Package 'anota'

May 5, 2024

Version 1.52.0

Date 2011-03-03

Title ANalysis Of Translational Activity (ANOTA).

Author Ola Larsson <ola.larsson@ki.se>, Nahum Sonenberg <nahum.sonenberg@mcgill.ca>, Robert Nadon <robert.nadon@mcgill.ca>

Maintainer Ola Larsson <ola.larsson@ki.se>

Description Genome wide studies of translational control is emerging as a tool to study verious biological conditions. The output from such analysis is both the mRNA level (e.g. cytosolic mRNA level) and the levl of mRNA actively involved in translation (the actively translating mRNA level) for each mRNA. The standard analysis of such data strives towards identifying differential translational between two or more sample classes - i.e. differences in actively translated mRNA levels that are independent of underlying differences in cytosolic mRNA levels. This package allows for such analysis using partial variances and the random variance model. As 10s of thousands of mRNAs are analyzed in parallell the library performs a number of tests to assure that the data set is suitable for such analysis.

Imports multtest, qvalue

Depends qvalue

LazyData yes

LazyLoad yes

License GPL-3

biocViews GeneExpression, DifferentialExpression, Microarray, Sequencing

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

git_branch RELEASE_3_19

git_last_commit 73a9d51

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-05

Contents

anotaDataSet 2
anotaGetSigGenes 3
anotaPerformQc 5
anotaPlotSigGenes 8
anotaResidOutlierTest 11

Index **14**

`anotaDataSet` *Sample data set for anota*

Description

6 samples with data from 2 sample categories, both cytosolic (anotaDataT) and translational (anotaDataP) together with a sample class vector (anotaPhenoVec).

Usage

```
data(anotaDataSet)
```

Format

Each data matrix (anotaDataT and anotaDataP) has 1000 rows (1000 first identifiers from complete data set) and 6 columns (noAA or rich). The anotaPhenoVec vector contains the sample class of each sample and anotaDataT, anotaDataP and phenoVec follow the same order.

Source

Ingolia, NT et al. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science, 2009, 10;324(5924):218-23

Examples

```
##load data set
data(anotaDataSet)
##check dimensions
dim(anotaDataP)
head(anotaDataP)

dim(anotaDataT)
head(anotaDataT)

anotaPhenoVec
```

anotaGetSigGenes *Identify genes that are under translational control independent of cytosolic mRNA levels*

Description

This function uses analysis of partial variance (APV) to identify genes that are under translational regulation independent of cytosolic mRNA levels.

Usage

```
anotaGetSigGenes(dataT=NULL, dataP=NULL, phenoVec=NULL, anotaQcObj=NULL,
correctionMethod="BH", contrasts=NULL, useRVM=TRUE, useProgBar=TRUE)
```

Arguments

dataT	A matrix with cytosolic mRNA data. Non numerical rownames are needed.
dataP	A matrix with translational activity data. Non numerical rownames are needed.
phenoVec	A vector describing the sample classes (each class should have a unique identifier). Note that dataT, dataP and phenoVec have to have the same sample order so that column 1 in dataP is the translational data for a sample, column 1 in dataT is the cytosolic mRNA data and position 1 in phenoVec describes the sample class.
anotaQcObj	The object returned by anotaPerformQc.
correctionMethod	anota corrects p-values for multiple testing using the multtest package. 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".
contrasts	When there is more than 2 sample categories it is possible to use custom contrasts. The order of the sample classes needs to be correct and can be seen in the object generated from anotaPerformQc in the phenoClasses slot (see details section).
useRVM	Should the Random Variance Model be applied. Default is TRUE.
useProgBar	Should the progress bar be shown. Default is TRUE, show progress bar.

Details

The function performs APV on two or more sample categories. When more than two sample classes are compared it is possible to set custom contrasts to compare the sample classes of interest. Otherwise "treatment" contrasts are used which follow the alphabetical order of the sample classes. The order of the sample classes which the contrast matrix should follow can be found in the output of the anotaPerformQc function in the phenoClasses slot. Contrasts are supplied as a matrix where the sample classes are rows (same order as phenoClasses) and the columns are the different contrasts used. 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 will follow the order of the contrasts, i.e. the anocovaStats slot in the output object is a list with positions 1...n where 1 is the first contrast and n is the last.

A rare error can occur when data within dataT or dataP from any gene and any sample class has no variance. This is reported as "ANOVA F-TEST on essentially perfect fit...". In this case those genes that show no variance for a sample class within either dataT or dataP need to be removed before analysis. Trying a different normalization method may fix the problem.

Value

anotaGetSigGenes creates a plot showing the fit of the inverse gamma distribution used in RVM (similar output as from anotaPerformQc). anotaGetSigGenes also returns a list object with the following slots:

`apvStats` A list object (each slot named from 1 to the number of contrasts) where each slot contains a matrix with statistics from the applied APV for that contrast. Columns are "apvSlope" (the common slope used in APV); "apvSlopeP" (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); "unadjustedResidError" (the residual error before calculating the effective residual error); "apvEff" (the group effect); "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).

`apvStatsRvm` A summary list object (each slot named from 1 to the number of contrasts) where each slot contains a matrix with RVM statistics from the applied APV. Columns are "apvSlope" (the common slope used in APV); "apvSlopeP" (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); "apvEff" (the group effect); "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).

`correctionMethod` The multiple testing correction method used to adjust the p-values.

`usedContrasts` A matrix with the contrasts used. Order is same as in the statistical outputs (column wise) so that the first contrast is found in the first slot of the apvStats and the apvStatsRvm lists.

`abList` A list object containing the a and b parameters from the inverse gamma fits. Same order as the contrasts.

Author(s)

Ola Larsson <ola.larsson@ki.se>, Nahum Sonenberg <nahum.sonenberg@mcgill.ca>, Robert Nadon <robert.nadon@mcgill.ca>

See Also

anotaPerformQc, anotaResidOutlierTest, anotaPlotSigGenes

Examples

```
## See example for \code{\link{anotaPlotSigGenes}}
```

---

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

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 class effects.

Usage

```
anotaPerformQc(dataT=NULL, dataP=NULL, phenoVec=NULL,
generatePlot=FALSE, file="ANOTA_Total_vs_Polysomal_regressions.pdf",
nReg=200, correctionMethod="BH", useDfb=TRUE, useDfbSim=TRUE,
nDfbSimData=2000, useRVM=TRUE, onlyGroup=FALSE, useProgBar=TRUE)
```

Arguments

dataT	A matrix with cytosolic mRNA data. Non numerical rownames are needed.
dataP	A matrix with translational activity data. Non numerical rownames are needed.
phenoVec	A vector describing the sample classes (each class should have a unique identifier). Note that dataT, dataP and phenoVec must have the same sample order so that column 1 in dataP is the translational activity data for a sample, column 1 in dataT is the cytosolic mRNA data and position 1 in phenoVec describes the sample class.
generatePlot	anota can plot the regression for each gene. However, as there are many genes, this output is normally not informative. Default is FALSE, no individual plotting.
file	If generatePlot is set to TRUE use file to set desired file name (prints to current directory as a pdf). Default is "ANOTA_Total_vs_Polysomal_regressions.pdf"
nReg	If generatePlot is set to TRUE, nReg can be used to limit the number of output plots. Default is 200.
correctionMethod	anota adjusts the omnibus interaction and sample class 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 anota assess the occurrence of highly influential data points (defult is 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 samplings that will be performed per step (10 steps with different correlations between the translationally activty and the cytosolic mRNA level). Default is 2000.
useRVM	The Random Variance Model (RVM) can be used for the omnibus sample class comparison. In this case the effect of RVM on the distribution of the interaction significances needs to be tested as well. Default (TRUE) leads to calculation of RVM p-values for both omnibus interactions and omnibus sample class effects.
onlyGroup	It is possible to suppress the omnibus interaction analysis and only perform the omnibus sample class effect analysis. Default is FALSE (analyse both interactions and sample class effects.)
useProgBar	Should the progress bar be shown. Default is TRUE, show progress bar.

Details

The anotaPerformQc 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 anota. First, anota 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). anota requires a uniform distribution of the raw interaction p-values for valid analysis of differential translation. anota also assesses if there are more data points with high influence on the regression analyses than would be expected by chance. anota identifies influential data points as data points that influence the slope of the regression using standardized dfbeta (dfbetas). In the literature there are multiple suggestions of what should be regarded as an outlier dfbetas (dfbetas>1, dfbetas>2, dfbetas>3, dfbetas>(2/sqrt(N)), dfbetas>(3/sqrt(N)), dfbetas>(3.5*IQR)). Independent of which threshold is preferred, what is of interest is the comparison to the underlying distribution. As this distribution is unknown, we simulate random data sets assuming that the cytosolic mRNA level and the translationally active mRNA levels are normally distributed and that there is a correlation between the cytosolic and the translationally active mRNA level. Following such simulation the frequencies of outlier dfbetas (using all thresholds) is compared to the frequencies found in the simulated data set. The function also performs an omnibus sample class effect test if there are more than 2 sample classes. It is possible to use RVM for the omnibus sample class statistics. If RVM is used, it is necessary to verify that the interaction RVM p-values also follow the expected NULL distribution. A rare error can occur when data within dataT or dataP from any gene and any sample class has no variance. This is reported as "ANOVA F-TEST on essentially perfect fit...". In this case those genes that show no variance for a sample class within either dataT or dataP need to be removed before analysis. Trying a different normalization method may fix the problem.

Value

anotaPerformQc generates several graphical outputs. One output ("ANOTA_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 ("ANOTA_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 "ANOTA_rvm_fit_for_....jpg") for both the omnibus sample class and the omnibus interaction test. The function also outputs a list object 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 sample class 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 sample classes); "groupDf" (the degrees of freedom for the sample classes); "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).

`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 identifying differential translation using anotaGetSigGenes.

sampleNames A vector with the sample names (taken from the translationally active 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 sample classes within RVM.

Author(s)

Ola Larsson <ola.larsson@ki.se>, Nahum Sonenberg <nahum.sonenberg@mcgill.ca>, Robert Nadon <robert.nadon@mcgill.ca>

See Also

anotaResidOutlierTest, anotaGetSigGenes,anotaPlotSigGenes

Examples

```
## See example for \code{\link{anotaPlotSigGenes}}
```

---

anotaPlotSigGenes *Filter and plot genes to visualize/summarize genes that are differentially translated.*

Description

This function filters the output from the anotaGetSigGenes function based on many user defined thresholds and flags to generate a summary table and optional per gene plots.

Usage

```
anotaPlotSigGenes(anotaSigObj, selIds=NULL, selContr=NULL, minSlope=NULL, maxSlope=NULL, slopeP=NULL,
```

Arguments

anotaSigObj The output from the anotaGetSigGenes function.

selIds The function can consider only a subset of the identifiers from the input data set (which can be further filtered) or used for custom plotting of identifiers of interest (leaving all filters as NULL). 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. Minimum length of selIds is currently 2. However, if only one identifier is of interest this identifier can be at position one and two of the supplied vector which will lead to that the data for the identifier of interested will be plotted twice.

selContr Which contrast should be evaulated during the filtering, sorting and plotting? Descriptions of the contrasts can be found in the output from the anotaGetSigGenes object in the usedContrasts slot. Indicate the contrast by the column number.

`minSlope` The output can be filtered so that genes whose identified slopes are too small can be excluded. Default is NULL i.e. no filtering based on lower boundary of the slope. To exclude genes with e.g. a slope <(-1) assign -1 to minSlope.

`maxSlope` The output can be filtered so that genes whose identified slopes are too large can be excluded. Default is NULL i.e. no filtering based on upper boundary of the slope. To exclude genes with e.g. a slope >2 assign 2 to maxSlope.

`slopeP` A p-value for the slope being <0 or >1 is calculated if the estimate for the slope is <0 or >1. This p-value can be used to filter the output based on unrealistic slopes. When set low fewer genes will be disqualified. Default is NULL i.e. no filtering based on slope p-value. We recommend setting slopeP between 0.01 and 0.1 depending on data set characteristics.

`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 anota analysis without RVM (i.e. smaller compared to assigned value). Default is NULL i.e. no filtering.

`maxPAdj` The output can be filtered based on adjusted p-values from the anota analysis without RVM (i.e. smaller compared to assigned value). The adjustment method that was used when running anotaGetSigGenes will be evaluated. Default is NULL i.e. no filtering.

`maxRvmP` The output can be filtered based on raw p-values from the anota analysis with RVM (i.e. smaller compared to assigned value). Default is NULL i.e. no filtering.

`maxRvmPAdj` The output can be filtered based on adjusted p-values from the anota analysis with RVM (i.e. smaller compared to assigned value). The adjustment method that was used when running anotaGetSigGenes will be evaluated. Default is NULL i.e. no filtering.

`selDeltaPT` The output can be filtered based on the mean log2(translational activity data / cytosolic mRNA data) between groups difference. The groups are defined by the selected contrast. Default is NULL i.e. no filtering.

`selDeltaP` The output can be filtered based on the translational activity data only so that the minimum absolute between groups delta translation is used for gene inclusion. The groups are defined by the selected contrast. Default is NULL i.e. no filtering.

`sortBy` The output can be sorted by effect ("Eff"), raw p-value("p") or raw RVM p-value ("apvRvmP"). Default is NULL i.e. no sorting.

`performPlot` The function can generate a graphical output per gene. Default is TRUE i.e. generate plots.

`fileName` The plots are printed to a file whose file name is given here. Default is "ANOTA_selected_significant_genes_plot.pdf".

`geneNames` When anotaPlotSigGenes plots the individual gene plots they will be named by the original row names supplied to the anotaGetSigGenes 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

This function allows the user to filter the output generated from the anotaGetSigGenes function to derive a reduced selection of genes that are considered for further evaluation. This is done by setting one or several of the filtering parameters described above. The function also generates a graphical output which helps when evaluating a single gene's regulation. In the graphical output, the results for each gene is displayed on separate rows. The first graph shows all samples and per sample class regression lines using the common slope with different colors for each sample class. The magnitude of the common slope is indicated. The second graph shows key statistics for the gene without the RVM model for all contrasts analyzed when running anotaGetSigGenes but any ordering and selection of genes is performed on the contrast given by the selContr argument as described above. The third graph is similar to the second but with RVM statistics instead (if RVM was used in the anotaGetSigGenes analysis).

Value

anotaPlotSigGenes generates a graphical output as described above and a list object containing summary data for those genes that passed the selected set of filters. The output list object contains the following slots:

`selectedData` A matrix containing non-RVM data for the filtered identifiers. Columns are "apvSlope" (the common slope used in APV); "apvSlopeP" (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); "unadjustedResidError" (the residual error before calculating the effective residual error); "apvEff" (the group effect); "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).

`selectedRvmData`
A matrix containing RVM data for the filtered identifiers. Columns are "apvSlope" (the common slope used in APV); "apvSlopeP" (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); "apvEff" (the group effect); "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).

`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. Data for all groups defined when using the anotaGetSigGenes function are supplied although the filtering is based on the contrast defined under the selContr argument.

`deltaData` Mean delta translational activity data ("deltaP"), mean delta cytosolic mRNA data ("deltaT") and mean delta log ratio data ("deltaPT") comparing the sample classes specified by the selected contrast.

`usedThresholds` A list object with the user set values for the filtering.

Author(s)

Ola Larsson <ola.larsson@ki.se>, Nahum Sonenberg <nahum.sonenberg@mcgill.ca>, Robert Nadon <robert.nadon@mcgill.ca>

See Also

anotaPerformQc, anotaResidOutlierTestanotaGetSigGenes

Examples

```
 ## Load the library and dataset (two phenotypes)
  library(anota)
  data(anotaDataSet)
 ## Quality control of the data set.
  anotaQcOut <- anotaPerformQc(dataT= anotaDataT[1:200,], dataP=anotaDataP[1:200,],
phenoVec=anotaPhenoVec, nDfbSimData=500)
 ##Test normality of residuals
  anotaResidOut <- anotaResidOutlierTest(anotaQcObj=anotaQcOut)
 ##Identify differentially translated genes.
  anotaSigGeneOut <- anotaGetSigGenes(dataT= anotaDataT[1:200,],
dataP=anotaDataP[1:200,], phenoVec=anotaPhenoVec, anotaQcObj=anotaQcOut)
 ##Plot some of the differentially expressed mRNAs
  anotSigGeneOutFiltered <-
anotaPlotSigGenes(anotaSigObj=anotaSigGeneOut, selContr=1,
maxP=0.05,slopeP=0.05, maxSlope=1.5, minSlope=(-0.5), selDeltaPT=0.5)
```

---

anotaResidOutlierTest *Test for normality of residuals*

Description

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

Usage

```
anotaResidOutlierTest(anotaQcObj=NULL, confInt=0.01, iter=5,
generateSingleGraph=FALSE, nGraphs=200, generateSummaryGraph=TRUE,
residFitPlot=TRUE, useProgBar=TRUE)
```

Arguments

anotaQcObj	The object returned by anotaPerformQc.
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.

`generateSingleGraph`
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. Default is FALSE.

`nGraphs` If generateSingleGraph is set to TRUE, nGraphs controls for how many identifiers such single gene graphs will be generated.

`generateSummaryGraph`
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 anotaResidOutlierTest function assesses whether the residuals from the per identifier linear regressions of translationally active mRNA level~cytosolic mRNA level+phenoType are normally distributed. anota 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, anota 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 and Ripley 1999).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

anotaResdiOutlierTest generates a graphical output ("ANOTA_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 the generateSingleGraph is set to TRUE, the function also generates individual plots (stored as "ANOTA_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 "ANOTA_residuals_vs_fitted.jpeg"). An output list object with the following slots is also generated:

`confInt` The selected confInt (see function arguments).

`inputResiduals` The residuals used.

`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.

Author(s)

Ola Larsson <ola.larsson@ki.se>, Nahum Sonenberg <nahum.sonenberg@mcgill.ca>, Robert Nadon <robert.nadon@mcgill.ca>

Source

Modern Applied Statistics with S-PLUS. Venables, B.N. and Ripley, B.D., Springer. 1999

See Also

anotaPerformQc, anotaGetSigGenes, anotaPlotSigGenes

Examples

```
## See example for \code{\link{anotaPlotSigGenes}}
```

Index

∗ **datasets**
anotaDataSet, 2
∗ **methods**
anotaGetSigGenes, 3
anotaPerformQc, 5
anotaPlotSigGenes, 8
anotaResidOutlierTest, 11

anotaDataP *(*anotaDataSet*)*, 2
anotaDataSet, 2
anotaDataT *(*anotaDataSet*)*, 2
anotaGetSigGenes, 3, *8*, *11*, *13*
anotaPerformQc, *4*, 5, *11*, *13*
anotaPhenoVec *(*anotaDataSet*)*, 2
anotaPlotSigGenes, *4*, *8*, 8, *13*
anotaResidOutlierTest, *4*, *8*, *11*, 11