Package 'prot2D'

April 23, 2016

Type Package

Title Statistical Tools for volume data from 2D Gel Electrophoresis

Version 1.8.0

Date 2013-10-25

Author Sebastien Artigaud

Maintainer Sebastien Artigaud <sebastien.artigaud@gmx.com>

Description The purpose of this package is to analyze (i.e. Normalize and select significant spots) data issued from 2D GEl experiments

Depends R (>=

2.15),fdrtool,st,samr,Biobase,limma,Mulcom,impute,MASS,qvalue

Suggests made4,affy

License GPL (≥ 2)

biocViews DifferentialExpression, MultipleComparison, Proteomics

NeedsCompilation no

R topics documented:

prot2D-package	2
ES.prot	3
Norm.qt	4
Norm.vsn	5
pecten	6
pecten.fac	
RIplot	7
robust.fdr	9
Sim.Prot.2D	0
variousT.Prot	12

Index

15

prot2D-package

Description

The purpose of this package is to analyze (i.e. Normalize and select significant spots) data issued from 2D GEI experiments

Details

This package provides a simple interface for analysing data from 2D gel experiments. Functions for performing normalization as well as selecting significant spots are provided. All the functions for selecting significant spots are adapted from functions for microarray analysis provided by Bioconductor and CRAN, and all credits go to the authors of these functions. For analyzing 2D gel experiments data, users are advised to follow theses steps :

- 1. Normalize data using Norm.qt
- 2. Coerce data into an ExpressionSet using ES.prot
- 3. Use normalized data to find differentially expressed proteins with FDR-controled functions: modT.Prot was find to be the more efficient for 2-DE (see Artigaud et al , 2013) but other functions are provided (ttest.Prot,samT.Prot,efronT.Prot, ,shrinkT.Prot)

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

References

- Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.
- Dudoit, S., Yang, Y.H., Callow, M.J., & Speed, T.P. (2002) "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments" Statistica Sinica, vol. 12: 111-139.
- Strimmer, K. (2008) "A unified approach to false discovery rate estimation." BMC Bioinformatics, vol. 9: 303.

See Also

ExpressionSet, fdrtool

ES.prot

Examples

```
data(pecten)
data(pecten)
data(pecten.fac)

pecten.norm <- Norm.qt(pecten, n1=6, n2=6, plot=TRUE) #Quantiles normalization of the data
ES.p <- ES.prot(pecten.norm, n1=6, n2=6, f=pecten.fac)
x <- modT.Prot(ES.p, fdr.thr=0.1, plot=TRUE)
featureNames(x) # Names of the spots selected for a moderated t-test with a fdr of 0.1
fData(x) # Displaying fold change (as log2(ratio)) for selected spots
exprs(x) # Normalized volume data for all the selected spots
## Not run: heatplot(x) #Great heatmap of the selected spots (requires made4 Bioconductor package )</pre>
```

ES.prot

Create an ExpressionSet containing 2D Gel Volume Data

Description

This function create an an ExpressionSet containing 2D Gel Volume Data and associated metadata. It requires a matrix of 2D Gel Volume data with Gels as columns and Spots as rows as well as a dataframe describing the data.

Usage

ES.prot(data, n1, n2, f)

Arguments

data	a matrix of spots intensities. data should be intensities displayed with gels as columns with the name of columns corresponding to the names of the gels and spots as rows with the names of the rows corresponding to the name of the spots. The replicates for each condition should be ordered in following columns.
n1	an integer. Number of replicates in condition 1.
n2	an integer. Number of replicates in condition 2.
f	a dataframe giving the factors for data. f dataframe should have only 1 column giving the 2 levels of factor with rownames corresponding to the names of gels (in the same order as data).

Details

Usually data matrix should be the output of Norm. qt and therefore is log2-transformed. In data the columns (i.e. Gels) must be ordered with replicates from condition 1 before those from condition 2.

Value

The function returns an ExpressionSet. The matrix of spots intensities (i.e. Volume) is stored in the assayData slot of the ExpressionSet and can be retrieved with exprs. In the featureData slot the log2-ratio is computed and can be retrieved with fData.

Norm.qt

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

See Also

Norm.qt,Norm.vsn,ExpressionSet

Examples

data(pecten)
data(pecten.fac)

pecten.norm <- Norm.qt(pecten, n1=6, n2=6, plot=TRUE) #Quantiles normalization of the data ES.p <- ES.prot(pecten.norm, n1=6, n2=6, f=pecten.fac) head(exprs(ES.p))

```
Norm.qt
```

Normalize 2D Gel Volume data using Quantiles Normalization

Description

This function allows to normalize 2D Gel Volume data using the Quantiles Normalization.

Usage

```
Norm.qt(data, n1, n2, plot = T)
```

Arguments

data	a dataframe of raw 2D Gel Volume data. data should be raw intensities dis- played with gel as columns with the name of columns corresponding to the names of the gels and spots as rows with the names of the rows corresponding to the name of the spots. The replicates for each condition should be ordered in following columns.
n1	an integer. Number of replicates in condition 1.
n2	an integer. Number of replicates in condition 2.
plot	logical. if TRUE (default) displaying two RIplot, one with the raw data, another with normalized data.

Details

2D Gel Volume data must be normalized in order to remove systemic variation prior to data analysis. The principle of the "quantiles normalization" is to set each quantile of each column (i.e. the spots volume data of each gels) to the mean of that quantile across gels. The intention is to make all the normalized columns have the same empirical distribution. This function is based on normalizeQuantiles from limma package.

Norm.vsn

Value

The function returns a matrix of log2 transformed quantiles normalized data

Author(s)

Sebastien.artigaud@gmx.com>

References

Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.

See Also

normalizeQuantiles,RIplot,ES.prot

Examples

data(pecten)

pecten.norm <- Norm.qt(pecten, n1=6, n2=6, plot=TRUE)</pre>

Norm.vsn

Normalize 2D Gel Volume data using VSN

Description

This function allows to normalize 2D Gel Volume data using the "Variance Stabilizing Normalization".

Usage

```
Norm.vsn(data, n1, n2, plot = T)
```

Arguments

data	a dataframe of raw 2D Gel Volume data. data should be raw intensities dis- played with gel as columns with the name of columns corresponding to the names of the gels and spots as rows with the names of the rows corresponding to the name of the spots. The replicates for each condition should be ordered in following columns.
n1	an integer. Number of replicates in condition 1.
n2	an integer. Number of replicates in condition 2.
plot	logical. if TRUE (default) displaying two RIplot, one with the raw data, another with normalized data.

Details

The "Variance Stabilizing Normalization" relies on a transformation h, of the parametric form h(x) = arsinh(a+bx) (for details see, Huber et al., 2002). The parameters of h together with those of the calibration between experiments are estimated with a robust variant of maximum-likelihood estimation. This function is based on normalizeVSN from limma package.

Value

The function returns a matrix of vsn normalized data

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

References

- Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.
- Huber, W., Heydebreck, von, A., Sultmann, H., Poustka, A., & Vingron, M. (2002) "Variance stabilization applied to microarray data calibration and to the quantification of differential expression" Bioinformatics, vol. 18 (Suppl 1): S96-S104.

See Also

normalizeVSN,RIplot,ES.prot

Examples

data(pecten)

pecten.norm <- Norm.vsn(pecten, n1=6, n2=6, plot=TRUE)</pre>

pecten

Pecten Dataset

Description

Raw Volume data from a 2-DE experiment performed on proteins from *Pecten maximus* gills subjected to a temperature challenge (15C = control vs 25C). 766 spots were identified with 6 replicates per condition, therefore the dataset is a dataframe of 12 columns and 766 rows (for details, see Artigaud et al, 2013).

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

pecten.fac

References

Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.

pecten.fac pecten.fac Dataset

Description

Dataframe with 1 column describing factors for pecten dataset. It gives the names of the gels (as rownames) and the condition for the temperature challenge (15C = control vs 25C) in the "Condition" column (for details, see Artigaud et al, 2013).

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

References

Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.

RIplot

Ratio-Intensity Plot for 2D Gel Volume data

Description

R-I plots consist in plotting the intensity log2-ratio (R) against mean log10 intensity (I) of spot's volume data from 2D Gel experiments.

Usage

RIplot(data, n1, n2, ...)

Arguments

data	either a dataframe or a matrix of 2D Gel Volume data. data should be spots intensities displayed with gel as columns with the name of columns correspond- ing to the names of the gels and spots as rows with the names of the rows corre- sponding to the name of the spots. The replicates for each condition should be ordered in following columns.
n1	an integer. Number of replicates in condition 1.
n2	an integer. Number of replicates in condition 2.
•••	Further arguments to be passed to methods, such as graphical parameters (see par).

Details

Dudoit et al (2002) proposed a method for visualization of artifacts in microarray datasets, called the MA-plot, which was transposed for proteomics data as the Ratio-Intensity plot by Meunier et al (2005). R-I plots allow to directly visualize artifacts in the original data set as well as the effects of normalization. The log2-ratio (R) is the log2 of the mean of volume data in condition 2 upon the mean of volume data in condition 1:

$$R = \log_2 \frac{mean(V_{Cond2})}{mean(V_{Cond1})}$$

The intensity (I) is the log10 of the mean of volume data in condition 2 by the mean of volume data in condition 1

$$I = log_10(mean(V_{Cond2}) \times mean(V_{Cond1}))$$

Value

Display the RI-plot of data, the values on the plot are the number of spots with a Ratio greater then 1 (upper values) or lesser then -1 (lower value). A list is invinsibly returns with components :

RI	a dataframe with ratio and intensity values calculated.
diff	a vector with number of spots with a Ratio greater then 1 or lesser then -1 (in that order).

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

References

- Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.
- Dudoit, S., Yang, Y.H., Callow, M.J., & Speed, T.P. (2002) "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments" Statistica Sinica, vol. 12: 111-139.
- Meunier, B., Bouley, J., Piec, I., Bernard, C., Picard, B., & Hocquette, J.-F. (2005) "Data analysis methods for detection of differential protein expression in two-dimensional gel electrophoresis" Analytical Biochemistry, vol. 340 (2): 226-230.

See Also

Norm.qt,Norm.vsn

Examples

data(pecten)

RIplot(pecten, 6, 6, main="Without Normalization")

robust.fdr

Description

Implements robust method of FDR estimation (Pounds and Cheng 2006, Bioinformatics)

Usage

```
robust.fdr(p, sides = 1, p2 = 1 - p, discrete = F, use8 = T)
```

Arguments

р	vector of p-values from the analysis.
sides	indicate whether p-values are 1-sided (set sides=1) or 2-sided (set sides=2), de-fault=1.
p2	for one-sided testing, p-values from testing the "other alternative", default=1-p.
discrete	logical. Indicates whether p-values are discrete
use8	indicates whether the constant 8 should be used if p-values are discrete, see Pounds and Cheng (2006) for more details.

Details

This function uses the code from Stan Pounds available at http://www.stjuderesearch.org/ depts/biostats/documents/robust-fdr.R and is included in prot2D package for convenience and comparison purpose.

Value

A list with components:

р	the vector of p-values provided by the user.
fdr	the vector of smoothed FDR estimates.
q	the vector of q-values based on the smoothed FDR estimates.
cdf	the vector with p-value empirical distribution function at corresponding entry of p.
loc.fdr	the local (unsmoothed) FDR estimates.
fp	the estimated number of false positives at p-value cutoff in p.
fn	the estimated number of false negatives at p-value cutoff in p.
te	the total of fp and fn.
pi	the null proportion estimate.
ord	a vector of indices to order the vectors above by ascending p-value.

Author(s)

Stan Pounds. Edited by Sebastien Artigaud for prot2D package.

References

Pounds, S. & Cheng, C. (2006) "Robust estimation of the false discovery rate" Bioinformatics, vol. 22 (16): 1979-1987.

Sim.Prot.2D	Simulate 2D Gel volume data	
-------------	-----------------------------	--

Description

This function simulates realistic 2D Gel volume data, based on parameters estimate from real dataset.

Usage

```
Sim.Prot.2D(data, nsp=nrow(data), nr=10, p0=0.1, s2_0=0.2 ,d0=3)
```

Arguments

data	either a dataframe or a matrix of 2D Gel Volume data. Usually the output of Norm.qt.
nsp	an integer. Number of spots (i.e. rows) to generate (should be equal or less than the number of rows of data).
nr	an integer. Number of replicates for each conditions.
p0	proportion of spots to be differentially expressed (defaults to 0.1).
s2_0	estimator of the standard deviation of the spots distribution (defaults to 0.2).
d0	degree of freedom of the Chi-squared distribution used in the calculation of the distribution (defaults to 3).

Details

Volume data are computed following these steps (see Smyth, 2004 and Artigaud et al, 2013 for details) :

- 1. Log2 mean volumes from data are computed for each spot.
- 2. Means are used as input parameters in order to simulate a normal distribution (with no differential expression between conditions) for each spots with standard deviations, computed as described by Smyth (2004).
- 3. p0 of the spots are randomly picked for introducing differential expression in both conditions (p0/2 in each condition).

If nsp is less than the number of rows in data, the nsp first columns of data are used as input parameters.

Sim.Prot.2D

Value

returns an ExpressionSet of simulated volume data (log2 transformed) with 2 conditions ("Cond1" and "Cond2" in phenoData) slot of the ExpressionSet.The spots differentially generated can be retrieve with notes.

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

References

- Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.
- Smyth, G.K. (2004) "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." Statistical Applications in Genetics and Molecular Biology, vol. 3: Article 3.

See Also

variousT.Prot,

Examples

```
#Simulate data based on "pecten"
data(pecten)
data(pecten.fac)
pecten.norm <- Norm.qt(pecten, n1=6, n2=6, plot=FALSE)</pre>
Sim.data <- Sim.Prot.2D(data=pecten.norm, nsp=700, nr=10, p0=0.1, s2_0=0.2, d0=3)
#Compare different methods for finding diffenrentially expressed proteins
res.stud <- ttest.Prot(Sim.data, fdr.thr=0.1, plot=FALSE)
res.mo <- modT.Prot(Sim.data, fdr.thr=0.1, plot=FALSE)</pre>
a <- featureNames(res.stud)</pre>
a # Names of the spots selected for a student's t-test with an FDR of 0.1
b <- featureNames(res.mo)</pre>
b # Names of the spots selected for a modT-test with an FDR of 0.1
c <- notes(Sim.data)$SpotSig</pre>
c # Names of the really differentially generated spots
# Compare between t-test and moderate t-test (requires made4 package)
# Just for illustration purpose (only 1 simulation !)
## Not run:
cl.t <- comparelists(c,a)</pre>
cl.m <- comparelists(c,b)</pre>
fp <- matrix(c(c(length(a)- (cl.t$Length.Y-length(cl.t$intersect))),</pre>
c(cl.t$Length.Y-length(cl.t$intersect)),
c(length(b)- (cl.m$Length.Y-length(cl.m$intersect))),
```

```
c(cl.m$Length.Y-length(cl.m$intersect))),
nrow=2,ncol=2,byrow=F)
barplot(fp, main="spots declared significant",names.arg=c("t-test","modT"))
legend("topleft",c("False Positives","True Positives"), fill=c("grey","black"), bty="n")
## End(Not run)
```

variousT.Prot Various Tests for finding differentially expressed proteins in 2-DE experiments

Description

These functions provides simple methods for finding differentially expressed proteins in 2-DE experiments.

Usage

ttest.Prot(data, plot = T, fdr.thr = 0.1, Fold2 = F, method.fdr = "BH", var.equal = F)
modT.Prot(data, plot = T, fdr.thr = 0.1, Fold2 = F, method.fdr = "BH", col=1)
samT.Prot(data, plot = T, fdr.thr = 0.1, Fold2 = F, method.fdr = "BH")
efronT.Prot(data, plot = T, fdr.thr = 0.1, Fold2 = F, method.fdr = "BH")
shrinkT.Prot(data, plot = T, fdr.thr = 0.1, Fold2 = F, method.fdr = "BH", var.equal = F)

Arguments

data	an ExpressionSet of volume data. Usually as returned by ES.prot.
fdr.thr	threshold value for the FDR (defaults to 0.1).
method.fdr	method for FDR estimator. 4 methods are implemented :"BH" for the Classi- cal FDR from Benjamini & Hochberg, "Strimmer", "Storey" and "Pounds" (see references for details).
Fold2	logical. If true only spots with an absolute ratio of 2 are selected.
plot	logical. If true (default) draws a plot fdr values depending on the p-values with a line indicating the FDR threshold.
var.equal	logical. If false (default) assume unequal variances in each condition and uses a proper correction. Only for ttest.Prot and shrinkT.Prot.
col	An integer (default to 1). Only for modT.Prot. If there is more than one factor in pData, indicates the column to use for the analysis.

12

variousT.Prot

Details

As described in Artigaud et al (2013) these are functions adapted from microarray analysis. 2-DE experiments analysis requires a variant of the t-statistic that is suitable for high-dimensional data and large-scale multiple testing. For this purpose, in the last few years, various test procedures have been suggested.

These functions provides:

- the classical Student's t-test (adapted from studentt.stat).

- two tests especially modified for micro-array analysis : Efron's t-test (adapted from efront.stat, Efron et al, 2001) and the modified t-test used in Significance Analysis for Microarray (adapted from samr, Tusher et al, 2001)

- two methods that take advantage of hierarchical Bayes methods for estimation of the variance across genes: the moderate t-test from Smyth (using limma; see Smyth, 2004) and the "Shrinkage t" statistic test from Opgen-Rhein & Strimmer (adapted from shrinkcat.stat; see Opgen-Rhein & Strimmer, 2007).

As statistical tests allowing the identification of differentially expressed proteins must take into account a correction for multiple tests in order to avoid false conclusions. These functions also provides different methods to estimate the False Discovery Rate :

- the classical FDR estimator of Benjamini & Hochberg (using p.adjust; see Benjamini & Hochberg, 1995)

- the Fdr estimator of Strimmer (based on local fdr calculation) (using fdrtool; see Strimmer 2008)

- the "robust FDR" estimator of Pounds & Cheng (implemented in robust.fdr for the prot2D package; see Pounds & Cheng, 2006)

- Fdr method of Storey and Tibshirani (2003), also known as "q-values" (using qvalue.

Value

returns an ExpressionSet containing only the significant spots (see Examples).

Author(s)

Sebastien Artigaud <sebastien.artigaud@gmx.com>

References

- Artigaud, S., Gauthier, O. & Pichereau, V. (2013) "Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools." Bioinformatics, vol.29 (21): 2729-2734.
- Benjamini, Y. & Hochberg, Y. (1995) "Controlling the false discovery rate: a practical and powerful approach to multiple testing" Journal of the Royal Statistical Society. Series B. Methodological.: 289-300.
- Efron, B., Tibshirani, R., Storey, J.D., & Tusher, V. (2001) "Empirical Bayes Analysis of a Microarray Experiment" Journal of the American Statistical Association, vol. 96 (456): 1151-1160.
- Tusher, V.G., Tibshirani, R., & Chu, G. (2001) "Significance analysis of microarrays applied to the ionizing radiation response"" Proceedings of the National Academy of Sciences of the United States of America, vol. 98 (9): 5116-5121.

- Smyth, G.K. (2004) "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." Statistical Applications in Genetics and Molecular Biology, vol. 3: Article 3.
- Pounds, S. & Cheng, C. (2006) "Robust estimation of the false discovery rate" Bioinformatics, vol. 22 (16): 1979-1987.
- Strimmer, K. (2008) "A unified approach to false discovery rate estimation." BMC Bioinformatics, vol. 9: 303.
- Opgen-Rhein, R. & Strimmer, K. (2007) "Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach" Statistical Applications in Genetics and Molecular Biology, vol. 6 (1).

See Also

Norm.qt,ES.prot,fdrtool,limma, samr,studentt.stat,shrinkt.stat, efront.stat,qvalue

Examples

data(pecten)
data(pecten.fac)

```
pecten.norm <- Norm.qt(pecten, n1=6, n2=6, plot=TRUE) #Quantiles normalization of the data
ES.p <- ES.prot(pecten.norm, n1=6, n2=6, f=pecten.fac)
ES.diff <- modT.Prot(ES.p, fdr.thr=0.1, plot=TRUE)
featureNames(ES.diff) # Names of the spots selected for a moderated t-test with a fdr of 0.1
fData(ES.diff) # Displaying fold change (as log2(ratio)) for selected spots
exprs(ES.diff) # Volume normalized data for all the selected spots
## Not run: heatplot(ES.diff) #Great heatmap of the selected spots (require made4 Bioconductor package )
```

Index

*Topic **2DGel** prot2D-package, 2 *Topic **biology** prot2D-package, 2 *Topic **package** prot2D-package, 2 *Topic prot2D ES.prot, 3 Norm.qt, 4Norm.vsn, 5 pecten, 6 pecten.fac, 7 prot2D-package, 2 RIplot, 7 robust.fdr,9 Sim.Prot.2D, 10 variousT.Prot, 12 *Topic **proteomic** prot2D-package, 2 assayData, 3 efronT.Prot, 2 efronT.Prot (variousT.Prot), 12 efront.stat, *13*, *14* ES.prot, 2, 3, 5, 6, 12, 14 ExpressionSet, 2-4, 11-13 exprs, 3fData, 3 fdrtool, 2, 13, 14 featureData, 3

limma, 13, 14

modT.Prot, 2
modT.Prot (variousT.Prot), 12

Norm.qt, 2-4, 4, 8, 10, 14 Norm.vsn, 4, 5, 8 normalizeQuantiles, 4, 5 normalizeVSN, 6 notes, 11 p.adjust, 13 par, 7 pData, 12 pecten, 6, 7pecten.fac,7 phenoData, 11 prot2D, 9, 10, 13 prot2D (prot2D-package), 2 prot2D-package, 2 qvalue, 13, 14 RIplot, 4-6, 7 robust.fdr, 9, 13 samr, 13, 14 samT.Prot, 2 samT.Prot(variousT.Prot), 12 shrinkcat.stat, 13 shrinkT.Prot, 2 shrinkT.Prot(variousT.Prot), 12 shrinkt.stat, 14 Sim.Prot.2D, 10 studentt.stat, 13, 14 ttest.Prot, 2 ttest.Prot (variousT.Prot), 12

variousT.Prot, 11, 12