Package

March 29, 2019

Type Package

Title Variance-Preserving Estimation and Normalization of M-A Values

Version 1.0

Author Adèle Helena Ribeiro

Maintainer Adèle Helena Ribeiro <adele@ime.usp.br>

Description This package provides methods for estimating and normalizing the
M (intensity log-ratio) and A (mean log intensity) values from two-channel
(or two-color) microarrays.
Unlike conventional estimation methods which take into account only measures of location
(e.g., mean and median) of the pixel intensities of each channel,
the provided estimation method takes into account pixel-level variability,
which may reflects uncertainties due noise and systematic artifacts.
To remove array-specific effects, intensity-dependent dye biases,
and other systematic trends of the microarray data, the M and A values have to be subjected
to a within-slide normalization. The most used within-slide normalization technique is LOWESS.
However, the choice of the LOWESS parameters, particularly the smoothing neighborhood parameter
(or bandwidth), critically affects the quality of the microarray data normalization.
Thus, to preserve relevant variation that may be removed in LOWESS normalization
with arbitrarily chosen parameters, it is provided a parameter selection method that is
parsimonious and considers intrinsic characteristics of microarray data, such as heteroskedasticity.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports limma, locfit, Rdpack, methods

Suggests knitr, rmarkdown, maigesPack, lattice

Depends R (>= 2.10)

RdMacros Rdpack

RoxygenNote 6.1.1

R topics documented:

`estimateMAValues` *Estimation of the M and A Values*

Description

Estimates the intensity values M and A, considering p spots/genes and n microarray slides, by using the conventional method (Yang et al. 2002) or the one that takes into account pixel-level variability (Ribeiro et al. 2019).

Usage

```
estimateMAValues(R.mean, G.mean, R.var, G.var, RG.cov, R.bckg, G.bckg,
  estimator = "second-order", bgcorr = "normexp", offset = 50,
  putaway.perc = 0.05, array.ids = NULL, gene.ids = NULL)
```

Arguments

`R.mean`	A p x n matrix of mean values for the red (test) foreground intensities.
`G.mean`	A p x n matrix of mean values for the green (reference) foreground intensities.
`R.var`	A p x n matrix of variances of the red foreground intensities.
`G.var`	A p x n matrix of variances of the green foreground intensities.
`RG.cov`	A p x n matrix of covariances between the green and red foreground intensities.
`R.bckg`	A p x n matrix of mean values of the red background intensities.
`G.bckg`	A p x n matrix of mean values of the green background intensities.
`estimator`	A character string indicating the estimation method. The options are 'conventional' and 'second-order'. If estimator is set as 'second-order', then an estimator that takes into account pixel-level variability is used.
`bgcorr`	A character string indicating the background correction method. The options are 'none' and 'normexp'.
`offset`	A numeric parameter for the normexp background correction.
`putaway.perc`	A scalar between 0 and 100 indicating percentage of genes expressions. with high variance to be discarded. Default is 0.05.
`array.ids`	A character vector with identifiers of the arrays.
`gene.ids`	A character vector with identifiers of the genes.

Value

A limma's MAlist object with the M and A values

References

Ribeiro AH, Soler JMP, Hirata Jr R (2019). "Variance-Preserving Estimation of Intensity Values Obtained from Omics Experiments." *Submitted for publication*.

Yang YH, Buckley MJ, Dudoit S, Speed TP (2002). "Comparison of methods for image analysis on cDNA microarray data." *J Comput Graph Stat*, **11**(1), 108–136.

Examples

```
data(metaplasia)
MA <- estimateMAValues(metaplasia$R.mean, metaplasia$G.mean,
                       metaplasia$R.var, metaplasia$G.var, metaplasia$RG.cov,
                       metaplasia$R.bckg, metaplasia$G.bckg,
                       array.ids=metaplasia$array.ids, gene.ids=metaplasia$gene.ids)
```

`getOptimalLowessFvalue` *Optimal Selection of the LOWESS Bandwidth Parameter Using the HRCp Criterion*

Description

LOWESS parameter selection method, proposed by (Ribeiro et al. 2019), that is parsimonious and considers intrinsic characteristics of microarray data, such as heteroskedasticity. Particularly, the best bandwidth is selected according to the HRCp criterion (Liu and Okui 2013).

Usage

```
getOptimalLowessFvalue(M.mean, A.mean, eva.values = seq(0.2, 1, by =
  0.05), debug = TRUE, save.objs = TRUE, save.plots = TRUE,
  dir.to.save = "./", array.id = "array")
```

Arguments

`M.mean`	A numeric vector with the M values for the p genes.
`A.mean`	A numeric vector with the A values for the p genes.
`eva.values`	A numeric vector with bandwith values to be considered by the LOWESS parameter selection method.
`debug`	A logical value indication if you want to view logs of the execution.
`save.objs`	A logical value indicating if you want to save objects with partial results.

`save.plots`	A logical value indicating if you want to generate the M plots illustrating the bandwidth parameter selection process.
`dir.to.save`	Path to the folder you want to save the output objects.
`array.id`	A character string identifying the array.

Value

A list with the following elements:

opt.value The optimal bandwith.

criterion.val A list with the mean square errors for the estimates obtained for each evaluated bandwith.

criterion.df A list with the effective degrees of freedom for the estimates obtained for each evaluated bandwith.

fits A list with the estimates obtained for each evaluated bandwith.

References

Liu Q, Okui R (2013). "Heteroscedasticity-Robust Cp Model Averaging." *The Econometrics Journal*, **16**(3), 463–472.

Ribeiro AH, Soler JMP, Hirata Jr R (2019). "Variance-Preserving Estimation of Intensity Values Obtained from Omics Experiments." *Submitted for publication*.

`metaplasia` *Intestinal Metaplasia Dataset*

Description

This is part of the dataset used in the study by (Ribeiro et al. 2019).

Usage

```
metaplasia
```

Format

A list with the following elements:

R.mean A 45015 x 10 matrix with the means of the red (test) foreground pixels in each spot of each array.

G.mean A 45015 x 10 matrix with the means of the green (reference) foreground pixels in each spot of each array.

R.var A 45015 x 10 matrix with the variances across red foreground pixels in each spot of each array.

G.var A 45015 x 10 matrix with the variances across green foreground pixels in each spot of each array.

RG.cov A 45015 x 10 matrix with the covariances between green and red foreground pixels in each spot of each array.

R.bckg A 45015 x 10 matrix with the means of the red background pixels in each spot of each array.

G.bckg A 45015 x 10 matrix with the means of the green background pixels in each spot of each array.

R.size A 45015 x 10 matrix with the number of red foreground pixels in each spot of each array.

G.size A 45015 x 10 matrix with the number of green foreground pixels in each spot of each array.

array.ids A vector with the ids for the 10 arrays.

gene.ids A vector with the ids for the 45015 spots.

gene.names A vector with the names of the 45015 genes.

Details

It contains data from 10 two-color microarrays, being 5 from tissues representing type II intestinal metaplasia and 5 from tissues representing the normal condition, obtained from the Tumor Bank at A.C. Camargo Cancer Center / Antonio Prudente Foundation.

Each sample is hybridized against a pool of normal tissues using the same orientation of dye labeling. Gene expression levels were measured on Agilent Whole Human Genome Microarrays 4x44K G4112F (design ID 014850).

Each slide contais 45015 spots (41093 unique probes) and each spot contains about 60 foreground pixels.

The scanned images of the microarray slides were processed by Agilent Feature Extraction software, version 9.5, where statistics (mean, standard deviation and covariance) of the foreground and local background pixels were computed for each spot, in both test and reference channels.

The experiment was conducted with financial support of the Foundation for Research Support of the State of São Paulo (FAPESP), grant 06/03227-2, and in accordance with the recommendations of the international guidelines for investigations involving human beings with written informed consent from all subjects. The protocol was approved by the Ethics Institutional Committee of the A.C. Camargo Cancer Center (process number 1023/07).

References

Ribeiro AH, Soler JMP, Hirata Jr R (2019). "Variance-Preserving Estimation of Intensity Values Obtained from Omics Experiments." *Submitted for publication*.

`normalizeWithinArraysByOptimalLowess`

Optimal LOWESS Within-Slide Normalization of a MAList

Description

Applies LOWESS for within-slide normalization of the MA values. The parameters are set by a data-driven parameter selection method, proposed by (Ribeiro et al. 2019), which is parsimonious and considers intrinsic characteristics of microarray data, such as heteroskedasticity. Particularly, the best bandwidth is selected according to the HRCp criterion (Liu and Okui 2013).

Usage

```
normalizeWithinArraysByOptimalLowess(MA, eva.values = seq(0.2, 1, by =
  0.05), debug = TRUE, save.objs = TRUE, save.plots = TRUE,
  dir.to.save = "./")
```

Arguments

MA	A MAList object, as in the limma R package, with the non-normalized MA values.
eva.values	A vector with values between 0 and 1 corresponding to the bandwith values to be considered by the LOWESS parameter selection method.
debug	A logical value indication if you want to view logs of the execution.
save.objs	A logical value indicating if you want to save objects with partial results.
save.plots	A logical value indicating if you want to generate the M plots illustrating the bandwidth parameter selection process.
dir.to.save	Path to the folder you want to save the output objects.

Value

A MAList object, as in the limma R package, with the MA values after applying within-slide normalization by LOWESS with optimal parameter settings.

References

Liu Q, Okui R (2013). "Heteroscedasticity-Robust Cp Model Averaging." *The Econometrics Journal*, **16**(3), 463–472.

Ribeiro AH, Soler JMP, Hirata Jr R (2019). "Variance-Preserving Estimation of Intensity Values Obtained from Omics Experiments." *Submitted for publication*.

Examples

```
data(metaplasia)
MA <- estimateMAValues(metaplasia$R.mean, metaplasia$G.mean,
                       metaplasia$R.var, metaplasia$G.var, metaplasia$RG.cov,
                       metaplasia$R.bckg, metaplasia$G.bckg,
                       array.ids=metaplasia$array.ids, gene.ids=metaplasia$gene.ids)
normMA <- normalizeWithinArraysByOptimalLowess(MA)
```

Index