Package 'ENmix'

April 15, 2019

Version 1.18.2

Date 2019-03-06

Title Data preprocessing and quality control for Illumina HumanMethylation450 and MethylationEPIC BeadChip

Type Package

Description The ENmix package provides a set of quality control and data pre-processing tools for Illumina HumanMethylation450 and MethylationEPIC Beadchips. It includes ENmix background correction, RELIC dye bias correction, RCP probe-type bias adjustment, along with a number of additional tools. These functions can be used to remove unwanted experimental noise and thus to improve accuracy and reproducibility of methylation measures. ENmix functions are flexible and transparent. Users have option to choose a single pipeline command to finish all data pre-processing steps (including background correction, dye-bias adjustment, inter-array normalization and probe-type bias correction) or to use individual functions sequentially to perform data pre-processing in a more customized manner. In addition the ENmix package has selectable complementary functions for efficient data visualization (such as data distribution plots); quality control (identifing and filtering low quality data points, samples, probes, and outliers, along with imputation of missing values); identification of probes with multimodal distributions due to SNPs or other factors; exploration of data variance structure using principal component regression analysis plot; preparation of experimental factors related surrogate control variables to be adjusted in downstream statistical analysis; and an efficient algorithm oxBS-MLE to estimate 5-methylcytosine and 5-hydroxymethylcytosine level.

Depends parallel,doParallel,foreach, SummarizedExperiment (>= 1.1.6),minfi (>= 1.22.0)

Imports MASS,preprocessCore,wateRmelon,sva,geneplotter,impute,grDevices,graphics,stats

Suggests minfiData (>= 0.4.1), RPMM, RUnit, BiocGenerics

biocViews ImmunoOncology, DNAMethylation, Preprocessing, QualityControl, TwoChannel, Microarray, OneChannel, MethylationArray, BatchEffect, Normalization, DataImport, Regression, PrincipalComponent,Epigenetics, MultiChannel, DifferentialMethylation

License Artistic-2.0

NeedsCompilation no

Maintainer Zongli Xu <xuz@niehs.nih.gov>

git_url https://git.bioconductor.org/packages/ENmix

git_branch RELEASE_3_8

git_last_commit e2cea15

git_last_commit_date 2019-03-06

Date/Publication 2019-04-15

Author Zongli Xu [cre, aut],
Liang Niu [aut],
Leping Li [ctb],
Jack Taylor [ctb]

R topics documented:

`B2M` *Convert Beta value to M value.*

Description

Convert Beta value to M value.

Usage

```
B2M(x)
```

Arguments

x	An numeric matrix with values between 0 and 1

Value

A matrix of M values

Author(s)

Zongli Xu

---

bmiq.mc *A multi-processor wrapper of BMIQ method*

Description

A multi-processor wrapper of BMIQ method. BMIQ is an intra-sample normalization procedure to correct the bias of Infinium 2 probe methylation beta values.

Usage

```
bmiq.mc(mdat, nCores = 1,...)
```

Arguments

mdat	An object of class MethylSet.
nCores	Number of cores used for computation.
...	See BMIQ in R package wateRmelon for more options.

Value

A data matrix of Methylation beta value.

Author(s)

Zongli Xu

References

Teschendorff AE et. al. *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data*. Bioinformatics. 2013

See Also

See BMIQ in R package watermelon for model details

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdatq1=norm.quantile(mdat,method="quantile1")
beta=bmiq.mc(mdatq1,nCores=10)
}}
```

ComBat.mc *A multi-processor wrapper for ComBat method.*

Description

A multi-processor wrapper for ComBat method. ComBat is a method to adjust batch effect where the batch covariate is known.

Usage

```
ComBat.mc(dat,batch,nCores = 1,...)
```

Arguments

dat	A data matrix with column for samples and row for probe.
batch	Batch covariate (multiple batches allowed)
nCores	Number of cores will be used for computation
...	See ComBat in sva package for extra options

Value

A data matrix with the same dimension as input data, adjusted for batch effects. Warning: Values for multimodal distributed CpGs could be over-adjusted.

Author(s)

Zongli Xu

References

Johnson, WE, Rabinovic, A, and Li, C (2007). *Adjusting batch effects in microarray expression data using Empirical Bayes methods. Biostatistics 8(1):118-127.*

See Also

See ComBat in sva package for details.

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdat=norm.quantile(mdat,method="quantile1")
beta=bmiq.mc(mdat,nCores=10)
batch=factor(pData(mdat)$Slide)
betaC=ComBat.mc(beta,batch,nCores=6,mod=NULL)
}}
```

ctrlsva *Non-negative control surrogate variables*

Description

Surrogate variables derived from intensity data for non-negative internal control probes. These variables can be modeled in association analysis to adjust for experimental batch effects.

Usage

```
ctrlsva(rgSet,percvar=0.9,npc=1,flag=1)
```

Arguments

rgSet	An object of class RGChannelSet.
percvar	Minimum percentage of data variations can be explained by surrogate variables, range from 0 to 1,default is 0.9
npc	Number of surrogate variables, default is 1
flag	1: select number of surrogate variables based on argument percvar; 2: select number of surrogate variables based on argument npc

Value

ctrlsva: a matrix of surrogate variables (columns) with row corresponding to samples

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015.

Examples

```
if(FALSE){
if (require(minfiData)) {
sheet <- read.metharray.sheet(file.path(find.package("minfiData"),"extdata"), pattern = "csv$")
rgSet <- read.metharray.exp(targets = sheet,extended = TRUE)
sva<-ctrlsva(rgSet)
}}
```

freqpoly *Frequency polygon plot*

Description

Similar to histogram, frequency polygon plot can be used to display data distribution.

Usage

```
freqpoly(mat, nbreaks=15, col="black", xlab="", ylab="Frequency",
         type="l",append=FALSE,...)
```

Arguments

mat	A numeric vector
nbreaks	Number of bins for frequency counting
col	color code
xlab	x-axis lable
ylab	y-axis lable
type	character indicating the type of plotting; actually any of the 'type's as in 'plot.default'.
append	TRUE or FALSE, whether to create a new figure or append to the current figure.
...	Further arguments that get passed to the function "plot"

Value

Frequency polygon plot.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015.

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat <- preprocessRaw(RGsetEx)
beta=getBeta(mdat, "Illumina")
freqpoly(beta[,1])
}}
```

getBeta *Extract Beta value.*

Description

Extract Methylation Beta value, Beta = Meth / (Meth + Unmeth + offset)

Usage

```
getBeta(mdat,type="Illumina",offset=100)
```

Arguments

mdat	An object of class MethylSet.
type	type="Illumina" sets offset=100 as per Genome Studio.
offset	offset in calculating beta ratio

Value

beta: a matrix of beta values

Author(s)

Zongli Xu

M2B *Convert M value to Beta value.*

Description

Convert M value to Beta value.

Usage

```
M2B(x)
```

Arguments

x	An numeric matrix

Value

A matrix of methylation Beta values.

Author(s)

Zongli Xu

`mpreprocess` *A pipeline to perform background correction, dye bias correction, inter-array normalization and probe type bias correction for HumanMethylation 450 and MethylationEPIC BeadChip data.*

Description

Fuction mpreprocess is a pipeline to easy preprocess steps for Illumina DNA methylation BeadChip. It comprehensively removes background noise and correct bias due to array design, including background correction, dye bias correction, inter-array normalization, probe type bias correction. It will also identify and exclude low quality samples and probes, remove outlier values, and perform imputation.

Usage

```
mpreprocess(rgSet,nCores=1,bgParaEst="oob",dyeCorr="RELIC",
            qc=FALSE,qnorm=TRUE,qmethod="quantile1",
            foutlier=TRUE,rmcr=FALSE,impute=FALSE)
```

Arguments

`rgSet`	An object of class `RGChannelSetExtended`, `RGChannelSet` or `MethylSet`.
`nCores`	Number of cores will be used for computation
`bgParaEst`	Method to estimate background normal distribution parameters. This must be one of the strings: "oob","est", or "neg".
`dyeCorr`	Dye bias correction, "mean": correction based on averaged red/green ratio; or "RELIC": correction with RELIC method; or "none": no dye bias correction. The default is RELIC
`qc`	If TRUE, QC will be performed. Low quality samples and CpGs will be excluded before background correction.
`qnorm`	If TRUE, inter-array quantile normalization will be performed.
`qmethod`	Quantile normalization method. This should be one of the following strings: "quantile1", "quantile2", or "quantile3". See details in function norm.quantile.
`foutlier`	If TRUE, outlier and low quality values will be filtered out.
`rmcr`	TRUE: excluded rows and columns with more than 5% of missing values. FALSE is in default
`impute`	Whether to impute missing values. If TRUE, k-nearest neighbor's methods will be used for imputation. FALSE is in default.

Details

Fuction mpreprocess is a pipeline that perform methylaiton data preprocessing and quality controls using functions: preprocessENmix, norm.quantile, rcp, QCinfo and rm.outlier. More customized preprocessing steps can be achieved using the individual functions.

Value

a methylation beta value matrix.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. Nucleic Acids Research 2015.

Zongli Xu, Sabine A. S. Langie, Patrick De Boever, Jack A. Taylor1 and Liang Niu, RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip, in review 2016

Liang Niu, Zongli Xu and Jack A. Taylor: RCP: a novel probe design bias correction method for Illumina Methylation BeadChip, Bioinformatics 2016

See Also

Package `minfi` for classes `RGChannelSet` and `MethylSet`

Examples

```
if(FALSE){
if (require(minfiData)) {
beta=mpreprocess(RGsetEx,nCores=6)

sheet <- read.metharray.sheet(file.path(find.package("minfiData"),"extdata"), pattern = "csv$")
rgSet <- read.metharray.exp(targets = sheet,extended = TRUE)
beta=mpreprocess(rgSet,nCores=6,qc=TRUE,foutlier=TRUE,rmcr=TRUE,impute=TRUE)
}}
```

---

multifreqpoly *Frequency polygon plot to display data distribution.*

Description

Produce Frequency polygon plot for each column of a numeric data matrix.

Usage

```
multifreqpoly(mat, nbreaks=100, col=1:ncol(mat), xlab="",
              ylab="Frequency",legend = list(x = "top", fill=col,
              legend = if(is.null(colnames(mat))) paste(1:ncol(mat))
              else colnames(mat)),...)
```

Arguments

`mat`	A numeric matrix
`nbreaks`	The number of bins for frequency counting
`col`	Line plot color code, the length should be equal to the number of columns in mat
`xlab`	x-axis lable
`ylab`	y-axis lable
`legend`	A list of arguments that get passed to the function "legend"
`...`	Further arguments that get passed to the function "plot"

Value

Frequency polygon plot.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015.

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat <- preprocessRaw(RGsetEx)
beta=getBeta(mdat, "Illumina")
multifreqpoly(beta,col=rep("black",ncol(beta)))
}}
```

`nmode.mc` *Estimating number of mode in methylaion data for each probe.*

Description

Due to SNPs in CpG probe region or other unknow factors, methylation beta values for some CpGs have multimodal distribution. This function is to identify this type of probes with obovious multimoal distribution.

Usage

```
nmode.mc(x, minN = 3, modedist=0.2, nCores = 1)
```

Arguments

`x`	A methylation beta value matrix with row for probes and column for samples.
`minN`	Minimum number of data points at each cluster
`modedist`	Minimum mode distance
`nCores`	Number of cores used for computation

Details

This function used an empirical approach to estimate number of mode in methylation beta value for each CpG probe. By default, the function requires the distance between modes have to be greater than 0.2 in methylation beta value, and each mode clusters should has at least 3 data points or 5% of data points whichever is greater.

Value

A vector of integers

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat <- preprocessRaw(RGsetEx)
beta=getBeta(mdat, "Illumina")
nmode=nmode.mc(beta, minN = 3,modedist=0.2, nCores = 5)
}}
```

---

norm.quantile *Quantile normalization.*

Description

Quantile normalization of methylation intensity data across samples for Illumina Infinium HumanMethylation 450 and MethylationEPIC BeadChip.

Usage

```
norm.quantile(mdat, method = "quantile1")
```

Arguments

mdat	An object of class MethylSet.
method	Quantile normalization method. This should be one of the following strings: "quantile1", "quantile2", or "quantile3".

Details

By default, method = "quantile1" will separately quantile normalize Methylated or Unmethylated intensities for Infinium I or II probes. The "quantile2" will quantile normalize combined Methylated or Unmethylated intensities for Infinium I or II probes. The "quantile3" will quantile normalize combined Methylated or Unmethylated intensities for Infinium I and II probes together.

Value

An object of class MethylSet.

Author(s)

Zongli Xu

References

Pidsley, R., CC, Y.W., Volta, M., Lunnon, K., Mill, J. and Schalkwyk, L.C. (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. BMC genomics, 14, 293.

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdatq1=norm.quantile(mdat,method="quantile1")
}}
```

`normalize.quantile.450k`

Quantile normalization.

Description

Quantile normalization of methylation intensity data across samples for Illumina Infinium HumanMethylation 450 BeadChip.

Usage

```
normalize.quantile.450k(mdat, method = "quantile1")
```

Arguments

`mdat`	An object of class `MethylSet`.
`method`	Quantile normalization method. This should be one of the following strings: "quantile1", "quantile2", or "quantile3".

Details

By default, method = "quantile1" will separately quantile normalize Methylated or Unmethylated intensities for Infinium I or II probes. The "quantile2" will quantile normalize combined Methylated or Unmethylated intensities for Infinium I or II probes. The "quantile3" will quantile normalize combined Methylated or Unmethylated intensities for Infinium I and II probes together.

Value

An object of class `MethylSet`.

Author(s)

Zongli Xu

References

Pidsley, R., CC, Y.W., Volta, M., Lunnon, K., Mill, J. and Schalkwyk, L.C. (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. BMC genomics, 14, 293.

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdatq1=normalize.quantile.450k(mdat,method="quantile1")
}}
```

---

oxBS.MLE *oxBS-MLE.*

Description

Find the Maximum Likelihood Estimate (MLE) of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) using sequencing/array data from paired bisulfite and oxidative bisulfite treated DNA experiments.

Usage

```
oxBS.MLE(beta.BS,beta.oxBS,N.BS,N.oxBS)
```

Arguments

beta.BS	A matrix of measurements (proportions of methylated signals in total signals) obtained from bisulfite (BS) experiments
beta.oxBS	A matrix of measurements obtained from oxidative bisulfite (oxBS) experiments
N.BS	A matrix of total signals from BS experiments
N.oxBS	A matrix of total signals from oxBS experiments

Details

For all the inputs (beta.BS, beta.oxBS, N.BS and N.oxBS), the rows should be corresponding to CpG loci and the columns should be corresponding to samples. The order of rows/columns in all four matrices should be consistent (otherwise oxBS.MLE will stop with error messages). Using a binomial model at each CpG locus in each sample, oxBS.MLE outputs a list with two matrices: a matrix of MLEs of 5mC levels and a matrix of MLEs of 5hmC levels. The rows and columns of both matrices are consistent with the rows and columns of the input matrices. For any CpG locus in any sample, if any of the four corresponding values (beta.BS, beta.oxBS, N.BS and N.oxBS) is NA, or N.BS is zero, or N.oxBS is zero, the MLE of both 5mC and 5hmC levels will be set as NA.

Value

A list with two elements:

5mC: a matrix of estimated 5mC levels.

5hmC: a matrix for estimated 5hmC levels.

Author(s)

Liang Niu and Zongli Xu

References

Zongli Xu, Jack A. Taylor, Yuet-Kin Leung, Shuk-Mei Ho and Liang Niu, *oxBS-MLE: An efficient method to estimate 5-methylcytosine and 5-hydroxymethylcytosine in paired bisulfite and oxidative bisulfite treated DNA*, under review.

Examples

```
# load example data
load(system.file("oxBS.MLE.RData",package="ENmix"))
# run oxBS.MLE
temp<-oxBS.MLE(beta.BS,beta.oxBS,N.BS,N.oxBS)
```

---

pcrplot *Principal component regression plot*

Description

First, principal component analysis will be performed in the standadized input data matrix (standadized for each row/CpG), and then the specified number of top principal components (that explain most data variation) will be used to perform linear regression with each specified variables. Regression P values will be plotted for exploration of methylation data variance structure or identification of possible confounding variables for association analysis.

Usage

```
pcrplot(beta, cov,npc=50)
```

Arguments

beta	A methylation beta value matrix with row for probes and column for samples.
cov	A data frame of covariates. Categorical variables should be converted to factors.
npc	The number of top principal components to plot

Value

A jpeg figure "svdscreeplot.jpg" to show the variations explained by each principal component.

A jpeg figure "pcr_diag.jpg" to show association strength between principal components and covariates with cell colors indicating different levels of association P values.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat <- preprocessRaw(RGsetEx)
beta=getBeta(mdat, "Illumina")
group=pData(mdat)$Sample_Group
slide=factor(pData(mdat)$Slide)
cov=data.frame(group,slide)
pcrplot(beta,cov,npc=6)
}}
```

---

plotCtrl *Plot internal controls of 450K or MethylationEPIC BeadChip.*

Description

Intensity data are ploted for all internal control probe types on the Illumina Infinium HumanMethylation450 or MethylationEPIC BeadChip. These figures can be used to check data quality and experimental procedures.

Usage

```
plotCtrl(rgSet,IDorder=NULL)
```

Arguments

rgSet	An object of class RGChannelSet.
IDorder	A list of sample ids in the order user specified. The list can be a subset of the samples in input dataset. If an id list is provided, all plots will be produced in the order of the list. The default is NULL.

Value

A set of jpeg figures.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015.

Examples

```
if(FALSE){
if (require(minfiData)) {
pinfo=pData(RGsetEx)
IDorder=rownames(pinfo)[order(pinfo$Slide,pinfo$Array)]
plotCtrl(RGsetEx,IDorder)
}}
```

`preprocessENmix` *The ENmix background correction for HumanMethylation 450 and MethylationEPIC BeadChip*

Description

ENmix models methylation signal intensities with a flexible exponential-normal mixture distribution, and models background noise with a truncated normal distribution. ENmix will split BeadChip intensity data into 6 parts and separately model methylated and unmethylated intensities, 2 different color channels and 2 different probe designs.

Usage

```
preprocessENmix(rgSet, bgParaEst = "oob", dyeCorr="RELIC", QCinfo=NULL, exQCsample=TRUE,
                    exQCcpg=TRUE, exSample=NULL, exCpG=NULL, nCores = 2)
```

Arguments

`rgSet`	An object of class `RGChannelSetExtended`, `RGChannelSet` or `MethylSet`.
`bgParaEst`	Optional method to estimate background normal distribution parameters. This must be one of the strings: "oob","est", or "neg".
`dyeCorr`	Dye bias correction, "mean": correction based on averaged red/green ratio; or "RELIC": correction with RELIC method; or "none": no dye bias correction. The default is RELIC
`QCinfo`	If QCinfo object from function QCinfo() is provided, low quality samples (if exQCsample=TRUE) and CpGs (if exQCcpg=TRUE) will be excluded before background correction.
`exQCsample`	If TRUE, low quality samples listed in QCinfo will be excluded.
`exQCcpg`	If TRUE, low quality CpGs listed in QCinfo will be excluded.
`exSample`	User specified sample list to be excluded before background correction
`exCpG`	User specified probe list to be excluded before background correction
`nCores`	Number of cores will be used for computation

Details

By default, ENmix will use out-of-band Infinium I intensities ("oob") to estimate normal distribution parameters to model background noise. Option "est" will use combined methylated and unmethylated intensities to estimate background distribution parameters separately for each color channel and each probe type. Option "neg" will use 600 chip internal controls probes to estimate background distribution parameters. If rgSet if a `MethylSet`, then only option "est" can be selected.

Value

An object of class `MethylSet`

Author(s)

Zongli Xu and Liang Niu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. Nucleic Acids Research 2015.

Zongli Xu, Sabine A. S. Langie, Patrick De Boever, Jack A. Taylor1 and Liang Niu, RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip, in review 2016

See Also

Package `minfi` for classes `RGChannelSet` and `MethylSet`

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,nCores=6)
}}
```

`QCfilter` *Sample or CpG probe filter.*

Description

Filter low quality samples or CpGs, outlier samples or user specified samples or CpGs.

Usage

```
QCfilter(mdat,qcinfo=NULL,detPthre=0.000001,nbthre=3,samplethre=0.05,CpGthre=0.05,
         bisulthre=NULL,outlier=FALSE,outid=NULL, outCpG=NULL,plot=FALSE)
```

Arguments

`mdat`	An object of class `MethylSet` or beta value matrix.
`qcinfo`	An object outputed from function `QCinfo`
`detPthre`	Detection P value threshold to identify low quality data point
`nbthre`	Number of bead threshold to identify low quality data point
`samplethre`	Threshold to identify low quality samples, the percentage of low quality methylation data points across probes for each sample
`CpGthre`	Threshold to identify low quality probes, percentage of low quality methylation data points across samples for each probe
`bisulthre`	Threshold of bisulfite intensity for identification of low quality samples. By default, Mean - 3 x SD of sample bisufite control intensities will be used as the threshold.
`outlier`	If TRUE, outlier samples will be excluded.
`outid`	A list of user specified samples to be excluded.
`outCpG`	A list of user specified CpGs to be excluded.
`plot`	TRUE or FALSE, whether to produce quality checking plots.

Value

An same type object as input object after excluding low quality samples and CpGs

Figure "qc_sample.jpg": scatter plot for Percent of low quality data per sample and Average bisulfite conversion intensity

Figure "qc_CpG.jpg": histogram for Percent of low quality data per CpG.

Figure "freqpolygon_beta_beforeQC.jpg": distribution plot before filtering.

Figure "freqpolygon_beta_afterQC.jpg": distribution plot after filtering.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015.

Examples

```
if(FALSE){
if (require(minfiData)) {
sheet <- read.metharray.sheet(file.path(find.package("minfiData"),"extdata"), pattern = "csv$")
rgSet <- read.metharray.exp(targets = sheet,extended = TRUE)
qcscore<-QCinfo(rgSet)
rgSet=QCfilter(rgSet,qcinfo=qcscore,outlier=TRUE)
}}
```

`QCinfo` *QC information.*

Description

Extract informations for data quanlity controls: detection P values, number of beads and averaged bisulfite conversion intensity. The function can also identify low quality samples and probes, as well as outlier samples based on total intensity or beta value distribution.

Usage

```
QCinfo(rgSet, detPthre=0.000001, nbthre=3, samplethre=0.05, CpGthre=0.05,
       bisulthre=NULL, outlier=TRUE, distplot=TRUE)
```

Arguments

`rgSet`	An object of class `RGChannelSetExtended`.
`detPthre`	Detection P value threshold to identify low quality data point
`nbthre`	Number of bead threshold to identify low quality data point
`samplethre`	Threshold to identify low quality samples, the percentage of low quality methylation data points across probes for each sample
`CpGthre`	Threshold to identify low quality probes, percentage of low quality methylation data points across samples for each probe

`bisulthre` Threshold of bisulfite intensity for identification of low quality samples. By default, Mean - 3 x SD of sample bisufite control intensities will be used as the threshold.

`outlier` If TRUE, outlier samples in total intensity or beta value distribution will be idenfied and classified as bad samples.

`distplot` TRUE or FALSE, whether to produce beta value distribution plots before and after QC.

Value

detP: a matrix of detection P values

nbead: a matrix for number of beads

bisul: a vector of averaged intensities for bisulfite conversion controls

badsample: a list of low quality or outlier samples

badCpG: a list of low quality CpGs

outlier_sample: a list of outlier samples

Figure "qc_sample.jpg": scatter plot for Percent of low quality data per sample and Average bisulfite conversion intensity

Figure "qc_CpG.jpg": histogram for Percent of low quality data per CpG.

Figure "freqpolygon_beta_beforeQC.jpg": distribution plot before filtering.

Figure "freqpolygon_beta_afterQC.jpg": distribution plot after filtering.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015.

Examples

```
if(FALSE){
if (require(minfiData)) {
sheet <- read.metharray.sheet(file.path(find.package("minfiData"),"extdata"), pattern = "csv$")
rgSet <- read.metharray.exp(targets = sheet,extended = TRUE)
qcscore<-QCinfo(rgSet)
}}
```

rcp *Regression on Correlated Probes(RCP)*

Description

Probe design type bias correction using Regression on Correlated Probes (RCP) method

Usage

```
rcp(mdat, dist=25, quantile.grid=seq(0.001,0.999,by=0.001), qcscore = NULL, nbthre=3, detPt
```

Arguments

mdat	An object of class MethylSet.
dist	Maximum distance in base pair between type I and type II probe pairs for regression calibration
quantile.grid	Quantile grid used in linear regression
qcscore	If the data quality infomation (the output from function QCinfo) is provied, low quality data points as defined by detection p value threshold (detPthre=0.000001) or number of bead threshold (nbthre=3) will be set to missing.
detPthre	Detection P value threshold to define low qualitye data points, detPthre=0.000001 in default.
nbthre	Number of beads threshold to define low qualitye data points, nbthre=3 in default.

Details

The function will first identify type I and type II probe pairs within specified distance, and then perform linear regression between the probe types to estimate regression coefficients. With the estimates the function will then calibrates type II data using type I data as references.

Value

A beta value matrix

Author(s)

Liang Niu, Zongli Xu

References

Liang Niu, Zongli Xu and Jack A. Taylor *RCP: a novel probe design bias correction method for Illumina Methylation BeadChip, Bioinformatics 2016*

Examples

```
if(FALSE){
if (require(minfiData)) {
mdat=preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6)
mdatq1=norm.quantile(mdat,method="quantile1")
beta=rcp(mdatq1)
}}
```

relic *RELIC dye bias correction method for Illumina HumanMethylation450 and MethylationEPIC BeadChip*

Description

REgression on Logarithm of Internal Control probes (RELIC) correct for dye bias on whole array by utilizing the intensity values of paired internal control probes that monitor the two color channels.

Usage

```
relic (mdat,at_red,cg_grn)
```

Arguments

mdat	An object of class MethylSet.
at_red	an intensity matrix for Illumina control probes "NORM_A" and "NORM_T"
cg_grn	an intensity matrix for Illumina control probes "NORM_C" and "NORM_G"

Details

The Illumina MethylationEPIC BeadChip contains 85 pairs of internal normalization control probes (name with prefix NORM_A, NORM_T, NORM_G or NORM_C), while its predecessor, Illumina HumanMethyl-ation450 BeadChip contains 93 pairs. RELIC first performs a regression on the logarithms of the intensity values of the normalization control probes to derive a quantitative relationship between red and green channels, and then uses the relation-ship to correct for dye-bias on intensity values for whole array.

Value

An object of class MethylSet

Author(s)

Zongli Xu and Liang Niu

References

Zongli Xu, Sabine A. S. Langie, Patrick De Boever, Jack A. Taylor and Liang Niu, RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip, in review 2016

See Also

Package preprocessENmix

Examples

```
if(FALSE){
if (require(minfiData)) {
    #background correction and dye bias correction
    mdat <- preprocessENmix(RGsetEx,bgParaEst="oob",nCores=6,dyeCorr =="RELIC")
    #dye bias correction only
    ctrls <- getProbeInfo(RGsetEx,type="Control")
    ctrls <- ctrls[ctrls$Address %in% featureNames(RGsetEx),]
    ctrl_r <- getRed(RGsetEx)[ctrls$Address,]
    ctrl_g <- getGreen(RGsetEx)[ctrls$Address,]
    CG.controls <- ctrls$Type %in% c("NORM_C","NORM_G")
    AT.controls <- ctrls$Type %in% c("NORM_A","NORM_T")
    cg_grn <- ctrl_g[CG.controls,]
    rownames(cg_grn) = ctrls$ExtendedType[CG.controls]
    at_red <- ctrl_r[AT.controls,]
    rownames(at_red) = ctrls$ExtendedType[AT.controls]
    mdat <- preprocessRaw(RGsetEx)
    mdat <- relic(mdat,at_red,cg_grn)
}}
```

`rm.outlier` *Filtering out outlier and/or low quality values*

Description

Setting outliers as missing value. Outlier was defined as value smaller than 3 times IQR from the lower quartile or larger than 3 times IQR from the upper quartile. If data quality information were provided, low quality data points will be set to missing first before looking for outliers. If specified, imputation will be performed using k-nearest neighbors method to impute all missing values.

Usage

```
rm.outlier(mat,byrow=TRUE,qcscore=NULL,detPthre=0.000001,nbthre=3,
           rmcr=FALSE,rthre=0.05,cthre=0.05,impute=FALSE,
           imputebyrow=TRUE,...)
```

Arguments

`mat`	An numeric matirx
`byrow`	TRUE: Looking for outliers row by row, or FALSE: column by column.
`qcscore`	If the data quality infomation (the output from function QCinfo) were provied, low quality data points as defined by detection p value threshold (detPthre) or number of bead threshold (nbthre) will be set to missing.
`detPthre`	Detection P value threshold to define low qualitye data points, detPthre=0.000001 in default.
`nbthre`	Number of beads threshold define low qualitye data points, nbthre=3 in default.
`rmcr`	TRUE: excluded rows and columns with too many missing values as defined by rthre and cthre. FALSE is in default
`rthre`	Minimum of percentage of missing values for a row to be excluded
`cthre`	Minimum of percentage of missing values for a column to be excluded

impute	Whether to impute missing values. If TRUE, k-nearest neighbors methods will used for imputation. FALSE is in default. Warning: imputed values for multi-modal distributed CpGs may not be correct.
imputebyrow	TRUE: impute missing values using similar values in row, or FALSE: in column
...	Arguments to be passed to the function impute.knn in R package "impute"

Value

An numeric matrix of same dimention as the input matrix.

Author(s)

Zongli Xu

References

Zongli Xu, Liang Niu, Leping Li and Jack A. Taylor, *ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip*. Nucleic Acids Research 2015.

Examples

```
if(FALSE){
if (require(minfiData)) {
sheet <- read.metharray.sheet(file.path(find.package("minfiData"),"extdata"), pattern = "csv$")
rgSet <- read.metharray.exp(targets = sheet,extended = TRUE)
qcscore<-QCinfo(rgSet)
mdat <- preprocessRaw(rgSet)
beta=getBeta(mdat, "Illumina")
#filter out outliers
b1=rm.outlier(beta)
#filter out low quality and outlier values
b2=rm.outlier(beta,qcscore=qcscore)
#filter out low quality and outlier values, remove rows and columns with too many missing values
b3=rm.outlier(beta,qcscore=qcscore,rmcr=TRUE)
#filter out low quality and outlier values, remove rows and columns with too many missing values, and then do
b3=rm.outlier(beta,qcscore=qcscore,rmcr=TRUE,impute=TRUE)
}}
```

Index