Package 'LVSmiRNA'

September 24, 2012

Type Package

Title LVS normalization for Agilent miRNA data

Version 1.6.0

Date 2010-12-28

Depends R (>= 2.10), Biobase,quantreg,splines,MASS,limma,affy,methods,SparseM, vsn

Imports BiocGenerics, stats4

Enhances multicore,snow, Rmpi

Author Stefano Calza, Suo Chen, Yudi Pawitam

Maintainer Stefano Calza <stefano.calza@biostatistics.it>

biocViews Microarray,AgilentChip,OneChannel,Preprocessing

Description Normalization of Agilent miRNA arrays.

License GPL-2

LazyLoad yes

R topics documented:

boxplot-methods *Methods for Function boxplot*

Description

Methods for function boxplot for objects of class EList and RGList

Methods

signature(x = "EList") boxplot for EList object

signature(x = "RGList") boxplot for RGList object

estVC *Robust Linear Model to Estimate Residual Variance and Array Effect*

Description

Given intensities from microRNA data, fits a robust linear model at probe level and return the residual standard deviations and the array effects.

Usage

```
estVC(object,method=c("joint","rlm"),cov.formula=c("weighted","asymptotic"),clName,verbose=FALSE
## S3 method for class 'RGList'
estVC(object,method=c("joint","rlm"),cov.formula=c("weighted","asymptotic"),clName,verbose=FALSE
## S3 method for class 'EList'
estVC(object,method=c("joint","rlm"),cov.formula=c("weighted","asymptotic"),clName,verbose=FALSE
```

Arguments

object	an object of class EList or RGList.
method	character string specifying the estimating algorithm to be used. Choices are "joint" and "rlm".
cov.formula	character string specifying the covariance formula to be used. Choices are "weighted" and "asymptotic".
clName	Cluster object produced by makeCluster function. Used only if snow is loaded.
verbose	Print some debug messages.

Details

estVC is the first step in LVS normalization. It fits a robust linear model at the probe-level data in order to estimate the variability of probe intensities due to array-to-array variability. Depending on whether probes show considerable differences in within-probe variance, user can choose the more complex joint model to accommodate the potential heteroscedasticity or standard robust linear model if within-probe variance can be ignored.

The array effects are then captured by the chi-square statistic. The covariance matrix can be estimated based either on the sandwich form of weighted covariance matrix or an asymptotic form.

Value

An object of class RA containing three components as follows:

ArrayEffects	a matrix containing the array effect with samples as columns and miRNAs as rows.
ArrayChi2	vector giving chi-square statisitcs of the miRNAs as a measure of array-to-array variability.
logStdDev	vector giving standard deviations of the genes on log scale.

Author(s)

Stefano Calza <stefano.calza@biostatistics.it>, Suo Chen and Yudi Pawitan.

References

Calza et al., 'Normalization of oligonucleotide arrays based on the least variant set of genes', (2008, BMCBioinformatics); Pawitan, Y. 'In All Likelihood: Statistical Modeling and Inference Using Likelihood', (2001, Oxford University Press); Huber, P. J., 'Robust estimation of a location parameter', (1964, Annuas of Mathematical Statistics).

See Also

read.mir, lvs

Examples

```
## Not run:

# Starting from an EList object called MIR
data("MIR-spike-in")
AA <- estVC(MIR,method="joint")

# Parellel execution using multicore

library(multicore)

# use this to set the desided number of
#cores. Otherwise multicore would use all the available
options(cores=8)

AA <- estVC(MIR,method="joint")

detach('package:multicore')

# Parellel execution using snow


library(snow)

cl <- makeCluster(8,type="SOCK")

# Or also...see ?makeCluster
# cl <- makeCluster(8,type="MPI")

AA <- estVC(MIR,method="joint",clName=cl)
```

```
## End(Not run)
```

`exprs-methods` *Methods for Function exprs*

Description

Methods for function `exprs` for objects of class `EList` and `RGList`

Methods

`signature(x = "EList")` exprs for `EList` object

`signature(x = "RGList")` exprs for `RGList` object

`exprs<--methods` *~~ Methods for Function exprs<- in Package 'Biobase' ~~*

Description

~~ Methods for function `exprs<-` in Package 'Biobase' ~~

Methods

```
signature(object = "AffyBatch", value = "ANY")
signature(object = "EList", value = "ANY")
signature(object = "ExpressionSet", value = "matrix")
signature(object = "RGList", value = "ANY")
signature(object = "SnpSet", value = "matrix")
```

`featureNames-methods` *Methods for Function featureNames*

Description

Methods for function `featureNames` for objects of class `EList` and `RGList`

Methods

`signature(x = "EList")` featureNames for `EList` object

`signature(x = "RGList")` featureNames for `RGList` object

lvs *Least Variant Set selection and Normalization Function(s)*

Description

Selects the Least Variant Set of mircoRNAs, according to the chosen proportion of miRNAs expected not to vary between arrays. Then performs normalization.

Usage

```
lvs(RG,RA,ref,proportion=0.7,df=3,method=c("joint","rlm"),
                cov.formula=c("weighted","asymptotic"),
                spar=NULL,normalize.method=c("vsn","smooth.spline","mixed"),
                summarize.args=NULL,stratify=TRUE,n.strata=3,
                level=c("mir","probe"),Atransf=c("sqrt","log"),keep.iset=FALSE,clName,
                verbose=FALSE,...)
## S3 method for class 'RGList'
lvs(RG,RA,ref,proportion=0.7,df=3,method=c("joint","rlm"),
                cov.formula=c("weighted","asymptotic"),
                spar=NULL,normalize.method=c("vsn","smooth.spline","mixed"),
                summarize.args=NULL,stratify=TRUE,n.strata=3,
                level=c("mir","probe"),Atransf=c("sqrt","log"),
                keep.iset=FALSE,clName,verbose=FALSE,...)
## S3 method for class 'EList'
lvs(RG,RA,ref,proportion=0.7,df=3,method=c("joint","rlm"),
                cov.formula=c("weighted","asymptotic"),
                spar=NULL,normalize.method=c("vsn","smooth.spline","mixed"),
                summarize.args=NULL,stratify=TRUE,n.strata=3,
                level=c("mir","probe"),Atransf=c("sqrt","log"),keep.iset=FALSE,clName,
                verbose=FALSE,...)
```

Arguments

RG	an object of class EList or RGList
RA	a list contaning components residual standard deviations, chi-square statistics and array effects. It can be computed by estVC. If not provided it will computed (slower),
proportion	the proportion below which miRNAs are expected not to vary between arrays. Default is set to 0.7.
ref	reference array to be used for normalization. Default is set to mean of array effects across samples.
df	the desired equivalent number of degrees of freedom(trace of the smooth matrix) in smoothing spline.
method	character string specifying the estimating algorithm to be used. Choices are "joint" and "rlm".
cov.formula	character string specifying the covariance formula to be used. Choices are "weighted" and "asymptotic".
spar	smoothing parameter, typicallly in (0,1].

normalize.method	character string specifying the normalization method to be used. Choices are "smooth.spline" and "vsn".
summarize.args	a named list containnig components from argument of summarize.
stratify	logical, if TRUE selection of least variant set will be stratified by expression level.
n.strata	integer giving the number of strata.
level	character string specifying the normalization performed at miRNA level or probe-level.
Atransf	Which transformation to use for Array Effect
keep.iset	return the LVS ids
clName	Cluster object. See estVC.
verbose	Verbose computation
...	...

Details

lvs works by first identifying least variant set (LVS) with the smallest array-to-array variation. The total information extracted from probe-level intensity data of all samples is modeled as a function of array and probe effect in order to select the reference set for normalization. If the residual variances and array effects are available, lvs runs faster because the step of robust linear modeling has already been done.

Once the LVS miRNAs are identified, the normalization is performed using VSN or smooth.spline.

Value

An object of the same class as RG.

G	matrix containing the normalized intensities for each array with miRNAs as rows and arrays as columns.
Gb	matrix containing the background intensities for each array with probes as rows and arrays as columns.
targets	data frame with column FileName giving the names of the files read, with column Sample giving the names of the samplse.
genes	data frame containing annotation information about the probes, for examples miRNA names and IDs and positions on the array.
source	character string giving the image analysis program name.
preprocessing	list with components Background, Normalization, is.log, Summarization indicate which pre-processing step has been done.

Author(s)

Stefano Calza <stefano.calza@biostatistics.it>, Suo Chen and Yudi Pawitan.

References

Calza et al., 'Normalization of oligonucleotide arrays based on the least variant set of genes' (2008, BMCBioinformatics).

See Also

estVC, summarize

Examples

```
## Not run:

# Starting from an Elist object called MIR
data("MIR-spike-in")
AA <- estVC(MIR,method="joint")
bb <- lvs(MIR,RA=AA,level="probe")

##It can also run with object RA missing, but taking longer time
cc <- lvs(MIR)

## End(Not run)
```

MIR-spike-in *Data example*

Description

Data from a micro-RNA spike-in experiment, extracted from scanned images using Agilent Feature Extraction Software.

Usage

```
data("MIR-spike-in")
data("MIR_RA")
```

Details

This dataset is derived from a library of synthetic RNA sequences, cor- responding to human mature miRNAs as well as in-house miRNAs with particularly similar sequences hybridized on an Agilent Human miRNA Mi- croarray 2.0. Data consist of a total of 799 miRNA species (excluding control features) for 4 samples organized in two groups A and B.

Data, colected with the Agilent Feature Extraction Software, are stored in a RGList object with the following components:

- MIR\$G: 'gMeanSignal' - MIR\$Gb: 'gProcessedSignal' - MIR\$gBGMedianSignal: 'gBGMedianSignal' - MIR\$targets 'targets' - MIR\$Row 'Row' - MIR\$Col 'Column' - MIR\$ProbeUID 'Probe ID' - MIR\$genes\$ControlType 'FLAG to specify the sort of feature' - MIR\$genes\$ProbeName 'Probe Name' - MIR\$genes\$GeneName 'microRNA Name' - MIR\$genes\$SystematicName 'microRNA Name' - MIR\$genes\$Description 'Description (not used)'

`MIR.RA` holds an object of class `RA` obtained from using `estVC` on the example data.

Author(s)

Stefano Calza

References

Willenbrock H, Salomon J, Barken KIMB, Nielsen FC, Litman T. 2009. Quantitative miRNA expression analysis: Comparing microarrays with next-generation sequencing. RNA 15: 2028-2034.

Data available from Genome Expression Omnibus (GEO) database under the series accession number GSE14511

See Also

`read.mir`, `estVC`

`plot-method` *Plot of Residual Variance and Array Effect*

Description

Plots results from `estVC`

Usage

```
## S4 method for signature 'RA,ANY'
plot(x,Atransf=c("both","sqrt","log"), abline=c("none","rq"),df=3,proportion=.7,
                    col="black",col.rq="red")
```

Arguments

`x`	An object of class `RA` resulting from `estVC`.
`Atransf`	Transformation to apply at Array Effect
`abline`	Add a line to the plot representing a quantile fit
`df`	Degrees of freedom of the quantile regression
`proportion`	Quantile to fit
`col`	Color for plotting points
`col.rq`	Color for plotting quantile line

Author(s)

Stefano Calza <stefano.calza@biostatistics.it>, Suo Chen and Yudi Pawitan.

References

Calza et al., 'Normalization of oligonucleotide arrays based on the least variant set of genes', (2008, BMCBioinformatics); Pawitan, Y. 'In All Likelihood: Statistical Modeling and Inference Using Likelihood', (2001, Oxford University Press); Huber, P. J., 'Robust estimation of a location parameter', (1964, Annuas of Mathematical Statistics).

See Also

`estVC`,`rq`

Examples

```
## Not run:

# Starting from an EList object called MIR
data("MIR-spike-in")
AA <- estVC(MIR,method="joint")
plot(AA)

## End(Not run)
```

---

preproc-methods *Methods for Function preproc*

Description

Methods for function preproc for objects of class EList and RGList

Methods

signature(x = "EList") preproc for EList object

signature(x = "RGList") preproc for RGList object

---

preproc<--methods *Methods for Function preproc*

Description

Methods for function preproc<- for objects of class EList and RGList

Methods

signature(x = "EList") preproc for EList object

signature(x = "RGList") preproc for RGList object

---

probeNames-methods *Methods for Function probeNames*

Description

Methods for function probeNames for objects of class EList and RGList

Methods

signature(x = "EList") probeNames for EList object

signature(x = "RGList") probeNames for RGList object

read.mir *Read in miRNA Data from Agilent Feature Extractiion Output Files*

Description

Reads intensitity data from a set of one-color microarray image analysis output files.

Usage

```
read.mir(files = NULL, path = NULL, ext = NULL, annotation=NULL,names = NULL, columns = list(E="
         other.columns = NULL, read.bg = TRUE,wt.fun = NULL, verbose = TRUE, sep = "\t", quote =
         remove.ctrl=TRUE, ...)
```

Arguments

files	data frame with column FileName giving the names of the files read, with column Sample giving the names of the samples.
path	character string giving the directory containing the files. The default is the current working directory.
ext	character string giving optional extension to be added to each file name
annotation	character vector of names of columns containing annotation information about the probes
names	character vector of names to be associated with each array as column name. Defaults to removeExt(files).
columns	character vector of names of the required columns.
other.columns	character vector of names of additional required columns to be read.
read.bg	logical, TRUE to indicate background corrected values are read.
wt.fun	function to calculate spot quality weights
verbose	logical, TRUE to report each time a file is read
sep	the field separator character. Values on each line of the file are separated by this character.
quote	character string of characters to be treated as quote marks
remove.ctrl	logical, if TRUE control probes will not be read
...	any other arguments are passed to read.table

Details

This is the main data input function for the LVSmiRNA package for one-color microRNA data. It extracts the green channel intensities from a series of files, produced by Agilent Feature Extractiion software, and assembles them into the components of one list. Data from some other image analysis programs can be read if the appropriate column names containing the intensities are specified using the columns argument. (This will work if the column names are unique and if there are no incomplete rows in the file after the last line of data. Header lines are ok, if appropriately skipped.)

The argument files should be a matrix with two columns at least. One column should contain the names of the samples and the other column should contain names of files containing intensity data.

The argument other.columns allows arbitrary columns of the image analysis output files to be reserved in the data object. These become matrices in the 'other' component.

Value

An Elist object.

G	matrix containing the intensities for each array with probes as rows and arrays as columns.
Gb	matrix containing the background intensities for each array with probes as rows and arrays as columns.
targets	data frame with column FileName giving the names of the files read, with column Sample giving the names of the samplse.
genes	data frame containing annotation information about the probes, for examples miRNA names and IDs and positions on the array.
source	character string giving the image analysis program name.
preprocessing	list with components Background, Normalization, is.log, Summarization indicate which pre-processing step has beendone.

Note

All image analysis files being read are assumed to contain data for the same genelist in the same order. No checking is done to confirm that this is true. Probe annotation information is read from one of the files only.

Author(s)

Stefano Calza <stefano.calza@biostatistics.it>, Suo Chen and Yudi Pawitan.

See Also

read.mir is based on "read.table" in the base package and modified from "read.maimages" in the limma package.

Examples

```
#  Read all intensity files from current working directory
## Not run:
dir.files <- system.file("extdata", package="LVSmiRNA")
taqman.data <- read.table(file.path(dir.files,"Comparison_Array.txt"),header=TRUE,as.is=TRUE)
MIR <- read.mir(taqman.data)

## End(Not run)
```

RLM *Robust Fitting of Linear Models*

Description

Fit a linear model by robust regression using the Huber estimator.

Usage

```
RLM(formula, maxit=20, k=1.345, data, model=TRUE,na.action,
method=c("joint","rlm"), x=TRUE, y=TRUE,
offset,cov.formula=c("weighted","asymptotic"), start=NULL,...)
```

Arguments

formula	a formula of the form y ~ x1 + x2 + ...
maxit	the limit on the number of IWLS iterations.
k	tuning constant used for Huber proposal 2 scale estimation.
data	data frame from which variables specified in formula are preferentially to be taken.
model	should the model frame be returned in the object?
na.action	A function to specify the action to be taken if NAs are found. The 'factory-fresh' default action in R is na.omit, and can be changed by options.
method	currently, method="rlm" and "joint" are supported.
x	should the model frame be returned in the object?
y	should the model matrix be returned in the object?
offset	numeric of length n. This can be used to specify an a priori known component to be included in the linear predictor during fitting.
cov.formula	are the methods to compute covariance matrix, currently either weighted or asymptotic.
start	vector containing starting values for the parameters in the predictor.
...	...

Details

Fitting is done by iterated re-weighted least squares (IWLS). This customized version of robust linear model deal with wild ouliers using log link in joint modelling heterogeneous variance of covariates.

Value

An object of class "RLM" inheriting from "lm".

Author(s)

Stefano Calza <stefano.calza@biostatistics.it>, Suo Chen and Yudi Pawitan.

References

Pawitan, Y. 'In All Likelihood: Statistical Modeling and Inference Using Likelihood', (2001, Oxford University Press); Huber, P. J. , Robust Statistics, (1981. Wiley).

See Also

RLM is modified from "rlm" in the MASS, "rlmFit"

Examples

```
set.seed(133)
n <- 9
p <- 3
X <- matrix(rnorm(n * p), n,p)
y <- rnorm(n)
```

```
fit <- RLM(y~X-1) #no intercept
```

---

rlmFit *Fitter Functions for Robust Linear Models*

Description

These are the basic computing engines called by RLM used to fit robust linear models. These should not be used directly unless by experienced users.

Usage

```
rlmFit(x, y, maxit=20L, k=1.345, offset=NULL,method=c("joint","rlm"),
cov.formula=c("weighted","asymptotic"),start=NULL, error.limit=0.01)
```

Arguments

x	design matrix of dimension n * p.
y	vector of observations of length n, or a matrix with n rows.
maxit	the limit on the number of IWLS iterations.
k	tuning constant used for Huber proposal 2 scale estimation.
offset	numeric of length n. This can be used to specify an a priori known component to be included in the linear predictor during fitting.
method	currently, only method="rlm.fit" is supported.
cov.formula	are the methods to compute covariance matrix, currently either weighted or asymptotic.
start	vector containing starting values for the paramter estimates.
error.limit	the convergence criteria during iterative estimation.

Value

a list with components

coeffecients	p vector
Std.Error	p vector
t.value	p vector
cov.matrix	matrix of dimension p*p
res.SD	value of residual standard deviation

...

Author(s)

Stefano Calza <stefano.calza@biostatistics.it>, Suo Chen and Yudi Pawitan.

References

Yudi Pawitan: In All Likelihood: Statistical modeling and inference using likelihood. Oxford University Press. 2001.

See Also

RLM which you should use for robust linear regression usually.

Examples

```
set.seed(133)
n <- 9
p <- 3
X <- matrix(rnorm(n * p), n,p) #no intercept
y <- rnorm(n)

RLM.fit <- rlmFit (x=X, y=y)
```

---

sampleNames-methods *Methods for Function sampleNames*

Description

Methods for function sampleNames for objects of class EList and RGList

Methods

signature(x = "EList") sampleNames for EList object

signature(x = "RGList") sampleNames for RGList object

---

summarize *LVSmiRNA Summarization Function(s) for microRNA Microarray*

Description

Summarize microRNA microarray data objects.

Usage

```
summarize(object, ...)
## S3 method for class 'EList'
summarize(object,RA,remove.ctrl=FALSE,is.log=!is.null(object$preprocessing$Normalization),
method=c("rlm","medianpolish","mean"),verbose=FALSE,make.exprs=FALSE,...)
## S3 method for class 'RGList'
summarize(object,RA,remove.ctrl=FALSE,is.log=!is.null(object$preprocessing$Normalization),
method=c("rlm","medianpolish","mean"),verbose=FALSE,make.exprs=FALSE,...)
```

Arguments

object	an object for which a summary is desired.
RA	an object from estVC.
remove.ctrl	logical, indicating whether to remove control probes.
is.log	Are data already logged?
method	currently, method "medianpolish","mean" and "rlm" are supported.
verbose	More output
make.exprs	Should the output be and exprSet object?
...	...

Details

For multi-probe, multi-replicate microarray, intensities need to be summarized into a single expression value for each miRNA. The data objects are summarized as if they were lists.

Value

An Elist object containing components as follows:

G	matrix containing the summarized intensities for each array with miRNAs as rows and arrays as columns.
Gb	matrix containing the background intensities for each array with probes as rows and arrays as columns.
targets	data frame with column FileName giving the names of the files read, with column Sample giving the names of the samplse.
genes	data frame containing annotation information about the probes, for examples gene names and IDs and positions on the array.
source	character string giving the image analysis program name.
preprocessing	list with components Background, Normalization, is.log, Summarization indicate which pre-processing step has been done.

Author(s)

Stefano Calza <stefano.calza@biostatistics.it>, Suo Chen and Yudi Pawitan.

References

Irizarry et al., 'Exploration, normalization, and summaries of high density oligonucleotide array probe level data', (2003a, Biostatistics); Huber, P. J., 'Robust estimation of a location parameter', (1964, Annuas of Mathematical Statistics)

See Also

lvs, estVC

Examples

```
## Not run:

data("MIR-spike-in")
AA <- estVC(MIR,method="joint")
dd <- summarize(MIR,RA=AA,method="rlm")

##summarization methods other than rlm, object RA is not required
dd1 <- summarize(MIR,method="medianpolish")
dd2 <- summarize(MIR,method="mean")

## End(Not run)
```

Index