Package 'ENmix'

October 17, 2020

Version 1.25.1

Date 2020-04-27

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; an efficient algorithm oxBS-MLE to estimate 5-methylcytosine and 5-hydroxymethylcytosine level; estimation of celltype proporitons; methlation age calculation and differentially methylated region (DMR) analysis.

Depends parallel,doParallel,foreach,SummarizedExperiment,stats

Imports grDevices,graphics,preprocessCore,matrixStats,methods,utils,irr, geneplotter,impute,minfi,RPMM,illuminaio,dynamicTreeCut,IRanges, Biobase,ExperimentHub,AnnotationHub,genefilter,gplots,quadprog,S4Vectors

Suggests minfiData, RUnit, BiocGenerics

biocViews DNAMethylation, Preprocessing, QualityControl, TwoChannel, Microarray, OneChannel, MethylationArray, BatchEffect, Normalization, DataImport, Regression,

PrincipalComponent,Epigenetics, MultiChannel,
DifferentialMethylation, ImmunoOncology

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_11

git_last_commit dea28bb

git_last_commit_date 2020-04-28

Date/Publication 2020-10-16

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

R topics documented:

`B2M` *Convert methylation beta value to M value.*

Description

Convert methylation beta value to M value.

Usage

```
B2M(x)
```

Arguments

`x`	An numeric matrix with values between 0 and 1

Details

Methylation beta value is calculated as beta=M/(M+U+a). M is methylated intensity, U is unmethylated intensity, and a is a constant offset (by default , a=100). M value is calculated as M=log2((M+a)/(U+a)). M or U is usually greater than 1000, so a is negligible for most probes. if a=0, then M=log2(beta/(1-beta).

Value

A matrix of M values

Author(s)

Zongli Xu

`calcdetP` *To calculate detection P values*

Description

Calculate detection P values based on negtive internal control probes or out of the band (oob) probes

Usage

```
calcdetP(rgSet,detPtype = "negative")
```

Arguments

`rgSet`	An object of class `rgDataSet`
`detPtype`	Calculate detection P values based on negtive internal control ("negative") probes or out of the band ("oob") probes

Value

An numerical matrix of detection P values, with row for CpGs and column for samples

Author(s)

Zongli Xu

References

Wanding Zhou et al. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions, Nucleic Acids Research, 2018

Examples

```
## Not run:
path <- file.path(find.package("minfiData"),"extdata")
rgSet <- readidat(path = path,recursive = TRUE)
detp=calcdetP(rgSet,detPtype = "negative")
detp2=calcdetP(rgSet,detPtype = "oob")

## End(Not run)
```

`combp` *Identification of differentially methylated regions*

Description

To identify differentially methylated regions using a modified comb-p method

Usage

```
combp(data,dist.cutoff=1000,bin.size=310,seed=0.01,
      region_plot=TRUE,mht_plot=TRUE,nCores=10)
```

Arguments

`data`	A data frame from bed format file with colname name "V1","V2", "V3","V4","V5", V1 indicate chromosome (1,2,3,...,X,Y), V2 is chromosome position, V4 is for P value and V5 for name of CpGs
`dist.cutoff`	Maximum distance in base pair to combine adjacent DMRs
`bin.size`	bin size for autocorrelation calculation
`seed`	FDR significance threshold for initial selection of DMR region
`region_plot`	Whether to draw regional plot
`mht_plot`	Whether to draw mahattan plot
`nCores`	Number of computer cores used in calculation

Details

The input should be a data frame with column name V1-V5, indicating chromosome, start position, end position, P values and probe names. The function will use a modified comb-p method to identify differentially methylated regions. DMR results will be stored in a file with name "resu_combp.csv". If plot options were selected, two figure files will be generated: mht.jpg and region_plot.pdf.

Author(s)

Liang Niu, Zongli Xu

References

Pedersen BS1, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinfomatics 2012

Zongli Xu, Changchun Xie, Jack A. Taylor, Liang Niu, ipDMR: Identification of differentially methyl-ated regions with interval p-values, in review

---

ctrlsva *Non-negative internal control surrogate variables*

Description

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

Usage

```
ctrlsva(rgSet,percvar=0.95,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: an numerical matrix with columns indicating surrogate variables and rows 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)
}}
```

---

dupicc *Evaluations of methylation duplicates*

Description

To calcuate ICC (consistency and twoway) and Pearson's correlation coefficient for each CpG probes, and to calculate centered/un-centered Pearson's correlation coefficient and methylation differences between duplicate samples.

Usage

```
dupicc(dat,dupid,mvalue=FALSE,center=FALSE,nCores=2,qcflag=FALSE,qc=NULL,
       detPthre=0.05,nbthre=3,skipicc=FALSE)
```

Arguments

dat	Methylation beta value matrix
dupid	A data frame with two variables, id1 and id2, for corresponding duplicates ids in methylation matrix
mvalue	Whether to convert to M value for calculation of ICC
center	Whether methylation beta values will be centered for calculation of correlation between duplicate samples
nCores	Number of core will be used for calculation of ICC
qcflag	Whether to perform QC before calculation of ICC
qc	QC object from ENmix package
detPthre	Detection P value threshold to identify low quality data point
nbthre	Number of bead threshold to identify low quality data point
skipicc	If TRUE, ICC calculation will be skipped

Value

icc: a data frame containing ICC and P values for each probes

dupcor: a data frame containing Pearson's correlation and averaged absolute difference between duplicates.

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)
mdat=preprocessRaw(rgSet)
beta=getBeta(mdat,"Illumina")
dupidx=data.frame(id1=c("5723646052_R02C02","5723646052_R04C01","5723646052_R05C02"),
id2=c("5723646053_R04C02","5723646053_R05C02","5723646053_R06C02"))
iccresu<-dupicc(dat=beta,dupid=dupidx)
}}
```

`estimateCellProp` *Cell type proportion estimation*

Description

To estimates relative proportion of underlying cell types of a sample based on reference methylation data of pure cell types.

Usage

```
estimateCellProp(userdata,refdata="FlowSorted.Blood.450k",
                 cellTypes=NULL,nonnegative = TRUE,nProbes=50,
                 normalize=TRUE,refplot=FALSE)
```

Arguments

`userdata`	The input can be `rgDataSet`,`methDataSet`\, codeMethylSet,`RGChannelSet` or methylation beta value matrix.
`refdata`	Reference data set will used: "FlowSorted.Blood.450k", "FlowSorted.DLPFC.450k","FlowSorted.Co "FlowSorted.CordBloodCombined.450k", "FlowSorted.CordBloodNorway.450k" or "FlowSorted.Blood.EPIC".
`cellTypes`	Specify cell type data in reference data will be used for deconvolution. if "NULL" all cell types will be used.
`normalize`	TRUE or FALSE, if TRUE, quantile normalization on methylated or unmethylated intensities will be performed.
`nonnegative`	TRUE or FALSE. If TRUE, the estimated proportions will be constrained to nonnegative values
`nProbes`	Number of best probes for each cell types will be used for the estimation.
`refplot`	TRUE or FALSE. IF TRUE, refdata distribution and heatmap will be plotted for inspection of reference dataset.

Details

This function use the method in Houseman et al (2012) to estimate cell type proportions based on DNA methylation data.

The following reference datasets can be used to assist the estimation. User should select a reference most resemble to user's data in tissue, age, and array type.

FlowSorted.Blood.450k: consisting of 450K methylation data for 60 blood samples from 6 male adults. 6 samples each for cell types: Bcell CD4T CD8T Eos Gran Mono Neu NK PBMC WBC; See Reinius et al. 2012 for details.

FlowSorted.CordBlood.450k: consisting of 450k methylation data for 104 cord blood samples from 17 male and female individuals. Cell type (# samples) are: Bcell(15) CD4T(15) CD8T(14) Gran(12) Mono(15) NK(14) nRBC(4) WholeBlood(15). See Bakulski et al. Epigenetics 2016 for details.

FlowSorted.CordBloodNorway.450k: consisting of 450K methylation data for 77 cord blood samples from 11 individuals (6 girls and 5 boys). 11 samples for each of the cell types: Bcell CD4T CD8T Gran Mono NK WBC. See P Yousefi et al Environ. Mol. Mutagen 2015 for details.

FlowSorted.Blood.EPIC: consisting of EPIC methylation data for 37 magnetic sorted blood cell references and 12 samples. See LA Salas et al. 2018 for details.

FlowSorted.DLPFC.450k: consisting of 450K methylation data for 58 brain tissue samples from 29 individuals. 15 females and 14 males, 6 Africans and 23 Caucasians, age between 13 to 79. 29 samples for each of the cell types: NeuN_neg and NeuN_pos. See Guintivano et al. 2013 for details.

FlowSorted.CordBloodCombined.450k: consisting of 289 combined umbilical cord blood cells samples assayed by Bakulski et al, Gervin et al., de Goede et al., and Lin et al. see https://github.com/immunomethylomic
details.

Value

A matrix of composition estimates with columns for cell types and rows for samples.

Author(s)

Zongli Xu

References

EA Houseman, WP Accomando, DC Koestler, BC Christensen, CJ Marsit, HH Nelson, JK Wiencke and KT Kelsey. *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC bioinformatics (2012) 13:86.

Examples

```
## Not run:
require(minfidata)
path <- file.path(find.package("minfiData"),"extdata")
#based on rgDataset
rgSet <- readidat(path = path,recursive = TRUE)
celltype=estimateCellProp(userdata=rgSet,refdata="FlowSorted.Blood.450k",
         nonnegative = TRUE,normalize=TRUE)
#using methDataSet
mdat<-preprocessENmix(rgSet, bgParaEst="oob", dyeCorr="RELIC",
                      QCinfo=qc, exCpG=outCpG, nCores=6)
celltype=estimateCellProp(userdata=mdat,refdata="FlowSorted.Blood.450k",
        nonnegative = TRUE,normalize=TRUE)
```

```
mdat<-norm.quantile(mdat, method="quantile1")
#using beta value
beta<-rcp(mdat,qcscore=qc)
celltype=estimateCellProp(userdata=beta,refdata="FlowSorted.Blood.450k",
         nonnegative = TRUE)

## End(Not run)
```

`freqpoly` *Frequency polygon plot*

Description

Similar to histogram, frequency polygon plot can be used to inspect 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])
}}
```

getB *Extract methylation Beta values.*

Description

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

Usage

```
getB(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

getCGinfo *CpG probe annotation inforamtion*

Description

Extract CpG probe annotation inforamtion from a rgDataSet

Usage

```
getCGinfo(rgSet, type="IandII")
```

Arguments

rgSet	An object of class rgDataSet
type	One of the following options "I","II","IandII","ctrl". "all"

Value

An object of dataframe class

Author(s)

Zongli Xu

Examples

```
## Not run:
require(minfidata)
path <- file.path(find.package("minfiData"),"extdata")
#based on rgDataset
rgSet <- readidat(path = path,recursive = TRUE)
info=getCGinfo(rgSet,type="all")

## End(Not run)
```

`getmeth` *Create a methDataSet*

Description

To create a methDataSet based on a rgDataset

Usage

```
getmeth(rgSet)
```

Arguments

`rgSet`	An object of class `rgDataSet`

Details

CpG annotation information is incorporated in the output object, and can be extracted using command `rowData`

Value

An object of class `methDataSet`

Author(s)

Zongli Xu

Examples

```
## Not run:
require(minfidata)
path <- file.path(find.package("minfiData"),"extdata")
#based on rgDataset
rgSet <- readidat(path = path,recursive = TRUE)
meth=getmeth(rgSet)

## End(Not run)
```

ipdmr *Differentially methylated region finder using interval P values*

Description

To identify differentially methylated regions using an interval P value method

Usage

```
ipdmr(data,include.all.sig.sites=TRUE,dist.cutoff=1000,bin.size=50,
      seed=0.01,region_plot=TRUE,mht_plot=TRUE)
```

Arguments

data	A data frame from bed format file with colname name "V1","V2", "V3","V4","V5", V1 indicate chromosome (1,2,3,...,X,Y), V2 is chromosome position, V4 is for P value and V5 for name of CpGs
include.all.sig.sites	Whether to use CpG singletons in calculate of FDR
dist.cutoff	Maximum distance in base pair to combine adjacent DMRs
bin.size	bin size for autocorrelation calculation
seed	FDR significance threshold for initial selection of DMR regions
region_plot	Whether to draw regional plot
mht_plot	Whether to draw mahattan plot

Details

The input should be a data frame with column names V1-V5, indicating chromosome, start position, end position, P value and probe name. The function will use a novel interval p value method to identify differentially methylated regions. DMR results will be stored in a file with name "resu_ipdmr.csv". If plot options were selected, two figure files will be generated: mht.jpg and region_plot.pdf.

Author(s)

Liang Niu, Zongli Xu

References

Zongli Xu, Changchun Xie, Jack A. Taylor, Liang Niu, ipDMR: Identification of differentially methyl-ated regions with interval p-values, in review

M2B *Convert methylation M value to methylation beta value.*

Description

Convert methylation M value to methylation beta value.

Usage

```
M2B(x)
```

Arguments

x	An numeric matrix

Details

Methylation beta value is calculated as beta=M/(M+U+a). M is methylated intensity, U is unmethylated intensity, and a is a constant offset (by default , a=100). M value is calculated as M=log2((M+a)/(U+a)). M or U is usually greater than 1000, so a is negligible for most probes. if a=0, then beta=2^M/(2^M+1).

Value

A matrix of methylation Beta values.

Author(s)

Zongli Xu

methDataSet-class *Class "methDataSet"*

Description

A class for storing Illumina methylation array methylated and unmethylated intensity data, and CpG annotation infomation.

Usage

```
methDataSet(Meth = new("matrix"), Unmeth = new("matrix"),
    rowData=new("DataFrame"),...)
```

Arguments

Meth	A matrix of methylated intensity values with row for CpGs and column for samples
Unmeth	A matrix of unmethylated intensity values with row for CpGs and column for samples
rowData	A dataframe contains CpG annotation information
...	Other arguments for class SummarizedExperiment

Details

CpG annotation information is incorporated in the object, and can be extracted using command `rowData`

Value

An object of class `methDataSet`

Examples

```
showClass("methDataSet")
```

`methyAge` *Methylation Age estimation*

Description

To calculate Methylation Age using Hovath, Hannum or PhenoAge methods

Usage

```
methyAge(beta,type="all",fastImputation=FALSE,normalize=TRUE,nCores=2)
```

Arguments

`beta`	Methylation beta value matrix with CpG names(row names) and sample ids(column names).
`type`	Which method will be used for methylation age calculation: "hovath", "hannum", "phenoAge" or "all"
`fastImputation`	If "TRUE" reference methylation values will used for imputation, if "FALSE", KNN nearest neighbor method will be used.
`normalize`	TRUE or FALSE, if TRUE, Hovath modified BMIQ method will be used to perform normalization.
`nCores`	Number of cores will be used for normalization

Value

A data frame with rows for sample and columns for estimates methylation age.

Author(s)

Zongli Xu

References

Horvath S. *DNA methylation age of human tissues and cell types*. Genome biology 2013 14:R115. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. *Genome-wide methylation profiles reveal quantitative views of human aging rates*. Molecular cell 2013 49:359-367. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. *An epigenetic biomarker of aging for lifespan and healthspan*. Aging (Albany NY) 2018 10:573-591.

Examples

```
## Not run:
require(minfidata)
path <- file.path(find.package("minfiData"),"extdata")
#based on rgDataset
rgSet <- readidat(path = path,recursive = TRUE)
meth=getmeth(rgSet)
beta=getB(meth)
mage=methyAge(beta)

## End(Not run)
```

`mhtplot` *P value manhattan plot*

Description

P value manhattan plot

Usage

```
mhtplot(dat,color="bg",sigthre=NULL,markprobe=NULL,
         markcolor="red", outf="mht.jpg")
```

Arguments

`dat`	A data frame from bed format file with colname name "V1","V2", "V3","V4","V5", V1 indicate chromosome (1,2,3,...,X,Y), V2 and V3 are chromosome position (V3 is an optional variable), V4 is for P value and V5 for name of CpGs
`color`	Color scheme of manhattan plot, "bg" indicate "black and gray"
`sigthre`	P value of significant threshold line
`markprobe`	A list of CpGs to be marked out
`markcolor`	Color code for marked probe
`outf`	figure file name, default "mht.jpg"

Details

Draw manhattan plot based on bed file format input file. Optional and colored marks fo a subset of probes

Author(s)

Zongli Xu

mpreprocess *methylation QC and preprocessing pipeline for Illuminal arrays*

Description

The pipeline performs background correction, dye bias correction, inter-array normalization and probe type bias correction for HumanMethylation 450 and MethylationEPIC BeadChip data. It removes or mitigates background noise and systematic experimental bias,It also perform quality controls, identifing and excluding low quality samples and probes, removing outlier values, and performing 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 rgDataSet,methDataSet, 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*

Description

Produce Frequency polygon plot for each column of a numeric data matrix. Similar to multidensity function, the plot can be used to inspect data distribution but with much faster speed.

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 for each row of data*

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 (so called gap 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 uses 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 indicating number of mode. Gap probes are probes with number of mode greater than 1.

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: "quantile1", "quantile2", or "quantile3".

Details

By default, method = "quantile1", which 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

To perform quantile normalization with methylation intensity data across samples

Usage

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

Arguments

mdat	An object of class MethylSet.
method	Quantile normalization method: "quantile1", "quantile2", or "quantile3".

Details

By default, method = "quantile1", which 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

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*, Bioinformatics. 2016

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)
```

`p.qqplot` *P value Q-Q plot*

Description

P value Q-Q plot with optional confidence interval

Usage

```
p.qqplot(pvalues,outf="qq.jpg",draw.conf=TRUE,
   conf.col="lightgray",conf.alpha=.95,pch=20,col="black")
```

Arguments

`pvalues`	An numeric list of P values
`outf`	figure file name, default "qq.jpg"
`draw.conf`	Whether to draw confidence interval of expected P values under NULL hypothesis
`conf.col`	Color code of confidence interval
`conf.alpha`	Confidence interval range, 0.95 in default
`pch`	Point type code
`col`	Point color code

Details

P value Q-Q plot with optional confidence interval

Author(s)

Zongli Xu

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 variable. 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 *Internal control plot*

Description

The function will generate a series of internal control plots, which are similar to the plots from Illumina genomic studio software. Users should refer to the software's online guide to interpret these figures. 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)
}}
```

predSex *Estimating sample sex*

Description

Estimating sample sex based on methylation data

Usage

```
predSex(mdat, cutoff = -2)
```

Arguments

mdat	An object of class MethDataSet or rgDataSet.
cutoff	The difference in log2 total intensity between X and Y chromosomes

Details

Estimation of sex is based on the difference of log2 median total intensity measures on the X and Y chromosomes.

Author(s)

Zongli Xu

Examples

```
## Not run:
path <- file.path(find.package("minfiData"),"extdata")
rgSet <- readidat(path = path,recursive = TRUE)
sex=predSex(rgSet)

## End(Not run)
```

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

Examples

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

QCfilter *Sample or CpG probe filter.*

Description

Filtering out 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, detPtype="negative", 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
detPtype	Calculate detection P values based on negtive internal control ("negative") probes or out of the band ("oob") probes
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, detPthre=0.000001)
```

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)
}}
```

---

readidat *Parsing IDAT files from Illumina methylation arrays.*

Description

Read in IDAT files and create a rgDataSet with probe annotation

Usage

```
readidat(path = NULL,manifestfile=NULL,recursive = FALSE, verbose = FALSE)
```

Arguments

path	Directory where idat files are located
manifestfile	The file name of array manifestfile, which can be downloaded from Illumina website. Bioconductor manifest package will be used if not provided
recursive	if TRUE, idat files in the subdirectories will also be read in
verbose	if TRUE, detailed running info will be printed on screen

Details

Probe annotation info can be extracted using command rowData

Value

An object of class rgDataSet,

Author(s)

Zongli Xu

Examples

```
## Not run:
require(minfidata)
path <- file.path(find.package("minfiData"),"extdata")
#based on rgDataset
rgSet <- readidat(path = path,recursive = TRUE)

## End(Not run)
```

---

readmanifest *Parsing Illumina methylation arrays manifest file.*

Description

Parsing Illumina methylation arrays manifest file.

Usage

```
readmanifest(file)
```

Arguments

file	Illumina methylation array manifest file, downloaded from Illuminal website

Value

An object of dataframe caintaining probe annotation information

Author(s)

Zongli Xu

Examples

```
## Not run:
manifestfile="manifest *.csv file path"
manifest=readmanifest(manifestfile)

## End(Not run)
```

`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)
}}
```

---

rgDataSet-class *Class* "rgDataSet"

Description

A class for storing Illumina methylation array raw intensity data of two color channels, and probe annotation infomation.

Usage

```
rgDataSet(Red = new("matrix"), Green = new("matrix"),
    NBeads = new("matrix"),rowData=new("DataFrame"),ictrl= new("DataFrame"),...)
```

Arguments

Red	A matrix of Red channel intensity values with row for methylation probes and column for samples
Green	A matrix of Green channel intensity values with row for methylation probes and column for samples
NBeads	A matrix contains the number of beads used to generate intensity values on the Red and Green channels.
rowData	A dataframe contains probe annotation information
ictrl	A dataframe contais detailed information for Illumina internal control probes
...	other arguments for class SummarizedExperiment

Value

An object of class rgDataSet

Examples

```
showClass("rgDataSet")
```

`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 imputation
b3=rm.outlier(beta,qcscore=qcscore,rmcr=TRUE,impute=TRUE)
}}
```

Index