency leading to altered protein levels that are independent of total mRNA levels (i.e. changes in translated mRNA that are independent of levels of total mRNA) or buffering, a mechanism regulating translational efficiency so that protein levels remain constant despite fluctuating total mRNA levels (i.e. changes in total mRNA that are independent of levels of translated mRNA). anota2seq applies analysis of partial variance and the random variance model to fulfill these tasks.

Depends R (>= 3.4.0), methods

License GPL-3

LazyData yes

LazyLoad yes

Imports multtest,qvalue,limma,DESeq2,edgeR,RColorBrewer, grDevices, graphics, stats, utils, SummarizedExperiment

biocViews ImmunoOncology, GeneExpression, DifferentialExpression, Microarray,GenomeWideAssociation, BatchEffect, Normalization, RNASeq, Sequencing, GeneRegulation, Regression

NeedsCompilation no

Suggests BiocStyle,knitr

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/anota2seq

git_branch RELEASE_3_23

git_last_commit 6b5afaf

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-11

Contents

anota2seq-plots . . . 2
anota2seqAnalyze . . . 5
Anota2seqDataSet-class . . . 8
anota2seqDataSetFromMatrix . . . 9
anota2seqGetContrasts . . . 13
anota2seqGetCovariates . . . 14
anota2seqGetDeltaData . . . 15
anota2seqGetNormalizedData . . . 17
anota2seqGetOutput . . . 18
anota2seqGetQualityControl . . . 21
anota2seqGetResidOutlierTest . . . 23
anota2seqGetThresholds . . . 24
anota2seqPerformQC . . . 25
anota2seqRegModes . . . 27
anota2seqResidOutlierTest . . . 29
anota2seqRun . . . 31
anota2seqSelSigGenes . . . 35
anota2seq_data . . . 38
show,Anota2seqDataSet-method . . . 39

Index **41**

`anota2seq-plots` *Visualization of anota2seq results*

Description

These functions generate graphical outputs aiming to provide an overview of the extents of different regulatory modes of gene expression. The graphical outputs consist of:

1. For all selected contrasts, P-value and FDR density plots indicating the distribution of P-values and FDRs for all analyzes found in the Anota2seqDataSet object.
2. Fold change plots where total mRNA log2FC is compared to translated mRNA (e.g. polysome-associated mRNA or RPF) log2FC (for all selected contrasts).
3. Per identifier and per treatment fitted regression curves between total and translated mRNA for all samples which helps evaluating regulation of single identifiers.

Usage

```
anota2seqPlotFC(object, visualizeRegModes = "all",
 selContrast, contrastName = NULL, fileStem = "ANOTA2SEQ_FoldchangePlot", plotToFile = TRUE, myYlim = N

anota2seqPlotPvalues(object, useRVM = TRUE, selContrast, contrastName = NULL,
  myBw = 0.05, plotToFile = TRUE, fileStem = "ANOTA2SEQ_pvalue_density", ...)

anota2seqPlotGenes(object, selContrast, analysis, geneNames = NULL,
plotToFile = TRUE, fileStem = "ANOTA2SEQ_significantGenes_plot")

## S4 method for signature 'Anota2seqDataSet'
anota2seqPlotFC(object,
 visualizeRegModes = "all", selContrast, contrastName = NULL, fileStem = "ANOTA2SEQ_FoldchangePlot",
  plotToFile = TRUE, myYlim = NULL, myXlim = NULL, ...)

## S4 method for signature 'Anota2seqDataSet'
anota2seqPlotPvalues(object,
  useRVM = TRUE, selContrast, contrastName = NULL, myBw = 0.05, plotToFile = TRUE,
  fileStem = "ANOTA2SEQ_pvalue_density", ...)

## S4 method for signature 'Anota2seqDataSet'
anota2seqPlotGenes(object,
  selContrast, analysis, geneNames = NULL,
  plotToFile = TRUE, fileStem = "ANOTA2SEQ_significantGenes_plot")
```

Arguments

`object`	An object of class Anota2seqDataSet. Should contain output of the anota2seqAnalyze function. To visualize significant identifiers in the fold change plot (`anota2seqPlotFC`) and for regression plots (`anota2seqPlotGenes`), the object should also contain the output of anota2seqSelSigGenes and/or anota2seqRegModes (see Details).
`visualizeRegModes`	Can be set to "all" (default, genes which were selected by `anota2seqSelSigGenes` are colored according to their allocated regulatory mode (change in mRNA abundance or change in translational efficiency leading to altered protein levels or buffering), "none" (no colors), "translation" (only identifiers regulated by change in translational efficiency leading to altered protein levels are colored), "buffering" (only identifiers regulated by change in translational efficiency leading to buffering are colored). The object will be required to contain different outputs depending on the value given to this parameter (see Details).
`selContrast`	Which contrast(s) should be considered? Descriptions of the contrasts can be found in the output from the anota2seqAnalyze object in the usedContrasts slot. Indicate the contrast(s) by a numeric vector of the column number(s).
`contrastName`	Custom name annotation for the selected contrast(s). Provide a character name for each selected contrast, this name will be used as plot title in the anota2seqPlotPvalues plots or as xlab and ylab annotation in the anota2seqPlotFC plots.
`fileStem`	if plotToFile is TRUE, this stem will be added in front of the output filename(s).

plotToFile	Boolean. If set to TRUE (default) the function will output the plot as a PDF file. If set to FALSE plots will be plotted to the R default graphics device.
myXlim	Specify the x-axis limits used for the scatterplots(s) in the anota2seqPlotFC function. Default is NULL.
myYlim	Specify the y-axis limits used for the scattersplot(s) in the anota2seqPlotFC function. Default is NULL
...	Graphical parameters that are passed to the par() function.
useRVM	Should the density plot be performed on RVM p-values/FDR (default) or no-RVM p-values/FDR?
myBw	Smoothing bandwidth parameter for used in the stats::density function. Within one plot, the same bandwidth will be used for all density curves.
analysis	For which analysis should anota2seqPlotGenes be performed. Can be set to "translation" or "buffering".
geneNames	When anota2seqPlotGenes performs the individual gene plots they will be named by the original row names supplied to the anota2seqAnalyze function. geneNames allows the user to add additional names when plotting to e.g. include gene symbols. Input is a matrix with one column where the original row names match the row names of the input matrix and the desired new names are given in column 1. Default is NULL i.e. no additional names.

Details

anota2seqPlotFC: if visualizeRegModes is set to "none", the Anota2seqDataSet object is only required to contains an output of anota2seqAnalyze. If visualizeRegModes is set to "translation" or "buffering", the object should contain the output of anota2seqAnalyze and anota2seqSelSigGenes for the corresponding analysis. Finally, if visualizeRegModes is set to "all" (default), anota2seqRegModes should have been called on the Anota2seqDataSet object so that identifiers can be colored according to the three regulatory modes (i.e. change in mRNA abundance or change in translational efficiency leading to altered protein levels or buffering).

anota2seqPlotGenes: requires an Anota2seqDataSet object containing the output of anota2seqAnalyze and anota2seqSelSigGenes for translation and/or buffering. In the graphical output of anota2seqPlotGenes, the results for each significant gene is displayed on a separate row. The first graph shows all samples and per treatment regression lines using the common slope with different colors for each treatment. The magnitude of the common slope is indicated. The second graph shows key statistics for the identifier without the RVM model for all contrasts analyzed when running anota2seqAnalyze. The third graph is similar to the second but with RVM statistics instead.

Value

No value is returned. These functions generate a graphical outputs as described above.

Examples

```
## Not run:
data(anota2seq_data)
# Initialize the Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(dataP = anota2seq_data_P[1:1000,],
```

```
                                    dataT = anota2seq_data_T[1:1000,],
                                    phenoVec = anota2seq_pheno_vec,
                                    dataType = "RNAseq",
                                    normalize = TRUE)
# Perform anota2seqRun function (performQC is set to FALSE here to limit running
# time of this example but the model assumptions should be assessed (see help of
# anota2seqPerformQC))
Anota2seqDataSet <- anota2seqRun(Anota2seqDataSet,
                                 performQC = FALSE,
                                 performROT = FALSE)
# Visualize results
anota2seqPlotPvalues(Anota2seqDataSet, plotToFile = FALSE, selContrast = 1)
anota2seqPlotFC(Anota2seqDataSet, plotToFile = FALSE, selContrast = 1)
anota2seqPlotGenes(Anota2seqDataSet, plotToFile = TRUE, analysis = "translation",
           selContrast = 1)

## End(Not run)
```

---

anota2seqAnalyze *Analysis of translational activity*

Description

Analysis of treatment effects for translated mRNA (polysome-associated mRNA or RPFs; i.e. the "P" samples in the SummarizedExperiment annotation or the samples in the dataP matrix) or total mRNA (i.e. the "T" samples in the SummarizedExperiment annotation or the samples in the dataT matrix); and/or identification of changes in translational efficiency leading to altered protein levels or buffering.

Usage

```
anota2seqAnalyze(Anota2seqDataSet, contrasts = NULL,
  correctionMethod = "BH", useProgBar = TRUE, fileStem = "ANOTA2SEQ",
  analysis = c("translation", "buffering", "translated mRNA", "total mRNA"))
```

Arguments

Anota2seqDataSet	An object of class Anota2seqDataSet
contrasts	If NULL (default), the contrasts will be created automatically and may not correspond to those of interest. It is therefore possible to use custom contrasts using this parameter. The input should be a matrix with row names corresponding to treatments (i.e. those in phenoVec or the treatment column of the SummarizedExperiment annotation) and columns corresponding to the different contrasts of interest (see Details section and examples for additional information on how to set up a custom contrast matrix).

`correctionMethod`
: Correction for multiple testing method. This parameter can be set to "Bonferroni", "Holm", "Hochberg", "SidakSS", "SidakSD", "BH", "BY", "ABH" or "TSBH" as implemented in the multtest package or "qvalue" as implemented in the qvalue package. Default is "BH".

`useProgBar`
: Should the progress bar be shown? Default is TRUE, show progress bar.

`fileStem`
: Stem for output file name (see Value section). Default is "ANOTA2SEQ".

`analysis`
: Specification of which analysis should be performed. Can be set to "translated mRNA","total mRNA", "translation", "buffering" or any combination of these. If set to translated mRNA or total mRNA, the function will perform differential expression analysis (based on default or supplied contrasts) for the specified mRNA source. If set to "translation", analysis of changes in translational efficiency leading to altered protein levels will be performed. If set to "buffering", analysis of changes in translational efficiency leading to buffering will be performed. Default is c("translated mRNA", "total mRNA", "translation", "buffering") i.e. all analyzes are performed.

Details

The function performs APV on two or more sample categories according to default contrasts or contrasts supplied by the user. The Random Variance Model (RVM) will be applied but the results without RVM will also be available in the output.

More details about setting a contrast matrix: By default (i.e. with contrasts = NULL), the order of the sample classes which are used to calculate differences between treatments will be in alphabetical order. To change the directionality of the contrasts (e.g. treatment b vs treatment a instead of treatment a vs treatment b) or to generate a custom set of contrasts when more than 2 treatments are included, a contrast matrix can be supplied to the "contrasts" parameter described above. The row names should be specified as indicated above. The contrasts are coded by using e.g. -1 for group a, 0 for group b and 1 for group c to compare group a and c; -2 for group a, 1 for group b and 1 for group c to compare group a to b & c. Each column of the contrast matrix should sum to 0 and to analyze orthagonal contrasts the products of all pairwise rows should sum to 0. The results in the Anota2seqDataSet object will follow the order of the contrasts (i.e. results for e.g. contrast 1 will correspond to the contrasts specified in column 1 of the contrast matrix).

Value

An Anota2seqDataSet. anota2seqAnalyze saves its output data in the 'translatedmRNA', 'totalmRNA', 'translation' and 'buffering' slots of the Anota2seqDataSet output object when analysis is set to "translated mRNA", "total mRNA", "translation" and "buffering" respectively. This output contains statistics from the applied APV model and can be accessed using `anota2seqGetOutput` (See ?`anota2seqGetOutput` for a detailed description). Additionally, `anota2seqAnalyze` saves fold changes and translational efficiency (TE) scores in the 'deltaData' slot of the Anota2seqDataSet object. (See ?`anota2seqGetDeltaData` for a detailed description of deltaData).

`anota2seqAnalyze` creates a plot showing the fit of the inverse gamma distribution used in RVM (i.e. a similar output as compared to that from anota2seqPerformQC) for each contrast.

See Also

anota2seqGetOutput, anota2seqGetDeltaData

Examples

```
data(anota2seq_data)
# Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Run analysis of changes in translational efficiency leading to
# altered protein levels and differential expression analysis on
# translated mRNA
## Not run:
Anota2seqDataSet <- anota2seqAnalyze(
    Anota2seqDataSet,
    analysis = c("translation"))

## End(Not run)
# Example to build a custom contrast matrix
# For the purpose of this example, we will use the first 6 samples of the
# simulated data provided with the package together with the following "dummy"
# sample classes:
phenoVec <- c("a","a","b","b","c","c")
contrastsEx_ads <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100, 1:6],
    dataT = anota2seq_data_T[1:100, 1:6],
    phenoVec = phenoVec,
    dataType = "RNAseq",
    normalize = TRUE)

# Get the levels of the phenoVec, these will be ordered as in anota2seq
phenoLev <- levels(as.factor(phenoVec))
# Construct the matrix with appropriate nrow and ncol
myContrast <- matrix(nrow =length(phenoLev),ncol=length(phenoLev)-1)
# Set the phenoLev as rownames for your contrast matrix
rownames(myContrast) <- phenoLev
# Now indicate the contrasts you want to analyse as explained above
# Compare a to c
myContrast[,1] <- c(-1,0,1)
# Compare a to b& c
myContrast[,2] <- c(2,-1,-1)
myContrast
#   [,1] [,2]
# a   -1    2
# b    0   -1
# c    1   -1
# The custom contrast matrix can then be used as input of anota2seqAnalyze.
contrastsEx_ads <- anota2seqAnalyze(contrastsEx_ads,
```

```
                                    analysis = "translation",
                                    contrasts = myContrast)
```

`Anota2seqDataSet-class` *Anota2seqDataSet class*

Description

A S4 class to store and collect output from the anota2seq workflow.

Details

The Anota2seqDataSet will be used to collect outputs of the `anota2seqPerformQC`, `anota2seqResidOutlierTest`,`anota2se` `anota2seqSelSigGenes` and `anota2seqRegModes` functions. The results that are collected in the Anota2seqDataSet can be accessed using the anota2seq.get methods. Furthermore show() gives an overview of the stored objects in the Anota2seqDataSet.

Slots

`dataT` a matrix containing normalized data for total mRNA. Can be accessed using `anota2seqGetNormalizedData`.

`dataP` a matrix containing normalized data for translated mRNA. Can be accessed using `anota2seqGetNormalizedData`.

`phenoVec` a vector describing the sample class for each column of dataT and dataP. Can be accessed using `anota2seqGetCovariates`.

`batchVec` an optional vector describing a batch class for each column of dataT and dataP. Can be accessed using `anota2seqGetCovariates`.

`contrasts` a matrix describing the contrasts to be used in the analysis. Can be accessed using `anota2seqGetContrasts`.

`qualityControl` after running `anota2seqPerformQC`, its output will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetQualityControl`.

`residOutlierTest` after running `anota2seqResidOutlierTest`, its output will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetResidOutlierTest`.

`translatedmRNA` after running `anota2seqAnalyze` with analysis parameter set to "translated mRNA" (or `anota2seqRun`), its output (results of the differential expression analysis) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "translated mRNA" and output = "full".

`totalmRNA` after running `anota2seqAnalyze` with analysis parameter set to "total mRNA" (or `anota2seqRun`), its output (results of the differential expression analysis) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "total mRNA" and output = "full".

`translation` after running `anota2seqAnalyze` with analysis parameter set to "translation" (or `anota2seqRun`), its output (results of the analysis of changes in translational efficiency leading to altered protein levels) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "translation" and output = "full".

`buffering` after running `anota2seqAnalyze` with analysis parameter set to "buffering" (or `anota2seqRun`), its output (results of the analysis of changes in translational efficiency leading to buffering) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "buffering" and output = "full".

`selectedTranslatedmRNA` after running `anota2seqSelSigGenes` with analysis parameter set to "translated mRNA" (or `anota2seqRun`), its output (results of the differential expression analysis on selected genes based on user defined thresholds in `anota2seqSelSigGenes`) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "translated mRNA" and output = "selected".

`selectedTotalmRNA` after running `anota2seqSelSigGenes` with analysis parameter set to "total mRNA" (or `anota2seqRun`), its output (results of the differential expression analysis on selected genes based on user defined thresholds in `anota2seqSelSigGenes`) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "total mRNA" and output = "selected".

`selectedTranslation` after running `anota2seqSelSigGenes` with analysis parameter set to "translation" (or `anota2seqRun`), its output (results of the analysis of changes in translational efficiency leading to altered protein levels on selected genes) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "translation" and output = "selected".

`selectedBuffering` after running `anota2seqSelSigGenes` with analysis parameter set to "buffering" (or `anota2seqRun`), its output (results of the analysis of changes in translational efficiency leading to buffering on selected genes) will be saved in this slot of the Anota2seqDataSet object. Can be accessed using `anota2seqGetOutput` with analysis = "buffering" and output = "selected".

`mRNAAbundance` after running `anota2seqRegModes` (or `anota2seqRun`), its output will be saved in this slot of the Anota2seqDataSet object. This output will contains results of the differential expression analysis on genes selected to be considered as regulated by mRNA abundance. The definition of regulation by mRNA abundance can be based on results from total mRNA, translated mRNA or both (this is set by the user on parameter mRNASelect in anota2seqRegModes). Statistics results from mRNA type(s) considered in this definition will be saved in this slot. Can be accessed using `anota2seqGetOutput` with analysis = "mRNA abundance" and output = "selected".

`deltaData` after running `anota2seqAnalyze`, an additional output called deltaData will be stored in the Anota2seqDataSet. deltaData is a list of matrices with one deltaData matrix for each contrast. For each matrix the rows correspond to the mRNA in the dataset and has the following columns: deltaP, deltaT, deltaPT and deltaTP. deltaP is translated mRNA fold change, deltaT is the total mRNA fold change, deltaPT is the within contrast difference difference of the log ratios with translated mRNA in the nominator and total mRNA in the denominator (deltaPT is commonly called TE score = Translational efficiency score) and deltaTP is the within contrast difference of the log ratios with total mRNA in the nominator and translated mRNA in the denominator.

`anota2seqDataSetFromMatrix`

Anota2seqDataSet constructors.

Description

Functions used to create an Anota2seqDataSet S4 object from user input. This object will be used to collect data and results from all steps of the anota2seq workflow and is initiated using one of the 2 available constructors: `anota2seqDataSetFromMatrix` (use when the input data is provided as a set of custom vectors and matrices) or `anota2seqDataSetFromSE` (use when the input is of the class SummarizedExperiment).

Usage

```
anota2seqDataSetFromMatrix(dataP, dataT, phenoVec, batchVec = NULL, dataType,
  normalize = FALSE, transformation = "TMM-log2",
  filterZeroGenes = ifelse(dataType == "RNAseq" & normalize == TRUE, TRUE, FALSE),
  varCutOff = NULL)

anota2seqDataSetFromSE(se, assayNum = 1, dataType, normalize = FALSE,
  transformation = "TMM-log2",
  filterZeroGenes = ifelse(dataType == "RNAseq" & normalize == TRUE, TRUE, FALSE),
  varCutOff = NULL)
```

Arguments

`dataP`	This parameter is used if selecting to initiate the Anota2seqDataSet using custom vectors and matrices, and the `anota2seqDataSetFromMatrix` function. A matrix containing data for translated mRNA (e.g. polysome-associated mRNA or RPF). Rows must correspond to identifiers (only non-numerical row names are allowed) and columns to samples.
`dataT`	This parameter is used if selecting to initiate the Anota2seqDataSet using custom vectors and matrices, and the `anota2seqDataSetFromMatrix` function. A matrix containing data for total mRNA. Rows must correspond to identifiers (only non-numerical row names are allowed) and columns to samples.
`phenoVec`	This parameter is used if selecting to initiate the Anota2seqDataSet using custom vectors and matrices, and the `anota2seqDataSetFromMatrix` function. A vector describing the treatments (each treatment should have a unique identifier). Note that dataT, dataP and phenoVec have to have the same sample order such that e.g. column 1 in dataP is the translated mRNA for a sample, column 1 in dataT is the total mRNA data and position 1 in phenoVec describes the sample treatment.
`batchVec`	This parameter is used if selecting to initiate the Anota2seqDataSet using custom vectors and matrices, and the `anota2seqDataSetFromMatrix` function. A vector describing annotation of the samples according to their batch (each batch should have a unique identifier). Note that dataT, dataP and batchVec have to have the same sample order such that e.g. column 1 in dataP is the polysome association data for a sample, column 1 in dataT is the total mRNA data and position 1 in batchVec describes the batch identity.
`dataType`	This parameter is used when selecting to initiate the Anota2seqDataSet using the `anota2seqDataSetFromMatrix` or the `anota2seqDataSetFromSE` functions.

Specify the platform on which data were acquired. Can be set to either "microarray" (i.e. already on a continuous log transformed scale) or "RNAseq" (i.e. count data). For pre-normalized RNAseq data that is on a continuous log scale, "microarray" or "RNAseq" in combination with setting the parameter "normalize" to FALSE could be used.

`normalize` This parameter is used when selecting to initiate the Anota2seqDataSet using the `anota2seqDataSetFromMatrix` or the `anota2seqDataSetFromSE` functions. Boolean (TRUE/FALSE) that defaults to TRUE. If TRUE, RNAseq data (or other count data) will be normalized and transformed according to the specified transformation. Microarray data should be normalized by the user before using it as input of `anota2seqDataSetFromMatrix` or `anota2seqDataSetFromSE`

`transformation` This parameter is used when selecting to initiate the Anota2seqDataSet using the `anota2seqDataSetFromMatrix` or the `anota2seqDataSetFromSE` functions. Selection of method for normalization. Must be a vector containing "rlog" or "TMM-log2" that is considered only when dataType = "RNAseq" and normalize = TRUE. The default is "TMM-log2". When using "TMM-log2", RNAseq data will be normalized using the TMM normalization prior to log2 counts per million computation using the voom function of the limma package.

`filterZeroGenes`

This parameter is used when selecting to initiate the Anota2seqDataSet using the `anota2seqDataSetFromMatrix` or the `anota2seqDataSetFromSE` functions. Boolean (TRUE/FALSE); if set to TRUE, genes with 0 counts in at least 1 sample will be removed prior to normalization.

`varCutOff` This parameter is a numeric value (or NULL) used when selecting to initiate the Anota2seqDataSet using the `anota2seqDataSetFromMatrix` or the `anota2seqDataSetFromSE` functions. This parameter indicates if and by which threshold variance filtering should be applied. The default is NULL, i.e. no filtering based on variance. If a cut off is applied, filtering will be performed by applying the threshold to the result of the var() function. Filtering is performed per mRNA source (i.e. translated mRNA and total mRNA) and treatment. This parameter can be used to avoid a rare error during anota2seq analysis (see details).

`se` This parameter is used if selecting to initiate the Anota2seqDataSet using a SummarizedExperiment object with the `anota2seqDataSetFromSE` function. Within the SummarizedExperiment object the expression data is supplied as one assay containing data for both translated mRNA (e.g. polysome-associated mRNA or RPFs) and total mRNA (rows correspond to identifiers and columns to samples). The annotation needed is supplied within "colData" of the SummarizedExperiment object (rows correspond to samples with identical names as in the assay while columns correspond to various annotation). The "colData" must contain the following annotation columns with their names within quotes:

- "RNA": under this column each sample must be annotated with one out of two RNA source identifiers: "P" indicates that the sample was obtained from translated mRNA (e.g. polysome-associated mRNA or RPFs) whereas "T" indicates that the sample was obtained from total mRNA.
- "treatment": under this column the treatment for each sample is indicated. Samples with the same treatment must have identical identifiers.

- "samplePairs": under this column the sample pair identity is indicated. This serves to identify pairs of data for translated mRNA (i.e. "P" under the "RNA" column) and total mRNA (i.e. "T" under the "RNA" column) that were derived from the same starting sample. Each pair of "P" and "T" must have a common identifier that that is unique for that pair (i.e. is not used by any other pair of "P" and "T"). This column will also be used to order columns of translated and total RNA data.
- "batch": under this optional column the batch identity of each sample is indicated (depending on whether the downstream analysis will include a batch parameter or not). A common batch used in downstream analysis is replicate but any other batch that does not overlap with analyzed treatments can be used. Each batch must be indicated by a unique identifier (i.e. not used by any other batch).

`assayNum` This parameter is used if selecting to initiate the Anota2seqDataSet using a SummarizedExperiment object and the `anota2seqDataSetFromSE` function and should specify the assay position (retrieved by "assays(se)") containing the expression data for analysis. By default, the first assay will be used.

Details

These functions initiate an Anota2seqDataSet and provide possibilities to filter, transform and normalize the data. The input can be either of the SummarizedExperiment class including the annotation as outlined above or as a set of matrices of vectors that together contain the same information.

If raw RNAseq data (or other count data) is provided, gene filtering for genes with 0 counts in at least one sample (optional) can be performed followed by normalization and transformation. Transformation algorithms that are available are rlog (DESeq2 package) and TMM-log2 (TMM normalization using the edgeR package followed by log2 counts per million computation using the voom function of the limma package). The relative performance of these methods have been described elsewhere.

A rare error can occur when data within translated mRNA (polysome-associated mRNA or RPF) or total mRNA data from any gene and any treatment has no variance. Users can use the varCutOff parameter to perform filtering based on variance per mRNA source (i.e. polysome- associated mRNA (RPFs) or total mRNA) and treatment. This will eliminate this error which is due to that statistics cannot be calculated in the absence of variance.

Value

an Anota2seqDataSet containing data and covariates ready for analysis using `anota2seqAnalyze` or `anota2seqRun`.

Examples

```
data(anota2seq_data)
Anota2seqDataSet <- anota2seqDataSetFromMatrix(dataP = anota2seq_data_P[1:500,],
                                    dataT = anota2seq_data_T[1:500,],
                                    phenoVec = anota2seq_pheno_vec,
                                    dataType = "RNAseq",
                                    normalize = TRUE)
```

anota2seqGetContrasts *Accessor for the 'contrasts' slot of an Anota2seqDataSet object.*

Description

Retrieve from an Anota2seqDataSet the contrast matrix that is used in the analysis

Usage

```
anota2seqGetContrasts(object)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetContrasts(object)
```

Arguments

object	A Anota2seqDataSet.

Value

A matrix containing information about the analyzed contrasts to be used for the analysis.

See Also

Anota2seqDataSet-class

Examples

```
data(anota2seq_data)
# Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Run analysis of differential translation
Anota2seqDataSet <- anota2seqAnalyze(Anota2seqDataSet,
                                     analysis = "translation")
contr <- anota2seqGetContrasts(Anota2seqDataSet)
contr
#           treatment
# ctrl             -1
# treatment         1
```

anota2seqGetCovariates *Accessor for the 'phenoVec' and 'batchVec' slots of an Anota2seqDataSet object.*

Description

Retrieve from an Anota2seqDataSet the covariate(s) to be used in the APV model i.e. the annotation of the samples according to their treatments and batch (when provided).

Usage

```
anota2seqGetCovariates(object)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetCovariates(object)
```

Arguments

object	An Anota2seqDataSet.

Value

A list of 2 elements named "phenoVec" and "batchVec" containing the corresponding covariates to be used for the analysis.

See Also

Anota2seqDataSet-class

Examples

```
data(anota2seq_data)
# Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    batchVec = c(1, 2, 3, 4, 1, 2, 3, 4),
    dataType = "RNAseq",
    normalize = TRUE)
# Run analysis of differential translation
Anota2seqDataSet <- anota2seqAnalyze(Anota2seqDataSet,
                                     analysis = "translation")
covariates <- anota2seqGetCovariates(Anota2seqDataSet)
covariates
# $phenoVec
# [1] "ctrl"      "ctrl"      "ctrl"      "ctrl"      "treatment" "treatment"
# [7] "treatment" "treatment"
```

```
#
# $batchVec
# [1] "1" "2" "3" "4" "1" "2" "3" "4"
```

`anota2seqGetDeltaData` *Accessor for the fold changes and TE scores stored in the 'deltaData' slot of an Anota2seqDataSet object.*

Description

anota2seqAnalyze, anota2seqSelSigGenes and anota2seqRun calculate the fold change values of translated mRNA (e.g. polysome-associated mRNA or RPFs) and total mRNA. They also provide deltaPT which is the mean log2(translated mRNA data / total mRNA data) between treatments difference (also referred to as translational efficiency (TE) scores) and similarly deltaTP log2(total mRNA data/translated mRNA data). anota2seqGetDeltaData allows access these delta data that can be useful for plotting or custom filtering.

Usage

```
anota2seqGetDeltaData(object, output, analysis, selContrast)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetDeltaData(object, output, analysis,
  selContrast)
```

Arguments

`object`	An Anota2seqDataSet. Should contain the output of the anota2seqAnalyze function (and of anota2seqSelSigGenes if the output parameter is set to "selected", see below).
`output`	A character string specifying whether the deltaData output of the anota2seqAnalyze ("full", this will output deltaData for all genes) or of the anota2seqSelSigGenes ("selected", this will output deltaData for selected genes only) function should be returned.
`analysis`	A character to specify for which analysis you want to retrieve the deltaData. Can be "translated mRNA", "total mRNA", "translation" or "buffering". Depending on the analysis parameter the returned deltaData will be influenced, i.e. when analysis is set to "total mRNA" a deltaT column will be returned, when analysis is set to "translated mRNA" a deltaP column will be returned; if analysis is set to "translation" a matrix with deltaP and deltaPT will be returned, if analysis is set to "buffering" a matrix with deltaT and deltaTP will be returned.
`selContrast`	A numeric vector specifying for which contrast the output should be retrieved. The contrast number corresponds to the position of the column in the automatically generated or specified contrast matrix.

Value

A matrix containing rows for all identifiers for the specified analysis and contrast. The matrix contains the following columns depending on analysis:

- deltaP: log2 fold change of translated mRNA across treatments (e.g. polysome-associated mRNA or RPF)
- deltaT: log2 fold changes of total mRNA across treatments,
- deltaPT: mean log2(translated mRNA data / total mRNA data) between treatments difference
- deltaTP: mean log2(total mRNA data / translated mRNA data) between treatments difference

When analysis is "translated mRNA" the matrix contains deltaP. When analysis is "total mRNA" the matrix contains deltaT. When analysis is "translation" the matrix contains deltaP and deltaPT. When analysis is "buffering" the matrix contains deltaT and deltaTP.

See Also

anota2seqAnalyze, anota2seqSelSigGenes

Examples

```
data(anota2seq_data)
# Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Run analysis of differential translation
Anota2seqDataSet <- anota2seqAnalyze(Anota2seqDataSet,
                                     analysis = "translation")
# Run anota2seqSelSigGenes
Anota2seqDataSet <- anota2seqSelSigGenes(Anota2seqDataSet,
                                         analysis="translation",
                                         selContrast = 1,
                                         maxPAdj = .15)

# Get delta data
delta_translation <- anota2seqGetDeltaData(Anota2seqDataSet,
                                           output = "selected",
                                           analysis = "translation",
                                           selContrast = 1)
```

anota2seqGetNormalizedData *Accessor for the 'normalizedDataP' and 'normalizedDataT' slots of an Anota2seqDataSet object.*

Description

Retrieve the data that is used in the analysis (i.e. if raw RNAseq counts were supplied, this function retrieves the normalized counts used in the analysis).

Usage

```
anota2seqGetNormalizedData(object)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetNormalizedData(object)
```

Arguments

object	A Anota2seqDataSet.

Value

A list of 2 elements named "dataP" and "dataT" containing the corresponding normalized data to be used for the analysis. Note that dataT and dataP have the same sample order such that e.g. column 1 in dataP is the translated mRNA data for a sample, column 1 in dataT is the total mRNA data.

See Also

Anota2seqDataSet-class

Examples

```
data(anota2seq_data)
# Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Get normalized dataP
normDataP <- anota2seqGetNormalizedData(Anota2seqDataSet)[["dataP"]]
head(normDataP)
```

anota2seqGetOutput *Retrieve output for analyzes stored in the Anota2seqDataSet*

Description

The function is used to access the outputs from the anota2seqAnalyze, anota2seqSelSigGenes and anota2seqRegModes stored in the Anota2seqDataSet

Usage

```
anota2seqGetOutput(object, analysis, output, selContrast, getRVM = TRUE)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetOutput(object, analysis, output,
  selContrast, getRVM = TRUE)
```

Arguments

object	An Anota2seqDataSet
analysis	A character string that can be "translated mRNA", "total mRNA", "translation", "buffering" or "mRNA abundance". This parameter is not needed when the output parameter is set to "singleDf".
output	A character string specifying whether the output of the anota2seqAnalyze ("full") or the anota2seqSelSigGenes ("selected") function should be returned. This parameter can also be set to "regModes" in which case the output of anota2seqRegModes is returned (i.e. identifiers which were given as single regulatory mode).Lastly, the parameter can be set to "singleDf" in this case the function returns a single data frame with apvEff and apvRvmPadj values for translated mRNA, total mRNA, translation and buffering. Additionally,there is a column "singleRegMode" indicating the regulatory mode the identifier belongs to and corresponds to the visualization with the anota2seqPlotFC function.
selContrast	A numeric vector specifying for which contrast the output should be retrieved. The contrast number corresponds to the position of the column in the automatically generated or specified contrast matrix.
getRVM	TRUE or FALSE indicating whether the RVM output be retrieved. Default is TRUE. NOTE: the output of anota2seqAnalyze (accessed with output = "full") is available both with RVM (getRVM = TRUE) and without RVM (getRVM = FALSE). However, when RVM was set to TRUE, respectively FALSE, in the anota2seqRun or anota2seqSelSigGenes functions, all subsequent filtering is done on the RVM, respectively no-RVM, output.

Details

This function is used to access outputs from the anota2seqRun, anota2seqAnalyze, anota2seqSelSigGenes and anota2seqRegModes functions stored within the Anota2seqDataSet. The output of the anota2seqAnalyze function holds analysis results on all identifiers (slope, p-value of test on slope

value, residual error, group effect, mean square error, F-value, residual degrees of freedom, p-value, adjusted p-value) and is accessed with output = "full". The output of the anota2seqSelSigGenes function is a reduced set based on filtering criteria and is accessed using output = "selected". In the case of accessing the output of anota2seqRegModes (output = "regModes"), the output will further be reduced by removal of the overlap between regulatory modes so that each gene has a unique regulatory mode. For this, we prioritize regulatory patterns so that translation > mRNA abundance > buffering (i.e. if an identifier is found as changing its translational efficiency leading to altered protein levels and mRNA abundance it will be removed from the mRNA abundance group). Lastly, to access a single data frame with merged apvEff and apvRvmPAdj results only from the analysis of translated mRNA, total mRNA, translation and buffering with indicated regulatory mode for each identifier the parameter can be set to output = "singleDf".

Value

Returns a data.frame with the following columns:

If getRVM is FALSE: A data.frame with statistics from the applied APV for that contrast. Columns are:

- apvSlope: the common slope used in APV;
- apvSlopeP: for translation if the slope is <0 or >1 a p-value for the slope being <0 or >1 is calculated; if the slope is >=0 & <=1 this value is set to 1. For buffering if the slope is <-1 or >0 a p-value for the slope being <-1 or >0 is calculated; if the slope is >=-1 & <=0 this value is set to 1;
- unadjustedResidError: (the residual error before calculating the effective residual error);
- apvEff: the group effect; log2FC. Please note that when analyzing changes in translational efficiency leading to buffering (and only for this analysis), an identifier with positive apvEff will be interpreted as being buffered down and a negative apvEff buffered up (see vignette for more details);
- apvMSerror: the effective mean square error;
- apvF: the F-value;
- residDf: the residual degrees of freedom;
- apvP: the p-value;
- apvPAdj: the adjusted p-value.

if getRVM is TRUE: A data.frame with RVM statistics from the applied APV. Columns are:

- apvSlope: the common slope used in APV;
- apvSlopeP: for translation if the slope is <0 or >1 a p-value for the slope being <0 or >1 is calculated; if the slope is >=0 & <=1 this value is set to 1. For buffering if the slope is <-1 or >0 a p-value for the slope being <-1 or >0 is calculated; if the slope is >=-1 & <=0 this value is set to 1;
- apvEff: the group effect; log2FC; Please note that when analyzing changes in translational efficiency leading to buffering (and only for this analysis), an identifier with positive apvEff will be interpreted as being buffered down and a negative apvEff buffered up (see vignette for more details)
- apvRvmMSerror: the effective mean square error after RVM;

- apvRvmF: the RVM F-value;
- residRvmDf: the residual degrees of freedom after RVM;
- apvRvmP: the RVM p-value;
- apvRvmPAdj: the adjusted RVM p-value.

The output of anota2seqSelSigGenes contains a matrix with the same columns and with a subset of rows corresponding to filtered identifiers.

When anota2seqRegModes has been run on the object, an additional column named singleRegMode is added indicating the allocated regulatory mode of each identifier.

While in the output of anota2seqSelSigGenes (accessed with output = "selected") a same identifier can be selected in several regulatory modes (e.g. both translation up and mRNA abundance up), the output of anota2seqRegModes (accessed with output = "regModes") shows a single regulatory mode allocated to each identifier according to the priority rule explained in the Details section.

See Also

anota2seqRun, anota2seqAnalyze, anota2seqSelSigGenes, anota2seqRegModes

Examples

```
# Initialize the Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Perform analysis of changes in translational efficiency leading to altered
# protein levels
Anota2seqDataSet <- anota2seqAnalyze(Anota2seqDataSet, analysis = "translation",
                                     useProgBar = FALSE)

# Filter
Anota2seqDataSet <- anota2seqSelSigGenes(Anota2seqDataSet, selContrast = 1,
                                         maxPAdj = .2, analysis = "translation")

# Get output for analysis of changes in translational efficiency leading to
# altered protein levels
translationResults <- anota2seqGetOutput(Anota2seqDataSet,
                                 output= "full",
                                 analysis="translation",
                                 selContrast = 1,
                                 getRVM = TRUE)

# Get the selected output for analysis of changes in translational efficiency
# leading to altered protein levels
translationResultsSig <- anota2seqGetOutput(Anota2seqDataSet,
                                 output= "selected",
                                 analysis="translation",
                                 selContrast = 1,
```

```
                                    getRVM = TRUE)
```

`anota2seqGetQualityControl` *Retrieve the anota2seqPerformQC output from the Anota2seqDataSet*

Description

Retrieves the anota2seqPerformQC output from the Anota2seqDataSet

Usage

```
anota2seqGetQualityControl(object)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetQualityControl(object)
```

Arguments

`object`	An Anota2seqDataSet.

Value

The function outputs a list containing the following data:

- omniIntStats: A matrix with a summary of the statistics from the omnibus interaction analysis containing the following columns:
  - intMS: the mean square for the interaction;
  - intDf: the degrees of freedom for the interaction;
  - residMS: the residual error mean square;
  - residDf: the degrees of freedom for the residual error;
  - residMSRvm: the mean square for the residual error after applying RVM;
  - residDfRvm: the degrees of freedom for the residual error after applying RVM;
  - intRvmFval: the F-value for the RVM statistics;
  - intP: the p-value for the interaction;
  - intRvmP: the p-value for the interaction using RVM statistics;
  - intPAdj: the adjusted [for multiple testing using the selected multiple testing correction method] p-value of the interaction;
  - intRvmPAdj: the adjusted [for multiple testing using the selected multiple testing correction method] p-value of the interaction using RVM statistics).
- omniGroupStats: A matrix with a summary of the statistics from the omnibus treatment analysis containing the following columns:
  - groupSlope: the common slope used in APV;

  - groupSlopeP: if the slope is <0 or >1 a p-value for the slope being <0 or >1 is calculated; if the slope is >=0 & <=1 this value is set to 1;
  - groupMS: the mean square for treatments;
  - groupDf: the degrees of freedom for the treatments;
  - groupResidMS: the residual error mean square);
  - groupResidDf: the degrees of freedom for the residual error;
  - residMSRvm: the mean square for the residual error after applying RVM;
  - groupResidDfRvm: the degrees of freedom for the residual error after applying RVM;
  - groupRvmFval: the F-value for the RVM statistics;
  - groupP: the p-value for the sample class effect;
  - groupRvmP: the p-value for the sample class effect using RVM statistics;
  - groupPAdj: the adjusted [for multiple testing using the selected multiple testing correction method] p-value of the sample class effect);
  - groupRvmPAdj: the adjusted [for multiple testing using the selected multiple testing correction method] p-value of the sample class effect using RVM statistics).

- groupIntercepts A matrix with the group intercepts, i.e. the translational activity for each group independent of cytosolic mRNA level. Can be used for e.g. clustering of translational activity.
- correctionMethod: The multiple testing correction method used to adjust the nominal p-values.
- dsfSummary: A vector with the obtained frequencies of outlier dfbetas without the interaction term in the model.
- dfbetas: A matrix with the dfbetas from the model without the interaction term in the model.
- residuals: The residuals from the regressions without the interaction term in the model.
- fittedValues: A matrix with the fitted values from the regressions without the interaction term in the model.
- phenoClasses: The sample classes used in the analysis. The sample class order can be used to create the contrast matrix when using anota2seqRun or anota2seqAnalyze.
- sampleNames: A vector with the sample names (from the translated mRNA [e.g. polysome-associated mRNA or RPF] samples).
- abParametersInt: The ab parameters for the inverse gamma fit for the interactions within RVM.
- abParametersGroup: The ab parameters for the inverse gamma fit for treatments within RVM.

See Also

See Also as anota2seqPerformQC

Examples

```
data(anota2seq_data)
#Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
```

```
    dataType = "RNAseq",
    normalize = TRUE)
#Run QC
Anota2seqDataSet <- anota2seqPerformQC(Anota2seqDataSet)

#Get QC
qc <- anota2seqGetQualityControl(Anota2seqDataSet)
```

---

anota2seqGetResidOutlierTest
Retrieve residOutlierTest

Description

Retrieves the output of the anota2seqResidualOutlierTest function from the Anota2seqDataSet.

Usage

```
anota2seqGetResidOutlierTest(object)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetResidOutlierTest(object)
```

Arguments

object	An Anota2seqDataSet.

Value

A list with the following items:

- confInt: The selected confInt (see function arguments).
- rnormIter: The number of sampled data sets.
- outlierMatrixLog: A logical matrix describing which residuals were outliers in the last iteration of the analysis.
- meanOutlierPerIteration: The fraction outliers per iteration.
- obtainedComparedToExpected: The ratio of the expected number of outlier residuals compared to the expected number of outliers given the selected confInt.
- nExpected Number of expected outlier residuals.
- nObtained Number of obtained outliers residuals.

See Also

anota2seqResidOutlierTest

Examples

```
data(anota2seq_data)
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Run QC, it is mandatory for anota2seqResidualOutlierTest
Anota2seqDataSet <- anota2seqPerformQC(Anota2seqDataSet)
# Run the Residual outlier testing
Anota2seqDataSet <- anota2seqResidOutlierTest(Anota2seqDataSet)

#Get resdidual outlier testing
rot <- anota2seqGetResidOutlierTest(Anota2seqDataSet)
```

---

anota2seqGetThresholds

Get filtering criteria thresholds

Description

Get the filtering criteria used for the anota2seqSelSigGenes function.

Usage

```
anota2seqGetThresholds(object, analysis, selContrast)

## S4 method for signature 'Anota2seqDataSet'
anota2seqGetThresholds(object, analysis,
  selContrast)
```

Arguments

`object`	An Anota2seqDataSet.
`analysis`	A vector containing "translated mRNA", "total mRNA", "translation" or "buffering" specifying for which analysis the selected thresholds should be returned.
`selContrast`	A numeric vector specifying for which contrast the output should be retrieved. The contrast number corresponds to the position of the column in the automatically generated or specified contrast matrix.

Value

A list with the filtering criteria applied when filtering the anota2seqAnalyze output using the anota2seqSelSigGenes function. For details on filtering criteria see anota2seqSelSigGenes function help.

See Also

anota2seqSelSigGenes

Examples

```
data(anota2seq_data)
#Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
#Run analysis of differential translation
Anota2seqDataSet <- anota2seqAnalyze(Anota2seqDataSet, analysis = "translation")
#Run anota2seqSelSigGenes
Anota2seqDataSet <- anota2seqSelSigGenes(Anota2seqDataSet,
                                         analysis="translation",
                                         selContrast = 1,
                                         maxPAdj = .15)

# Get delta thresholds
thresholds <- anota2seqGetThresholds(Anota2seqDataSet,
                                     analysis = "translation",
                                     selContrast = 1)
```

---

anota2seqPerformQC *Perform quality control to ensure that the supplied data set is suitable for Analysis of Partial Variance (APV) within anota2seq.*

Description

Generates a distribution of interaction p-values which are compared to the expected NULL distribution. Also assesses the frequency of highly influential data points using dfbetas for the regression slope and compares the dfbetas to randomly generated simulation data. Calculates omnibus treatment effects.

Usage

```
anota2seqPerformQC(Anota2seqDataSet, generateSingleGenePlots = FALSE,
  fileName = "ANOTA2SEQ_translation_vs_mRNA_individual_regressions.pdf",
  nReg = 200, correctionMethod = "BH", useDfb = TRUE, useDfbSim = TRUE,
  nDfbSimData = 2000, useRVM = TRUE, onlyGroup = FALSE,
  useProgBar = TRUE, fileStem = "ANOTA2SEQ")
```

Arguments

Anota2seqDataSet	An object of class Anota2seqDataSet.
generateSingleGenePlots	anota2seq can plot the regression for each gene. However, as there are many genes, this output is normally not informative. TRUE/FALSE with default FALSE, no individual plotting.
fileName	If generateSingleGenePlots is set to TRUE use file to set desired file name (prints to current directory as a pdf). Default is "ANOTA2SEQ_translation_vs_mRNA_individual_regressions.pd "
nReg	If generateSingleGenePlots is set to TRUE, nReg can be used to limit the number of output plots. Default is 200. NOTE: this parameter plots the top "n" genes in the same order as the input data.
correctionMethod	anota2seq adjusts the omnibus interaction and treatment p-values for multiple testing. Correction method can be "Bonferroni", "Holm", "Hochberg", "SidakSS", "SidakSD", "BH", "BY", "ABH" or "TSBH" as implemented in the multtest package or "qvalue" as implemented in the qvalue package. Default is "BH".
useDfb	Should anota2seq assess the occurrence of highly influential data points (TRUE/FALSE with default TRUE).
useDfbSim	The random occurrence of dfbetas can be simulated. Default is TRUE. FALSE represses simulation which reduces computation time but makes interpretation of the dfbetas difficult.
nDfbSimData	If useDfbSim is TRUE the user can select the number of sampling that will be performed per step (10 steps with different correlations between the polysome association and the total mRNA level). Default is 2000.
useRVM	The Random Variance Model (RVM) can be used for the omnibus treatment analysis. In this case the effect of RVM on the distribution of the interaction significances needs to be tested as well. TRUE/FALSE where default (TRUE) leads to calculation of RVM p-values for both omnibus interactions and omnibus treatment effects.
onlyGroup	It is possible to suppress the omnibus interaction analysis and only perform the omnibus treatment analysis. TRUE/FALSE with default FALSE (analyze both interactions and treatment effects.)
useProgBar	Should the progress bar be shown. TRUE/FALSE with default TRUE, show progress bar.
fileStem	This stem will be added in front of each output filename. Default is "ANOTA2SEQ".

Details

The anota2seqPerformQC performs the basic quality control of the data set. Two levels of quality control are assessed, both of which need to show good performance for valid application of anota2seq. First, anota2seq assumes that there are no interactions (for slopes). The output for this analysis is both a density plot and a histogram plot of both the raw p-values and the p-values adjusted by the selected multiple correction method (if RVM was used, the second page shows the

same presentation using RMV p-values). anota2seq requires a uniform distribution of the raw interaction p-values for valid analysis of changes in translational efficiency affecting protein levels and buffering. anota2seq also assesses if there are more data points with high influence on the regression analyzes than would be expected by chance. anota2seq identifies influential data points as data points that influence the slope of the regression using standardized dfbeta (dfbetas). The function also performs an omnibus treatment effect test if there are more than 2 treatments. It is possible to use RVM for the omnibus treatment statistics. If RVM is used, it is necessary to verify that the interaction RVM p-values also follow the expected NULL distribution.

Value

An Anota2seqDataSet. anota2seqPerformQC saves its output data in the 'qualityControl' slot of the Anota2seqDataSet, see anota2seqGetQualityControl for a detailed description of this output.

anota2seqPerformQC also generates several graphical outputs. One output ("ANOTA2SEQ_interaction_p_distribution.pdf") shows the distribution of p-values and adjusted p-values for the omnibus interaction (both using densities and histograms). The second page of the pdf displays the same plots but for the RVM statistics if RVM is used. One output ("ANOTA2SEQ_simulated_vs_obtained_dfbs.pdf") shows bar graphs of the frequencies of outlier dfbetas using different dfbetas thresholds. If the simulation was enabled (recommended) these are compared to the frequencies from the random data set. One optional graphical output shows the gene by gene regressions with the sample classes indicated. In the case where RVM is used, a Q-Q plot and a comparison of the CDF of the variances to the theoretical CDF of the F-distribution is generated (output as "ANOTA2SEQ_rvm_fit_for_....jpg") for both the omnibus sample class and the omnibus interaction test.

See Also

anota2seqGetQualityControl

Examples

```
## Not run:
data(anota2seq_data)
Anota2seqDataSet <- anota2seqDataSetFromMatrix(dataP = anota2seq_data_P[1:100,],
                                    dataT = anota2seq_data_T[1:100,],
                                    phenoVec = anota2seq_pheno_vec,
                                    dataType = "RNAseq",
                                    normalize = TRUE)

Anota2seqDataSet <- anota2seqPerformQC(Anota2seqDataSet)

## End(Not run)
```

`anota2seqRegModes` *Categorize genes into regulatory modes of gene expression (mRNA abundance or changes in translational efficiency leading to altered protein levels or buffering)*

Description

Polysome or ribosome profiling allows the user to distinguish between three regulatory modes of gene expression: changes mRNA abundance or translational efficiency leading to altered protein levels or buffering (see references). Following a complete analysis, this function categorizes each identifier into one of these regulatory patterns.

Usage

```
anota2seqRegModes(Anota2seqDataSet, mRNASelect = c(TRUE, TRUE))
```

Arguments

`Anota2seqDataSet`	An Anota2seqDataSet containing the output of the anota2seqAnalyze and anota2seqSelSigGenes function for all analyzes.
`mRNASelect`	A logical vector with length 2 that specifies how mRNA abundance genes will be identified. The first position of the vector indicates whether translated mRNA output (e.g. polysome-associated mRNA or RPFs) should be taken into consideration when selecting mRNA abundance genes. The second position specifies whether the total mRNA output should be taken into consideration when selecting mRNA abundance genes. (See details) Default is c(TRUE,TRUE), i.e. abundance identifiers need to pass thresholds for both changes in translated mRNA and total mRNA.

Details

The regulatory modes are given a priority order. First, changes in translational efficiency leading to altered protein levels is considered such that if an identifier passed filtering for this analysis it will be allocated to the translation group. Identifiers that did not pass those thresholds are then considered for belonging to the mRNA abundance group. For this step, the user can decide to base inclusion into the abundance group based on results from analysis of translated mRNA (e.g. polysome-associated mRNA or RPF), total mRNA or both (see below). Finally, the remaining identifiers can be assessed for belonging to the translational buffering group.

An mRNA abundance identifier can be defined based on analysis of translated mRNA (e.g. polysome associated mRNA or or RPF) and total mRNA (and also assuring that fold changes follow the same direction). To perform such analysis, mRNASelect should be set to c(TRUE, TRUE). Alternatively one can identify mRNA abundance identifiers based on analysis of only translated mRNA (e.g. polysome-associated mRNA or RPFs; mRNASelect = c(TRUE, FALSE)) or only total mRNA (mRNASelect = c(FALSE, TRUE)).

Value

An Anota2seqDataSet. anota2seqRegModes adds a column specifying the regulatory modes of each significant identifier to the dataframe in the 'selectedTranslatedmRNA', 'selectedTotalmRNA', 'selectedTranslation' and 'selectedBuffering' slots of the Anota2seqDataSet. It also saves the output of anota2seqSelSigGenes for identifier regulated by mRNA abundance in the mRNAAbundance slot of the object. See `anota2seqGetOutput` for a detailed description of these outputs.

References

Oertlin C. et al. Genome-wide analysis of differential translation and differential translational buffering using anota2seq, bioRxiv, 2017.

See Also

anota2seqAnalyze, anota2seqSelSigGenes, anota2seqRun

Examples

```
# Initialize the Anota2seqDataSet
data(anota2seq_data)
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)

# Perform analysis of changes in translational efficiency leading to altered
# protein levels or buffering; and differential expression of translated mRNA
# and total mRNA
Anota2seqDataSet <- anota2seqAnalyze(Anota2seqDataSet,
                                     analysis = c("translated mRNA","total mRNA",
                                                  "translation","buffering"))

# Select as significant genes of all analyzes, genes with a FDR < 0.15
Anota2seqDataSet <- anota2seqSelSigGenes(Anota2seqDataSet,
                                         maxPAdj = .15,
                                         analysis = c("translated mRNA",
                                                      "total mRNA",
                                                      "translation",
                                                      "buffering"),
                                         selContrast = 1)

# Determine which identifiers belong to which regulatory modes.
Anota2seqDataSet <- anota2seqRegModes(Anota2seqDataSet)
```

---

anota2seqResidOutlierTest

Test for normality of residuals

Description

One assumption when performing APV is that the residuals from the regressions are normally distributed. anota2seq assesses this by comparing the Q-Q plots of the residuals to envelopes derived by sampling from the normal distribution.

Usage

```
anota2seqResidOutlierTest(Anota2seqDataSet, confInt = 0.01,
  iter = 5, generateSingleGenePlots = FALSE, nGraphs = 200,
  generateSummaryPlot = TRUE, residFitPlot = TRUE, useProgBar = TRUE)
```

Arguments

Anota2seqDataSet	An object of class Anota2seqDataSet that also contains the output of the anota2seqPerformQC function.
confInt	Controls how many samples from the normal distribution will be used to generate the envelope to which the residuals are compared. Default is 0.01 which will generate 99 samples from the normal distribution to compare to the actual residuals.
iter	How many times should the analysis be performed? Default is 5 meaning that 5 sets of samples (each with the size controlled by confInt) will be generated. Notice that the summary plotting is only performed for the last set but the percentage of outliers for each iteration can be found in the output object.
generateSingleGenePlots	The analysis is performed per identifier and plots can be generated for each identifier. However, due to the high number of identifiers, a large number of plots will typically be generated. TRUE/FALSE with default FALSE.
nGraphs	If generateSingleGenePlots is set to TRUE, nGraphs controls for how many identifiers such single gene graphs will be generated. Default is 200. NOTE: this parameter plots the top "n" genes in the same order as the input data.
generateSummaryPlot	The function can generate a summary graph that shows the envelopes generated by sampling from the normal distribution compared to the obtained values for all genes. Default is TRUE, thus the graph is generated but only from the last iteration.
residFitPlot	Generates an output of the fitted values and residuals. Default is TRUE, generate the plot.
useProgBar	Should the progress bar be shown. Default is TRUE, show progress bar.

Details

The anota2seqResidOutlierTest function assesses whether the residuals from the per identifier linear regressions of translated mRNA level~total mRNA level+treatment are normally distributed. anota2seq generates normal Q-Q plots of the residuals. If the residuals are normally distributed, the data quantiles will form a straight diagonal line from bottom left to top right. Because there are typically relatively few data points, anota2seq calculates "envelopes" based on a set of samplings from the normal distribution using the same number of data points as for the true data *Venables,Ripley*.To enable a comparison both the actual and the sampled data are centered (mean=0) and scaled (sd=1). The data (both true and sampled) are then sorted and the true sample is compared to the envelopes of the sampled data at each sort position. The result is presented as a Q-Q plot of the true data where the envelopes of the sampled data are indicated. If there are 99 samplings we expect that 1/100 values to be outside the envelopes obtained from the samplings. Thus it is possible to assess

if approximately the expected number of outlier residuals are obtained. The result is presented as both a graphical output and an output object.

Value

An Anota2seqDataSet. anota2seqResidOutlierTest saves its output data in the 'residOutlierTest' slot of the Anota2seqDataSet, see anota2seqGetResidOutlierTest for a detailed description of this output.

anota2seqResidOutlierTest also generates a graphical output ("ANOTA2SEQ_residual_distribution_summary.pdf") showing the Q-Q plots from all genes as well as the envelopes from the sampled data. The obtained percentage of outliers is shown at each rank position and all combined. Optionally, when generateSingleGenePlots is set to TRUE, the function also generates individual plots (stored as "ANOTA2SEQ_residual_distributions_single.pdf") for n genes (set by nGraphs). When residFitPlot is set to TRUE an output comparing the fitted values to the residuals is generated (stored as "ANOTA2SEQ_residuals_vs_fitted.jpeg").

References

Venables, W.N. and Ripley, B.D., Modern Applied Statistics with S-PLUS, *springer* (1999).

See Also

anota2seqGetResidOutlierTest

Examples

```
## Not run:
data(anota2seq_data)
# Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Perform anota2seqPerformQC function. This must be performed prior the running
# the anota2seqResidualOutlierTest function.
Anota2seqDataSet <- anota2seqPerformQC(Anota2seqDataSet)
# Perform anota2seqResidualOutlierTest function
Anota2seqDataSet <- anota2seqResidOutlierTest(Anota2seqDataSet)

## End(Not run)
```

---

anota2seqRun *Wrapper for several functions which provide one-step analysis*

Description

anota2seqRun is a wrapper function running the following steps of the anota2seq workflow: assessing model assumptions (by calling anota2seqPerformQC and anota2seqResidOutlierTest), performing analysis of changes in translational efficiency leading to altered protein levels or buffering; and differential expression of translated mRNA (e.g. polysome-associated mRNA or RPF) and total mRNA (by calling anota2seqAnalyze sequentially with analysis parameter set to "translation", "buffering", "translated mRNA", "total mRNA" respectively). Gene filtering is performed by calling anota2seqSelSigGenes and anota2seqRegModes categorizes regulated genes into regulatory modes (mRNA abundance or changes in translational efficiency leading to altered protein levels or buffering).

Usage

```
anota2seqRun(Anota2seqDataSet, contrasts = NULL, performQC = TRUE,
  onlyGroup = FALSE, performROT = TRUE, generateSingleGenePlots = FALSE,
  analyzeBuffering = TRUE, analyzemRNA = TRUE, thresholds = NULL,
  useRVM = TRUE, correctionMethod = "BH", useProgBar = TRUE)
```

Arguments

Anota2seqDataSet	Object of class Anota2seqDataSet
contrasts	If NULL (default), the contrasts will be created automatically and may not correspond to those of interest. It is therefore possible to use custom contrasts using this parameter. The input should be a matrix with row names corresponding to treatments (i.e. those in phenoVec or the treatment column of the SummarizedExperiment annotation) and columns corresponding to the different contrasts of interest (additional details on how to set up the contrast matrix is indicated in details section).
performQC	Boolean that defaults to TRUE. Used to specify if the anota2seqPerformQC function should be run.
onlyGroup	Boolean parameter of the anota2seqPerformQC function (default: FALSE). In anota2seqPerformQC, it is possible to suppress the omnibus interaction analysis and only perform the omnibus treatment analysis. Typically, when the data contains less than 3 samples in each sample class and more than 2 sample classes, the interaction analysis cannot be performed but the onlyGroup mode (i.e. with onlyGroup = TRUE) can be used to assess omnibus group effects.
performROT	Boolean that defaults to TRUE. Used to specify if the anota2seqResidOutlierTest function should be run.
generateSingleGenePlots	Should the single gene graphical outputs from the anota2seqPerformQC and anota2seqResidOulierTest functions be generated. Default is set to FALSE.
analyzeBuffering	Boolean that defaults to TRUE. Used to specify if changes in translational efficiency leading to buffering should be be analyzed.
analyzemRNA	Boolean that defaults to TRUE. Used to specify if translated mRNA (e.g. polysome-associated mRNA or RPFs) and total mRNA should be analyzed.

thresholds A list containing thresholds that are applied during filtering of several parameters as described for the anota2seqSelSigGenes function. This list can contain the following name slots and if different from the below default values will update such defaults:

- minSlopeTranslation: The output can be filtered so that identifiers whose identified slopes in analysis of changes in translational efficiency leading to altered protein levels are too small can be excluded. Default is -1 i.e. excludes identifiers with a slope <(-1).
- maxSlopeTranslation: The output can be filtered so that identifiers whose identified slopes in analysis of changes in translational efficiency leading to altered protein levels are too large can be excluded. Default is 2 i.e. excludes identifiers with a slope > 2.
- minSlopeBuffering: The output can be filtered so that identifiers whose identified slopes in analysis of changes in translational efficiency leading to buffering are too small can be excluded. Default is -2 i.e. excludes identifiers with a slope <(-2).
- maxSlopeBuffering: The output can be filtered so that genes whose identified slopes in analysis of changes in translational efficiency leading to buffering are too large can be excluded. Default is 1 i.e. excludes identifiers with a slope > 1.
- maxPAdj: The output can be filtered based on adjusted p-values from the anota2seq analysis (i.e. smaller compared to assigned value). If useRVM is set to TRUE, filtering will be performed on RVM values, otherwise filtering will be performed on non-RVM values. Multiple testing adjustment method is set by argument correctionMethod. Default is 0.15.
- maxP: The output can be filtered based on unadjusted p-values from the anota2seq analysis (i.e. smaller compared to assigned value). If useRVM is set to TRUE, filtering will be performed on RVM values, otherwise filtering will be performed on non-RVM values. Default is NULL, i.e. no filtering.
- minEff: The output can be filtered based on minimum effect for inclusion. The value is applied both to negative and positive effects: e.g. a value of 1 will evaluate if the effects are >1 OR <(-1). Default is NULL i.e. no filtering based on effect.
- deltaPT: The output can be filtered based on the mean log2(translated mRNA data [e.g. polysome-associated mRNA or RPFs] / total mRNA data) difference between treatments. The treatments are defined by the selected contrast. Default is log2(1.2). Only applied when analysis is set to "translation".
- deltaTP: The output can be filtered based on the mean log2(total mRNA / translated mRNA data [e.g. polysome-associated mRNA or RPFs]) difference between treatments. The treatments are defined by the selected contrast. Default is log2(1.2). Only applied when analysis is set to "buffering".
- deltaP: The output can be filtered based on a minimum effect between the treatment and control groups in polysome-associated mRNA. Default is NULL (i.e. no filtering). Use when analysis parameter is set to "translation" (i.e. changes in translational efficiency leading to altered protein levels) or "translated mRNA"

- deltaT: The output can be filtered based on a minimum effect between the treatment and control groups in total mRNA. Default is NULL (i.e. no filtering). Use when analysis parameter is set to "buffering" or "total mRNA".

`useRVM` Should the Random Variance Model be applied. Default is TRUE.

`correctionMethod`
Correction for multiple testing method. This parameter can be set to "Bonferroni", "Holm", "Hochberg", "SidakSS", "SidakSD", "BH", "BY", "ABH" or "TSBH" as implemented in the multtest package or "qvalue" as implemented in the qvalue package. Default is "BH".

`useProgBar` Should the progress bar be shown. Default is TRUE, show progress bar.

Details

At the step of analysis, by default (i.e. with contrasts = NULL) the order of the sample classes which are used to calculate differences between treatments will be in alphabetical order. To change the directionality of the contrasts (e.g. treatment b vs treatment a instead of treatment a vs treatment b) or to generate a custom set of contrasts when more than 2 treatments are included, a contrast matrix can be supplied to the "contrasts" parameter described above. The row names should be specified as indicated above. The contrasts are coded by using e.g. -1 for group a, 0 for group b and 1 for group c to compare group a and c; -2 for group a, 1 for group b and 1 for group c to compare group a to b & c. Each column of the contrast matrix should sum to 0 and to analyze orthagonal contrasts the products of all pairwise rows should sum to 0. The results in the Anota2seqDataSet object will follow the order of the contrasts (i.e. results for e.g. contrast 1 will correspond to the contrasts specified in column 1 of the contrast matrix).

Value

An Anota2seqDataSet containing normalized data, model covariates and contrasts as well as outputs of all functions called by `anota2seqRun`. anota2seqRun will also output all diagnostic plots provided by `anota2seqPerformQC`, `anota2seqResidOutlierTest` and `anota2seqAnalyze`.

See Also

`anota2seqPerformQC`, `anota2seqResidOutlierTest`, `anota2seqAnalyze`, `anota2seqSelSigGenes` `anota2seqRegModes`

Examples

```
data(anota2seq_data)
# Initialize Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Perform anota2seqRun function
# The quality control and residual outlier testing are not
# performed in order to limit the running time of this example, but the model
# assumptions should be assessed (see help of anota2seqPerformQC)
```

```
Anota2seqDataSet <- anota2seqRun(Anota2seqDataSet,
    performQC = FALSE,
    performROT = FALSE,
    useProgBar = FALSE)

## Not run:
# Example to build a custom contrast matrix
# For the purpose of this example, we will use the first 6 samples of the
# simulated data provided with the package together with the following "dummy"
# sample classes:
phenoVec <- c("a","a","b","b","c","c")
contrastsEx_ads <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:300, 1:6],
    dataT = anota2seq_data_T[1:300, 1:6],
    phenoVec = phenoVec,
    dataType = "RNAseq",
    normalize = TRUE)

# Get the levels of the phenoVec, these will be ordered as in anota2seq
phenoLev <- levels(as.factor(phenoVec))
# Construct the matrix with appropriate nrow and ncol
myContrast <- matrix(nrow =length(phenoLev),ncol=length(phenoLev)-1)
# Set the phenoLev as rownames for your contrast matrix
rownames(myContrast) <- phenoLev
# Now indicate the contrasts you want to analyse as explained above
# Compare a to c
myContrast[,1] <- c(-1,0,1)
# Compare a to b& c
myContrast[,2] <- c(2,-1,-1)
myContrast
#   [,1] [,2]
# a   -1    2
# b    0   -1
# c    1   -1
# The custom contrast matrix can then be used as input of anota2seqRun. Because
# these data have only 2 samples per sample class, the onlyGroup mode of
# anota2seqPerformQC is the only available mode for assessment of model
# assumptions so we also set onlyGroup to TRUE.
contrastsEx_ads <- anota2seqRun(contrastsEx_ads,
                                contrasts = myContrast,
                                performQC = FALSE,
                                performROT = FALSE,
                                onlyGroup = TRUE,
                                thresholds = list(
                                    maxPAdj = 0.25,
                                    deltaPT = log2(2)))

## End(Not run)
```

---

anota2seqSelSigGenes *Select identifiers from the output of* anota2seqAnalyze.

Description

This function filters the output based on significance threshold, effect sizes (of total mRNA, translated mRNA, translation, buffering), identifier names and/or APV regression slopes.

Usage

```
anota2seqSelSigGenes(Anota2seqDataSet, useRVM = TRUE,
  analysis = anota2seqGetAvailableAnalyzes(Anota2seqDataSet),
  selIds = NULL, selContrast = seq(along =
  1:dim(anota2seqGetContrasts(Anota2seqDataSet))[2]),
  minSlopeTranslation = NULL, maxSlopeTranslation = NULL,
  minSlopeBuffering = NULL, maxSlopeBuffering = NULL, slopeP = NULL,
  minEff = NULL, maxP = NULL, maxPAdj = NULL, selDeltaPT = NULL,
  selDeltaTP = NULL, selDeltaP = NULL, selDeltaT = NULL,
  sortBy = c("rvmP", "none", "Eff", "p"))
```

Arguments

`Anota2seqDataSet`
: An Anota2seqDataSet containing the output from the anota2seqAnalyze function performed on the analysis to be filtered by anota2seqSelSigGenes.

`useRVM`
: If a significance threshold is provided (maxP or maxPAdj, see below), should the output be filtered on Random Variance Model p-values or on non-RVM p-values? Default is TRUE i.e. RVM p-values.

`analysis`
: A vector of character string(s) specifying which analysis(es) should be considered for filtering. Should be at least one of the following: "translated mRNA", "total mRNA", "translation" or "buffering". By default, filtering will be performed on all available analyzes in the Anota2seqDataSet object.

`selIds`
: The function can consider only a subset of the identifiers from the input data set. For custom selection of identifiers, supply a vector of identifiers (row names from the original data set) to be included. Default is NULL i.e. filtering is performed on all identifiers.

`selContrast`
: Which contrast(s) should be evaluated during the filtering and sorting? Descriptions of the contrasts can be found in the output from the anota2seqAnalyze object in the usedContrasts slot. Indicate the contrast(s) by a numeric vector of the column number(s). By default, all available contrasts in the Anota2seqDataSet object are filtered.

`minSlopeTranslation`
: The output can be filtered so that identifiers whose identified slopes in analysis of changes in translational efficiency leading to altered protein levels are too small can be excluded. Default is NULL i.e. no filtering based on lower boundary of the slope. To exclude identifiers with e.g. a slope <(-1), assign -1 to minSlopeTranslation. Only applied when analysis is set to "translation".

`maxSlopeTranslation`
: The output can be filtered so that identifiers whose identified slopes in analysis of changes in translational efficiency leading to altered protein levels are too large can be excluded. Default is NULL i.e. no filtering based on upper

boundary of the slope. To exclude identifiers with e.g. a slope >2, assign 2 to maxSlopeTranslation. Only applied when analysis is set to "translation".

`minSlopeBuffering`
: The output can be filtered so that identifiers whose identified slopes in analysis of changes in translational efficiency leading to buffering are too small can be excluded. Default is NULL i.e. no filtering based on lower boundary of the slope. To exclude identifiers with e.g. a slope <(-2), assign -2 to minSlopeBuffering Only applied when analysis is set to "buffering".

`maxSlopeBuffering`
: The output can be filtered so that genes whose identified slopes in analysis of changes in translational efficiency leading to buffering are too large can be excluded. Default is NULL i.e. no filtering based on upper boundary of the slope. To exclude identifiers with e.g. a slope > 1, assign 1 to maxSlopeBuffering. Only applied when analysis is set to "buffering".

`slopeP`
: A p-value for the slope being <0 or >1 if the estimate for the slope is <0 or >1 (for translation) and <-1 or >0 if the estimate for the slope is <-1 or >0 (for buffering). This p-value can be used to filter the output based on unrealistic slopes. When set low, fewer identifiers will be disqualified. Default is NULL i.e. no filtering based on slope p-value. Only applied for when analysis is set to "translation" or "buffering".

`minEff`
: The output can be filtered based on minimum effect for inclusion. The value is applied both to negative and positive effects: e.g. a value of 1 will evaluate if the effects are >1 OR <(-1). Default is NULL i.e. no filtering based on effect.

`maxP`
: The output can be filtered based on raw p-values from the anota2seq analysis (i.e. smaller compared to assigned value). If useRVM is set to TRUE, filtering will be performed on RVM values, otherwise filtering will be performed on non-RVM values. Default is NULL i.e. no filtering.

`maxPAdj`
: The output can be filtered based on adjusted p-values from the anota2seq analysis (i.e. smaller compared to assigned value). If useRVM is set to TRUE, filtering will be performed on RVM values, otherwise filtering will be performed on non-RVM values. The adjustment method that was used when running anota2seqAnalyze will be evaluated. Default is NULL i.e. no filtering.

`selDeltaPT`
: The output can be filtered based on the mean log2(translation data (e.g. polysome-associated mRNA) / total mRNA data) between treatments difference. The treatments are defined by the selected contrast. Default is NULL i.e. no filtering. Only applied when analysis is set to "translation".

`selDeltaTP`
: The output can be filtered based on the mean log2(total mRNA / translation data (e.g. polysome-associated mRNA)) between treatments difference. The treatments are defined by the selected contrast. Default is NULL i.e. no filtering. Only applied when analysis is set to "buffering".

`selDeltaP`
: The output for analysis changes in translational efficiency leading to altered protein levels can be filtered based on the translation data only so that the minimum absolute difference between treatments for translated mRNA data (e.g. polysome-associated mRNA) is used for inclusion. The treatments are defined by the selected contrast. Default is NULL i.e. no filtering. Use when analysis parameter is set to "translation" (i.e. changes in translational efficiency leading to altered protein levels) or "translated mRNA".

selDeltaT — The output for analysis of changes in translational efficiency leading to buffering can be filtered based on the total mRNA data only so that the minimum absolute difference between treatments for buffering is used for inclusion. The treatments are defined by the selected contrast. Default is NULL i.e. no filtering. Use when analysis parameter is set to "buffering" or "total mRNA".

sortBy — The output can be sorted by effect ("Eff"), raw p-value ("p"), raw RVM p-value ("rvmP") or not sorted ("none"). Default is "rvmP" when useRVM is set to TRUE and "p" when useRVM is set to FALSE.

Value

An Anota2seqDataSet. anota2seqSelSigGenes saves its output data in the 'selectedTranslatedmRNA', 'selectedTotalmRNA', 'selectedTranslation' and 'selectedBuffering' slots of the Anota2seqDataSet when analysis is set to "translated mRNA", "total mRNA", "translation" and "buffering" respectively. This output contains statistics from the applied APV model on selected genes, see anota2seqGetOutput for a detailed description.

See Also

anota2seqAnalyze, anota2seqGetOutput

Examples

```
data(anota2seq_data)
# Initialize the Anota2seqDataSet
Anota2seqDataSet <- anota2seqDataSetFromMatrix(
    dataP = anota2seq_data_P[1:100,],
    dataT = anota2seq_data_T[1:100,],
    phenoVec = anota2seq_pheno_vec,
    dataType = "RNAseq",
    normalize = TRUE)
# Run analysis of changes in translational efficiency leading to
# altered protein levels
Anota2seqDataSet <- anota2seqAnalyze(
    Anota2seqDataSet,
    analysis = c("translation"))
# Select significant genes (with a FDR threshold of 0.15) for changes
# in translational efficiency leading to altered protein level
Anota2seqDataSet <- anota2seqSelSigGenes(Anota2seqDataSet,
                                         maxPAdj = .15)
```

---

anota2seq_data *Sample data set for anota2seq*

Description

Simulated data used to illustrate the anota2seq workflow.

Usage

```
data(anota2seq_data)
```

Format

Eight samples were simulated from 2 sample classes ("control" and "treatment"); both total mRNA (anota2seq_data_T) and paired translated mRNA (anota2seq_data_P) are provided together with a sample class vector (anota2seq_pheno_vec).

anota2seq_data_T and anota2seq_data_P are 2 matrices with 10748 rows and 8 columns simulated to represent raw RNAseq counts from 8 samples (4 replicates of each condition) and 10748 mRNAs. Columns are ordered according to the RNA type pairing i.e. column 1 of anota2seq_data_T is from the same sample as column 1 of anota2seq_data_P etc.

Source

These data originate from the study by *Oertlin et al.* which compared methods for analysis of translatomes quantified by RNAseq.

References

Oertlin C. et al. Genome-wide analysis of differential translation and differential translational buffering using anota2seq, bioRxiv, 2017.

Examples

```
##load data set
data(anota2seq_data)
##check dimensions
dim(anota2seq_data_P)
head(anota2seq_data_P)
dim(anota2seq_data_T)
head(anota2seq_data_T)
anota2seq_pheno_vec
```

---

show,Anota2seqDataSet-method

Show information for an Anota2seqDataSet

Description

This function gives basic information on the slots within your Anota2seqDataSet.

Usage

```
## S4 method for signature 'Anota2seqDataSet'
show(object)
```

Arguments

object	An Anota2seqDataSet.

Value

No value is returned; this function is called for its side effects.

Examples

```
data(anota2seq_data)
Anota2seqDataSet <- anota2seqDataSetFromMatrix(dataP = anota2seq_data_P[1:1000,],
                                       dataT = anota2seq_data_T[1:1000,],
                                       phenoVec = anota2seq_pheno_vec,
                                       dataType = "RNAseq",
                                       normalize = TRUE)

show(Anota2seqDataSet)
```

Index

∗ **datasets**
anota2seq_data, 38

anota2seq-plots, 2
anota2seq_data, 38
anota2seq_data_P *(*anota2seq_data*)*, 38
anota2seq_data_T *(*anota2seq_data*)*, 38
anota2seq_pheno_vec *(*anota2seq_data*)*, 38
anota2seqAnalyze, *4*, 5, *6*, *8, 9*, *12*, *16*, *20*, *29*, *32*, *34*, *38*
Anota2seqDataSet *(*Anota2seqDataSet-class*)*, 8
Anota2seqDataSet-class, 8
anota2seqDataSetFromMatrix, 9
anota2seqDataSetFromSE *(*anota2seqDataSetFromMatrix*)*, 9
anota2seqGetContrasts, *8*, 13
anota2seqGetContrasts,Anota2seqDataSet-method *(*anota2seqGetContrasts*)*, 13
anota2seqGetCovariates, *8*, 14
anota2seqGetCovariates,Anota2seqDataSet-method *(*anota2seqGetCovariates*)*, 14
anota2seqGetDeltaData, *6, 7*, 15
anota2seqGetDeltaData,Anota2seqDataSet-method *(*anota2seqGetDeltaData*)*, 15
anota2seqGetNormalizedData, *8*, 17
anota2seqGetNormalizedData,Anota2seqDataSet-method *(*anota2seqGetNormalizedData*)*, 17
anota2seqGetOutput, *6–9*, 18, *28*, *38*
anota2seqGetOutput,Anota2seqDataSet-method *(*anota2seqGetOutput*)*, 18
anota2seqGetQualityControl, *8*, 21, *27*
anota2seqGetQualityControl,Anota2seqDataSet-method *(*anota2seqGetQualityControl*)*, 21
anota2seqGetResidOutlierTest, *8*, 23, *31*
anota2seqGetResidOutlierTest,Anota2seqDataSet-method *(*anota2seqGetResidOutlierTest*)*, 23
anota2seqGetThresholds, 24
anota2seqGetThresholds,Anota2seqDataSet-method *(*anota2seqGetThresholds*)*, 24
anota2seqPerformQC, *8*, *22*, 25, *32*, *34*
anota2seqPlotFC, *18*
anota2seqPlotFC *(*anota2seq-plots*)*, 2
anota2seqPlotFC,Anota2seqDataSet-method *(*anota2seq-plots*)*, 2
anota2seqPlotGenes *(*anota2seq-plots*)*, 2
anota2seqPlotGenes,Anota2seqDataSet-method *(*anota2seq-plots*)*, 2
anota2seqPlotPvalues *(*anota2seq-plots*)*, 2
anota2seqPlotPvalues,Anota2seqDataSet-method *(*anota2seq-plots*)*, 2
anota2seqRegModes, *8, 9*, *20*, 27, *32*, *34*
anota2seqResidOutlierTest, *8*, *23*, 29, *32*, *34*
anota2seqRun, *8, 9*, *12*, *20*, *29*, 31
anota2seqSelSigGenes, *3, 4*, *8, 9*, *16*, *20*, *25*, *29*, *32*, *34*, 35

show *(*show,Anota2seqDataSet-method*)*, 39
show,Anota2seqDataSet-method, 39