

Package ‘anapuce’

April 17, 2009

Type Package

Title Tools for microarray data analysis

Version 2.1

Date 2009-02-18

Author J. Aubert (INRA)

Maintainer J. Aubert <julie.aubert@agroparistech.fr>

Description This package contains functions for normalisation, differentially analysis of microarray data and others functions implementing recent methods developed by the Statistic and Genom Team from UMR 518 AgroParisTech/INRA Appl. Math. Comput. Sc.

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2009-02-18 10:20:51

R topics documented:

anapuce-package	2
DiffAnalysis	2
DiffAnalysis.unpaired	4
est.varmixt	6
fdr.estimate.eta0	7
filterByDefault	8
Intensity.Norm	9
LBI	10
LocalFDR	10
MeanBySpot	11
normalisation	12

Index	14
--------------	-----------

anapuce-package *Tools for microarray data analysis*

Description

This package contains functions for normalisation, differentially analysis of microarray data and others functions implementing recent methods developed by the Statistic and Genom Team from UMR 518 AgroParisTech/INRA Appl. Math. Comput. Sc.

Details

Package: anapuce
Type: Package
Version: 2.1
Date: 2008-10-02
License: GPL-2
LazyLoad: yes

Author(s)

INRA, J. Aubert <julie.aubert@agroparistech.fr>

References

<http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:anapuce> <http://www.agroparistech.fr/mia/>

See Also

[normalisation](#), [DiffAnalysis](#), [DiffAnalysis.unpaired](#), [LBI](#), [LocalFDR](#)

DiffAnalysis *Differential Analysis for paired data*

Description

Performs the differential analysis from normalised paired microarray data according to different ways of variance modelling and computes lists of differentially expressed genes according different multiple testing procedures.

Usage

```
DiffAnalysis(fileIN = "resNorm.txt", n = 3, ind.array = 1:2, name.A = "A", name.M =
```

Arguments

<code>fileIN</code>	normalized data set.
<code>n</code>	number of identificant columns.
<code>ind.array</code>	the indices of arrays to analyze.
<code>name.A</code>	character string containing a regular expression contained in the columnnames corresponding to the A values.
<code>name.M</code>	character string containing a regular expression contained in the columnnames corresponding to the M values.
<code>fileOUT</code>	output data file.
<code>fileDelete</code>	output data file containing the list of withdrawn genes.
<code>procs</code>	adjusting p-values procedures to be used.
<code>alpha</code>	values of the first type error to be used in the different procedures.
<code>dyeswap</code>	logical indicating whether the design is a dye swap.
<code>indDS</code>	index of arrays related to dye swap.
<code>fileID</code>	file giving information about genes.
<code>function.trt</code>	function to be applied before differential analysis.
<code>by.var</code>	argument passed to <code>function.trt</code> .
<code>varmixt.meth</code>	logical indicating whether to perform the variance modelling proposed by Delmar et al. (2005).
<code>header</code>	a logical value indicating whether the file contains the names of the variables as its first line. Used in the read.table function.
<code>sep</code>	the field separator string to use in the read.table function.
<code>sep.write</code>	the field separator string to use in the write.table function.
<code>dec.write</code>	the string to use for decimal points in the write.table function.
<code>...</code>	Further arguments to be passed to <code>read.table</code> .

Details

This function performs a differential analysis in the gene-specific and homoscedastic cases. If `varmixt.meth = TRUE`, the method used is the method proposed by Delmar et al. (2005), (the same as coded in the `vm.analysis.paired` function from the `varmixt` package.)

Value

By default these following files are created

- one list of differentially expressed genes by procedure
- The list of withdrawn genes
- The list of all the genes

An R object is returned if function return value is assigned.

Author(s)

J. Aubert

References

Delmar, P., Robin, S. and Daudin, J.J., (2005), VarMixt: efficient variance modelling for the differential analysis of replicated gene expression data, *Bioinformatics*, **21**,(4), 502–8

Dudoit, S., Yang, Y. H., Callow, M. J. and Speed, T.P., (2002), Statistical methods for identifying differentially expressed genes in replicated cdna microarray experiments, *Statistica Sinica*, **12**, 111–139

See Also

[p.adjust](#), [MeanBySpot](#), [DiffAnalysis.unpaired](#)

DiffAnalysis.unpaired

Differential Analysis for unpaired data

Description

Performs the differential analysis from normalised unpaired data according to different ways of variance modelling and computes lists of differentially expressed genes according different multiple testing procedures.

Usage

```
DiffAnalysis.unpaired(fileIN = "resNorm.txt", n = 3, cond1 = "cond1.", cond2 = "con
```

Arguments

<code>fileIN</code>	normalized data set.
<code>n</code>	number of identificant columns.
<code>cond1</code>	a regular expression corresponding to the first condition.
<code>cond2</code>	a regular expression corresponding to the second condition.
<code>fileOUT</code>	output data file.
<code>fileDelete</code>	output data file containing the list of withdrawn genes.
<code>procs</code>	adjusting p-values procedures to be used.
<code>alpha</code>	values of the first type error to be used in the different procedures.
<code>fileID</code>	file giving information about genes.
<code>function.trt</code>	function to be applied before differential analysis.
<code>by.var</code>	argument passed to <code>function.trt</code> .
<code>varmixt.meth</code>	logical indicating whether to perform the variance modelling proposed by Delmar et al. (2005).

header	a logical value indicating whether the file contains the names of the variables as its first line. Used in the read.table function.
sep	the field separator string to use in the read.table function.
sep.write	the field separator string to use in the write.table function.
dec.write	the string to use for decimal points in the write.table function.
...	Further arguments to be passed to read.table.

Details

This function performs a differential analysis in the gene-specific and homoscedastic cases.

If `varmixt.meth = TRUE`, the method used is the method proposed by Delmar et al. (2005), (the same as coded in the `vm.analysis` function from the `varmixt` package.)

Value

By default these following files are created

- one list of differentially expressed genes by procedure
- The list of withdrawn genes
- The list of all the genes

An R object is returned if function return value is assigned.

Author(s)

J. Aubert

References

Delmar, P., Robin, S. and Daudin, J.J., (2005), VarMixt: efficient variance modelling for the differential analysis of replicated gene expression data, *Bioinformatics*, **21**,(4), 502–8

Dudoit, S., Yang, Y. H., Callow, M. J. and Speed, T.P., (2002), Statistical methods for identifying differentially expressed genes in replicated cdna microarray experiments, *Statistica Sinica*, **12**, 111–139

See Also

[p.adjust](#), [MeanBySpot](#), [DiffAnalysis](#)

 est.varmixt

Variance Mixture Estimation

Description

Performs variance mixture analysis

Usage

```
est.varmixt(VAR, Kmax, dfreedom)
```

Arguments

VAR	vector of estimated variance.
Kmax	maximal number of variance components.
dfreedom	degrees of freedom of the estimated variance.

Value

a LIST with the following components :

BIC.crit	value of the BIC criterion
p.i	the probability of each variance component
vars	variances
loglike	value of the criterion based on Loglikelihood
nmixt	number of variance components
tau	the matrix of posterior probability that a gene belongs to each variance component. One row per gene, one column per variance component.
VM2	the variance attributed to each gene according to the MAP rule - Delmar et al. (2005) <i>JRSS</i>
VM	the variance attributed to each gene taking into account the tau values - Delmar et al. (2005) <i>Bioinformatics</i>

Author(s)

M-L Martin-Magniette and J. Aubert

References

Delmar P, Robin S, Le Roux D, Daudin J.J (2005), Mixture model on the variance for the differential analysis of gene expression, *JRSS series C*, **54:1**, 31-50.

Delmar P, Robin S, Daudin J.J (2005), VarMixt: efficient variance modelling for the differential analysis of replicated gene expression data, *Bioinformatics*, **21(4)**, 502-8.

See Also

[DiffAnalysis](#), [DiffAnalysis.unpaired](#)

fdr.estimate.eta0 *Estimate the Proportion of Null p-Values*

Description

fdr.estimate.eta0 estimates the proportion η_0 of null p-values in a given vector of p-values. This quantity is an important parameter when controlling the false discovery rate (FDR). A conservative choice is $\eta_0 = 1$ but a choice closer to the true value will increase efficiency and power - see Benjamini and Hochberg (1995, 2000) and Storey (2002) for details.

Usage

```
fdr.estimate.eta0(p, method=c("conservative", "adaptive", "bootstrap", "smoother"),  
lambda=seq(0,0.95,0.05) )
```

Arguments

p	vector of p-values.
method	algorithm used to estimate the proportion of null p-values. Available options are "conservative" (default), "adaptive", "bootstrap", and "smoother".
lambda	optional tuning parameter vector needed for "bootstrap" and "smoothing" methods (defaults to <code>seq(0, 0.95, 0.05)</code>) - see Storey (2002) and Storey and Tibshirani (2003).

Details

The function `fdr.estimate.eta0` provides four algorithms: the "conservative" method always returns $\eta_0 = 1$ (Benjamini and Hochberg, 1995), "adaptive" uses the approach suggested in Benjamini and Hochberg (2000), "bootstrap" employs the method from Storey (2002), and "smoother" uses the smoothing spline approach in Storey and Tibshirani (2003).

Value

The estimated proportion η_0 of null p-values.

Author(s)

Konstantinos Fokianos (<http://www.ucy.ac.cy/~fokianos/>) and Korbinian Strimmer (<http://www.stat.uni-muenchen.de/~strimmer/>).

Adapted in part from S-PLUS code by Y. Benjamini (http://www.math.tau.ac.il/~roee/FDR_Splus.txt) and R code from J.D. Storey (<http://faculty.washington.edu/~jstorey/>).

References

"conservative" procedure: Benjamini, Y., and Y. Hochberg (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B*, **57**, 289–300.

"adaptive" procedure: Benjamini, Y., and Y. Hochberg (2000) The adaptive control of the false discovery rate in multiple hypotheses testing with independent statistics. *J. Behav. Educ. Statist.*, **25**, 60–83.

"bootstrap" procedure: Storey, J. D. (2002) A direct approach to false discovery rates. *J. Roy. Statist. Soc. B.*, **64**, 479–498.

"bootstrap" procedure: Storey, J. D., and R. Tibshirani (2003) Statistical significance for genome-wide experiments. *Proc. Nat. Acad. Sci. USA*, **100**, 9440-9445.

filterByDefault *Filter applied on data before normalisation*

Description

Apply a filter on input data by deleting spots with values for filter.var are mentionned in *filter.fic* by deleting spots with flags equal to *flag0* (-100 by default).

Usage

```
filterByDefault(fileIN, flag0, m, filter.fic = filter.fic, filter.var = filter.var,
```

Arguments

fileIN	input dataframe.
flag0	value(s) of flags to delete.
m	number of identificant columns.
filter.fic	name of the file containing the values of the variable named filter.var to delete.
filter.var	name of the variable concerned with deletion.
sep	the field separator string to use in the read.table function.
sep.write	the field separator string to use in the write.table function.
dec.write	the string to use for decimal points in the write.table function.
...	Further arguments to be passed to read.table.

Value

A subset of the input dataframe

Author(s)

J. Aubert

See Also

[normalisation](#)

Intensity.Norm *Function to compute normalized intensity for each channel*

Description

Computes normalized intensity for each channel and each array from the file containing normalized data.

Usage

```
Intensity.Norm(fileIN = "resNorm.txt", n = 3, ind.array = NULL, name.A = "A", name.M = "M", sep = "\t", center = FALSE, log.transf = FALSE, ...)
```

Arguments

fileIN	input dataframe.
n	number of identificant columns.
ind.array	indices of arrays to use.
name.A	character string containing a regular expression contained in the columnnames corresponding to the A values.
name.M	character string containing a regular expression contained in the columnnames corresponding to the M values.
sep	the field separator string to use in the read.table function.
center	logical indicating wether data will be centered by array.
log.transf	logical indicating wether data will stay log-transformed.
...	Further arguments to be passed to read.table.

Value

A R dataframe containing the n first columns of the input dataframe and the "normalized" intensity for the red and green channels of each chosen array.

Author(s)

J. Aubert

References

Thorne, N.P. (2004) Single-channel normalisation and analysis of twocolour cDNA microarray data, PhD thesis

See Also

[normalisation](#)

LBI *Label Bias Index*

Description

This function calculates the LBI (Label Bias Index) for two self-self hybridization microarrays.

Usage

```
LBI(infile, name.M = "M.norm", ind.array = 1:2, graph = TRUE, graphout = "FigM1M2")
```

Arguments

<code>infile</code>	input data file with normalized log-ratio base 2.
<code>name.M</code>	character string containing a regular expression contained in the columnnames corresponding to the M values.
<code>ind.array</code>	the indices of arrays to analyze.
<code>graph</code>	logical indicating wether to perform graphics.
<code>graphout</code>	name of the graphic output.

Author(s)

J. Aubert

References

Martin-Magniette M.L, Aubert J, Cabannes E, Daudin J.J (2005) Evaluation of the gene-specific dye bias in cDNA microarray experiments, *Bioinformatics*, **21(9)**, 1995-2000.

LocalFDR *Local FDR*

Description

Estimation of local false discovery rates

Usage

```
LocalFDR(dataf = dataf, graph = TRUE, method = NULL, lambda0 = 0.5, smoothing = "1")
```

Arguments

dataf	input data file with two columns (1 = gene name, 2= ordered raw pvalues).
graph	if TRUE a pdf graphic is created.
method	method for estimating m0. This must be one of the strings "adaptive", "conservative", "bootstap", "smoother" or NULL.
lambda0	value used in calculating m0.
smoothing	"1" for the initial published method, "2" for the PAVA method isotonic (monotonely increasing nonparametric) least squares regression - see P. Broberg (2005).
thres	threshold defining intervals used for the moving average smoothing.
mm	parameter defining intervals used for the moving average smoothing.

Value

- if graph = TRUE, a file of graphics named LocalFDRGraph.pdf
- A data file (LocalFDRFile.txt) with 4 columns : gene name, raw pvalues and two columns corresponding to smoothed FDR values.

Author(s)

J. Aubert

References

Aubert J, Bar-Hen A, Daudin J.J, Robin S (2004) Determination of the differentially expressed genes in microarrays experiments using local FDR, *BMC Bioinformatics*, **5:125**.

Aubert J , Bar-Hen A, Daudin J.J, Robin S (2005) Correction: Determination of the differentially expressed genes in microarray experiments using local FDR, *BMC Bioinformatics*, **6:42**.

Per Broberg (2005) A comparative review of estimates of the proportion unchanged genes and the false discovery rate, *BMC Bioinformatics* **6:199**

MeanBySpot

Mean By Spot

Description

Computes the mean on the different values of spots grouping by *by.var*.

Usage

```
MeanBySpot(fileIN, n = 3, name.A = "A", name.M = "M.norm", by.var = "ID")
```

Arguments

<code>fileIN</code>	input dataframe.
<code>n</code>	number of identificant columns.
<code>name.A</code>	character string containing a regular expression contained in the columnnames corresponding to the A values.
<code>name.M</code>	character string containing a regular expression contained in the columnnames corresponding to the M values.
<code>by.var</code>	name of the grouping variable.

Value

A R dataframe

Author(s)

J. Aubert

See Also

[DiffAnalysis](#), [DiffAnalysis.unpaired](#)

`normalisation` *Function to normalize microarray data*

Description

Performs the data normalization using a general loess and a block effect correction via the subtraction of the median by block if there is a Block column.

Usage

```
normalisation(fileIN = NULL, Red = "F635.Median", Green = "F532.Median", n = 3, fla
```

Arguments

<code>fileIN</code>	on optional regular expression contained in names of the files to analyze
<code>Red</code>	name of the column containing the Cy5 values
<code>Green</code>	name of the column containing the Cy3 values
<code>n</code>	number of identificant columns
<code>flag</code>	list of flags values to delete
<code>graph</code>	logical indicating wether to perform graphics
<code>filter.function</code>	function to perform before beginning the normalization

<code>filter.fic</code>	argument passed to <code>filter.funtion</code> - file containing the values of the variable named <code>filter.var</code> to delete
<code>filter.var</code>	argument passed to <code>filter.funtion</code>
<code>sep.write</code>	the field separator string to use in the <code>write.table</code> function
<code>dec.write</code>	the string to use for decimal points in the <code>write.table</code> function
<code>header</code>	a logical value indicating whether the file contains the names of the variables as its first line. Used in the <code>read.table</code> function
<code>sep</code>	the field separator string to use in the <code>read.table</code> function
<code>skip</code>	integer: the number of lines of the data file to skip before beginning to read data.
<code>...</code>	Further arguments to be passed to <code>read.table</code> .

Details

By default `fileIN=NULL` and the files to analyze have to be chosen among all the files in the working directory. If `fileIN` is not `NULL`, only file names which match the [regular expression](#) `fileIN` will be proposed to the normalization step.

Value

A file `resNorm.txt` containing the normalized data and if `graph=TRUE` a file `.pdf` per array with graphics

Author(s)

J. Aubert

References

Yang, Y., Dudoit, S., Luu, P., Lin, D., Peng, V., Ngai, J. and Speed, T., (2002), Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, *Nucleic Acids Research* **30**.

See Also

[filterByDefault](#), [read.table](#)

Index

- *Topic **htest**
 - fdr.estimate.eta0, 6
- *Topic **manip**
 - filterByDefault, 8
 - MeanBySpot, 11
- *Topic **package**
 - anapuce-package, 1
- *Topic **programming**
 - DiffAnalysis, 2
 - DiffAnalysis.unpaired, 4
 - est.varmixt, 5
 - Intensity.Norm, 9
 - LBI, 10
 - LocalFDR, 10
 - normalisation, 12

anapuce (*anapuce-package*), 1

anapuce-package, 1

DiffAnalysis, 2, 2, 5, 6, 12

DiffAnalysis.unpaired, 2, 4, 4, 6, 12

est.varmixt, 5

fdr.estimate.eta0, 6

filterByDefault, 8, 13

Intensity.Norm, 9

LBI, 2, 10

LocalFDR, 2, 10

MeanBySpot, 4, 5, 11

normalisation, 2, 8, 9, 12

p.adjust, 4, 5

read.table, 3, 4, 8, 9, 13

regular expression, 2, 4, 9, 10, 12, 13

write.table, 3, 4, 8, 13