Package 'INCATome'

October 5, 2017

Type Package

Title Internal Control Analysis of Translatome Studies by Microarrays

Version 1.0

Date 2017-10-03

Author Sbarrato T. [cre,aut], Spriggs R.V. [cre,aut], Wilson L. [ctb], Jones C. [ctb], Dudek K. [ctb], Bastide A. [ctb], Pichon X. [ctb], Poyry T. [ctb] and Willis A.E. [ctb]

Maintainer Thomas Sbarrato <thomas.sbarrato@inserm.fr>

Description Data analysis, normalisation and differential expression for Translatome studies by microarrays (T Sbarrato et al. RNA. 2017 Aug 25; <DOI:10.1261/rna.060525.116>).

License CC BY-NC 4.0

RoxygenNote 6.0.1

Depends R (>= 2.10)

Imports limma, siggenes, multtest, genefilter, RankProd

LazyData true

NeedsCompilation no

R topics documented:

INCA.DEG   *INCATome Deregulated Genes Identification*

Description

Performs the INCATome DEG identification for microarray data, consisting of an overlap of at least two out of four DEG tests (TTest, Limma, RankProd and SAM).

Usage

```
INCA.DEG(x, cl, wcol, filt = TRUE, selneg, base = 2, highlight = NULL)
```

Arguments

x	an RGList object
cl	a vector specifying type of samples, 0 being control and 1 being condition.
wcol	an integer specifying the number of the column where Gene Names can be found in the gene annotation table.
filt	logical, TRUE if a set of negative control probes are to be used for filtering. Filtering is performed by removing any probes for which the average intensities are lower than the "negative" mean +/- 1.5 "negative" deviation.
selneg	a character or vector containing the negative control probe names for filtering.
base	an integer specifying the log base. Default is 2.
highlight	a character vector specifying a set of genes of interest. These will be highlighted in the graphical representations.

Value

A List object containing the INCA DEG output for significant DEGs with INCA DEG Score >= 2, as well as all individual outputs from the different tests. Additionally, volcanoplots for each test will be generated.

Examples

```
#Load the INCATome Dataset
data(INCATomeData)
attach(INCATomeData)
out=INCA.DEG(RGdataDS,c(0,0,0,1,1,1),8,filt=TRUE,
 selneg="NegativeControl", highlight=c("ACTB","PABPC1"))
```

INCA.DyeSwap *DyeSwap Method for two-color microarray data*

Description

Performs a dyeswap correction by an averaging method for two-color microarray data.

Usage

```
INCA.DyeSwap(x, dsvect)
```

Arguments

x   an RGList object

dsvect   an integer vector specifying dyeswapped microarrays. Needs to be of same
            length as number of arrays contained in the RGList object. Labelling should
            start from 1 and associates dyeswapped microarray with "-i".

Value

a new RGList object containing the dyeswapped array data.

Examples

```
#Load the INCATome Dataset
data(INCATomeData)
attach(INCATomeData)
ds=INCA.DyeSwap(RGdataNM,c(1,2,3,4,5,6,-1,-2,-3,-4,-5,-6))
```

INCA.MAPlot *MA Plot for Array data, with optional highlighting of set of genes*

Description

Plots MA plot for microarray data and highlights sets of genes and/or SpikeIn probes.

Usage

```
INCA.MAPlot(x, wcol, spikeIn = TRUE, SpikeFile, prefix = "",
  highlight = NULL)
```

Arguments

x   an RGList object

wcol   an integer specifying the number of the column where Gene Names can be found
            in the gene annotation table.

spikeIn   logical, TRUE to highlight SpikeIn Probes. Requires input in SpikeFile.

SpikeFile   a data.frame specifying the Spike In probe names if spikeIn=TRUE in a column
                   called "Probe" and the expected relative amounts for each dye, respectively in a
                   "Cy5" and "Cy3" column. For example, a given probe might be expected in a
                   3:1 ratio thus column "Cy5" would specify 3 and column "Cy3" would specify
                   1.

prefix   a character specifying the prefix to be used when saving the plot in a jpeg file.

highlight   a character vector specifying a set of genes of interest to be highlighted in the
                   plot.

Value

Generates jpeg files of MA plots for each arrays.

Examples

```
#Load the INCATome Dataset
data(INCATomeData)
attach(INCATomeData)
INCA.MAPlot(RGdata,8,spikeIn=TRUE,SpikeFile=sdata, highlight=c("ACTB","PABPC1"))
```

INCA.NormIC   *INCATome Normalisation by Internal Control Probes*

Description

Performs the INCATome normalisation using invariance of Internal Control probes selected by the
user for microarray data.

Usage

```
INCA.NormIC(x, InternalFile, wcol, base = 2, mva = TRUE)
```

Arguments

x   an RGList object

InternalFile   a data.frame specifying the names of the array files (in a column called "File-
                   Name") and the expected log2 ratios for two internal control genes selected by
                   the user (respectively in columns headed with the gene names). Expected log2
                   ratios are to be acquired experimentally, for each corresponding sample (typi-
                   cally by northern blotting or qPCR).

wcol   an integer specifying the number of the column where Gene Names can be found
                   in the gene annotation table.

base   an integer specifying the log base. Default is 2.

mva   logical, TRUE to plot MA plots before and after normalisation for each array.

Value

A new RGList object containing the normalised array data. Additionally, if mva is TRUE, MA plots
before and after normalisations will be generated for each arrays.

Examples

```
#Load the INCATome Dataset
data(INCATomeData)
attach(INCATomeData)
dc=INCA.NormIC(RGdataBG,idata,8)
```

INCA.NormSI   *INCATome Normalisation by Spike In Probes*

Description

Performs the INCATome normalisation using invariance of Spike In probes for microarray data.

Usage

```
INCA.NormSI(x, SpikeFile, wcol, base = 2, mva = TRUE, highlight = NULL)
```

Arguments

x	an RGList object
SpikeFile	a data.frame specifying the Spike In probe names in a column called "Probe" and the expected relative amounts for each dye, respectively in a "Cy5" and "Cy3" column. For example, a given probe might be expected in a 3:1 ratio thus column "Cy5" would specify 3 and column "Cy3" would specify 1.
wcol	an integer specifying the number of the column where Gene Names can be found in the gene annotation table.
base	an integer specifying the log base. Default is 2.
mva	logical, TRUE to plot MA plots before and after normalisation for each array.
highlight	a character vector specifying a set of genes of interest. These will be highlighted in the graphical representations.

Value

A new RGList object containing the normalised array data. Additionally, if mva is TRUE, MA plots before and after normalisations will be generated for each arrays.

Examples

```
#Load the INCATome Dataset
data(INCATomeData)
attach(INCATomeData)
dc=INCA.NormSI(RGdataBG,sdata,8,highlight=c("ACTB","PABPC1"))
```

INCA.PreProcess   *Background Correction of two-color microarray data*

Description

Performs a background correction by substraction method of two-color microarray data.

Usage

```
INCA.PreProcess(x, method, offset = 0)
```

Arguments

x	an RGList object
method	a character specifying the method to employ for background correction. Choices are: "subtract" or "normexp".
offset	a numerical value to add to intensities

Value

A new RGList object containing the background corrected array data. Of note, negative values generated from the correction are transformed to NA values.

Examples

```
#Load the INCATome Dataset
data(INCATomeData)
attach(INCATomeData)
db=INCA.PreProcess(RGdata,method="subtract")
```

INCA.SpikePlot   *Linearity Plot for Spike In Probes*

Description

Plots a linearity plot for Spike In probes for microarray data.

Usage

```
INCA.SpikePlot(x, SpikeFile, wcol, base = 2)
```

Arguments

x	an RGList object
SpikeFile	a data.frame specifying the Spike In probe names in a column called "Probe" and the expected relative amounts for each dye, respectively in a "Cy5" and "Cy3" column. For example, a given probe might be expected in a 3:1 ratio thus column "Cy5" would specify 3 and column "Cy3" would specify 1.
wcol	an integer specifying the number of the column where Gene Names can be found in the gene annotation table.
base	an integer specifying the log base. Default is 2.

Value

Generates jpeg files of SpikeIn Linearity plots for each arrays.

Examples

```
#Load the INCATome Dataset
data(INCATomeData)
attach(INCATomeData)
INCA.SpikePlot(RGdata,sdata,8)
```

INCATome   *Internal Control Analysis of Translatome Studies by Microarrays*

Description

Data analysis, normalisation and differential expression for Translatome studies by microarrays by means of a new statistical workflow which avoids interfering with data skewness in the identification of deregulated genes.

Details

INCATome Package Overview

Common microarray processing procedures (including for normalisation and statistical identification of deregulation) assume that deregulation must occur in low proportion (below 10%) and in equal symmetry (approx. equal number of upregulated and downregulated genes). However, we have shown that translatome studies in general violate these assumptions (Sbarrato *et al*, **RNA**, 2017 Aug 25, DOI:10.1261/rna.060525.116). This package can be implemented for the processing and statistical analysis of microarray datasets presenting inherent skewness due to the samples' nature causing violation of the aforementioned assumptions. INCATome workflow can be segmented as follows:

1. *Preprocessing and Quality Check:* INCATome workflow requires an RGList object containing the array data (for example output from `read.maimages` function of the limma package). First, the RGList is corrected for background with `INCA.PreProcess`, based on the limma package but insuring correct formatting of the output for the rest of the workflow. Users can select their correction method of choice in the arguments. Two graphical tools are at the user's disposal to perform quality checks on the data: 1) `INCA.MAPlot`, which allows to plot MA plots for each array and highlights a given set of control genes (SpikeIn probes and/or Internal References for example) and 2) `INCA.SpikePlot` which allows visual verification of linearity of SpikeIn probe signals on each array.

2. *Normalisation and Dyeswapping:* The normalization approach implemented with INCATome for translatome analysis is based on the root mean square deviation (RMSD) of internal controls. These can be represented by either the use of 1) the `INCA.NormSI` function requiring Spike-In controls that are independent of the sample and of known concentrations or by the use of 2) the `INCA.NormIC` function requiring Internal References chosen by the user and experimentally validated. The main advantage of this implementation is that the expected values for these given probes are already at hand to the user before the experiment is performed (Spike-In expected values given by Spike-In concentration ratios or Internal References expected values given by at least two northern blotting/qPCR quality controls for subpolysomal and polysomal associations i.e. ACTB and PABP respectively). As a consequence, the RMSD

values can be computed between expected and observed values for these probes in order to normalise the data. This procedure results in a within sample normalisation (to the expected levels of the given INCA probes for each sample) as well as a general scaling method across the samples (all tied to the same set of INCA probes). Finally, optional dyeswapping implementation by the INCA.DyeSwap function can be used whereby arrays dyeswapped will be reduced by averaging corresponding paired channels.

3. *Statistical Identification of Deregulation:* The aim of INCATome statistical pipeline with INCA.DEG is to reduce the false positive hits by combining four different statistical approaches to assess deregulation: a Welch T-Test rowFtests, the parametric Linear Models for Microarray (limma lmFit and eBayes), the non-parametric rank-based approach (RankProd RP) and the nonparametric variance-based Significance Analysis of Microarrays (SAM d.stat). The improved identification of significantly deregulated genes delivered by INCATome consists of selecting significant candidates (pvalue<=0.05) from each statistical test and assigning a confidence score corresponding to the number of tests concurring on the deregulation (high confidence: score=4; low confidence: score=2). Genes identified in only one test out of four implemented or with a fold change between -0.5 and 2 are not considered as being candidates for deregulation under INCATome implementation. Additionally, users have the opportunity prior of performing the statistical testing to filter out a set of genes which possess a ratio ranging between the (mean +/- 1.5*standard deviation) of negative control probes.

Author(s)

Sbarrato T. [cre,aut], Spriggs R.V. [cre,aut], Wilson L. [ctb], Jones C. [ctb], Dudek K. [ctb], Bastide A. [ctb], Pichon X. [ctb], Poyry T. [ctb] and Willis A.E. [ctb]

Maintainer: Thomas Sbarrato <thomas.sbarrato@inserm.fr>

References

Sbarrato T., Spriggs R.V., Wilson L., Jones C., Dudek K., Bastide A., Pichon X., Poyry T. and Willis A.E. **RNA**, (2017 Aug 25), An Improved Analysis Methodology for Translational Profiling by Microarray, DOI:10.1261/rna.060525.116

INCATomeData	*Translatome microarray dataset simulated from experimentally acquired translatome of Hela cells.*

Description

An RGList object for raw translatome data (based on one array and its dyesawapped array) simulated to produce n=12 arrays (n[CTRL]=3, n[CDT]=3, n[CTRLdyeswapped]=3, n[CDTdyeswapped]=3). Simulation was performed so that 25% of genes are deregulated and that this deregulation is skewed by 70% towards downregulation.

Usage

```
data("INCATomeData")
```

Format

The data contain a list of 6 objects: 4 RGLists and two dataframes.

RGdata an RGList object containing R, G, Rb, Gb, targets and source. The main data dimensions are ncol=12 arrays and nrow=2000 probes. The geneset can be fragmented as follows: from 1 to 1664: general probes, from 1665 to 1677: ACTB probes, from 1678 to 1680: PABPC1 probes and from 1681 to 2000: SpikeIn probes

RGdataBG an RGList object of background corrected data containing R, G, targets and source

RGdataNM an RGList object of INCA normalised data containing R, G, targets and source

RGdataDS an RGList object of dyeswapped data containing R, G, targets and source

idata a dataframe containing the Internal Reference (ACTB and PABPC1) Expected logged Ratios for each array as determined experimentally.

sdata a dataframe containing the SpikeIn Expected Ratios for each probe as defined experimentally by the manufacturer.

Source

Sbarrato T., Spriggs R.V., Wilson L., Jones C., Dudek K., Bastide A., Pichon X., Poyry T. and Willis A.E., **RNA**, 2017 Aug 25, An Improved Analysis Methodology for Translational Profiling by Microarray, DOI:10.1261/rna.060525.116

Examples

```
data(INCATomeData)
```

Index