The ENmix User's Guide

Modified: April 27 2020. Compiled: April 28, 2020

1 Introduction

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.

The ENmix package can also support the data structure used by several other related R packages, such as minfi, wateRmelon and ChAMP, providing straightforward integration of ENmix-corrected datasets for subsequent data analysis.

The software is designed to support large scale data analysis, and provides multi-processor parallel computing options for most functions.

2   Citation

The following publications can be refered to learn more about the methods implemented in this package.

Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina Human-Methylation450 BeadChip, Nucleic Acids Research, 2015

Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics, 2016

Xu Z, Langie SA, De Boever P, Taylor JA, Niu L. RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip. BMC Genomics. 2017

Xu Z, Taylor JA, Leung YK, Ho SM, Niu L. oxBS-MLE: an efficient method to estimate 5-methylcytosine and 5-hydroxymethylcytosine in paired bisulfite and oxidative bisulfite treated DNA. Bioinformatics. 2016

Xu Z, Xie C, Taylor JA, Niu L. ipDMR: Identification of differentially methylated regions with interval P values in review, 2020

3   List of functions

4   Example Analysis

4.1   Example 1: using pipeline

Pipeline function `mpreprocess` can be used to perform background correction, dye-bias adjustment, inter-array normalization, probe-type bias correction, as well as outlier removal and imputation in a single step.

```
> library(ENmix)
> #read in data
> require(minfiData)
> #data pre-processing
> beta=mpreprocess(RGsetEx,nCores=6)
> #or
> #read in IDAT files
> path <- file.path(find.package("minfiData"),"extdata")
> rgSet <- readidat(path = path,recursive = TRUE)
> #quality control and data pre-processing
```

Function	Description
Data acquisition	
readidat	Read idat files into R
readmanifest	Read array manifest file into R
Quality control	
QCinfo	Extract QC info and identify outlier samples
plotCtrl	Generate internal control plots
getCGinfo	Extract CpG probe annotation information
calcdetP	Estimate detection P values
Qcfilter	Filter out low quality CpGs or samples
rm.outlier	Remove low quality values, samples or CpGs; remove outlier samples and do imputation
nmode.mc	Identify gap probes
dupicc	Calculate ICC based on duplicates
multifreqpoly	Frequency polygon plot
Preprocessing	
mpreprocess	Preprocessing pipeline
preprocessENmix	ENmix background correction and dye bias correction
relic	RELIC dye bias correction
norm.quantile	Quantile normalization
rcp	RCP probe design type bias correction
Differential analysis	
ipdmr	ipDMR differentially methylated region analysis
combp	Combp differentially methylated region analysis
Other functions	
oxBS.MLE	MLE estimates of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC)
estimateCellProp	Cell type proportion Estimation
methyAge	Methylation age estimates
predSex	Sample sex estimation
ctrlsva	Estimates of non-negative control surrogate variables
pcrplot	Principal component regression plot
mhtplot	P value manhattan plot
p. qqplot	P value Q-Q plot with optional confidence interval
B2M	Convert Beta value to M value
M2B	Convert M value to Beta value

```
> beta=mpreprocess(rgSet,nCores=6,qc=TRUE,foutlier=TRUE,
+      rmcr=TRUE,impute=TRUE)
```

4.2   Example 2: using individual function

This example code is basically doing the same thing as in example 1, but more transparent, and
has more customized options.

```
> library(ENmix)
> #read in data
> path <- file.path(find.package("minfiData"),"extdata")
> rgSet <- readidat(path = path,recursive = TRUE)
> #QC info
> qc<-QCinfo(rgSet)
> #background correction and dye bias correction
> #if provide qc info, the low quality samples and probes
> #will be excluded before background correction
> mdat<-preprocessENmix(rgSet, bgParaEst="oob", dyeCorr="RELIC",
+ QCinfo=qc, nCores=6)
> #low quality samples and probes can also be excluded after
> #background correction using QCfilter.
> #mdat <- QCfilter(mdat,qcinfo=qc,outlier=TRUE)
> #inter-array normalization
> mdat<-norm.quantile(mdat, method="quantile1")
> #probe-type bias adjustment
> beta<-rcp(mdat,qcscore=qc)
> beta <- rm.outlier(beta,qcscore=qc,impute=TRUE,rmcr=TRUE)
```

4.3   Example 3: A more elaborated example

This example is to demonstrate the flexibility of the package functions. Getting detailed informa-
tion about data quality, SNP-like probes, excluding user specified probes and/or samples, principal
component regression analysis, outlier removal, sex prediction, cell sub-type proportion estimates
and control surrogate variables.

```
> library(ENmix)
> #read in data
> path <- file.path(find.package("minfiData"),"extdata")
> rgSet <- readidat(path = path,recursive = TRUE)
> sheet <- read.metharray.sheet(file.path(find.package("minfiData"),"extdat
+          pattern = "csv$")
> rownames(sheet)=paste(sheet$Slide,sheet$Array,sep="_")
```

```
> colData(rgSet)=as(sheet[colnames(rgSet),],"DataFrame")
> #control plots
> plotCtrl(rgSet)
> #QC info
> qc<-QCinfo(rgSet)
> #calculate detection P values
> detp=calc_detP(rgSet,detPtype = "negative")
> detp2=calc_detP(rgSet,detPtype = "oob")
> #get methDataSet
> mraw <- getmeth(rgSet)
> #get raw methyaltion beta values
> beta<-getB(mraw)
> #distribution plot
> multifreqpoly(beta,main="Methylation Beta value distribution")
> #Search for multimodal CpGs
> #sample size in this example data is too small for this purpose!
> #exclude low quality data first
> bb=beta; bb[qc$detP>0.05 | qc$nbead<3]=NA
> nmode<-nmode.mc(bb, minN = 3, modedist=0.2, nCores = 6)
> outCpG = names(nmode)[nmode>1]
> #background correction and dye bias correction
> mdat<-preprocessENmix(rgSet, bgParaEst="oob", dyeCorr="RELIC",
+                       QCinfo=qc, exCpG=outCpG, nCores=6)
> #inter-array normalization
> mdat<-norm.quantile(mdat, method="quantile1")
> colData(mdat)=as(sheet[colnames(mdat),],"DataFrame")
> #probe-type bias adjustment
> beta<-rcp(mdat,qcscore=qc)
> # Principal component regression analysis plot
> cov<-data.frame(group=colData(mdat)$Sample_Group,
+     slide=factor(colData(mdat)$Slide))
> pcrplot(beta, cov, npc=6)
> #filter out low quality and outlier data points for each probe;
> #rows and columns with too many missing value can be removed
> #if specify; Do imputation to fill missing data if specify.
> beta <- rm.outlier(beta,qcscore=qc,rmcr=TRUE,impute=TRUE)
> #Non-negative control surrogate variables
> csva<-ctrlsva(rgSet)
> #estimate cell type proportions
> #based on rgDataset
> celltype=estimateCellProp(userdata=rgSet,refdata="FlowSorted.Blood.450k",
+          nonnegative = TRUE,normalize=TRUE)
```

```
> #using methDataSet
> celltype=estimateCellProp(userdata=mdat,refdata="FlowSorted.Blood.450k",
+          nonnegative = TRUE,normalize=TRUE)
> #using beta value
> celltype=estimateCellProp(userdata=beta,refdata="FlowSorted.Blood.450k",
+          nonnegative = TRUE,normalize=TRUE)
> #predic sex based on rgDataSet or methDataset
> sex=predSex(rgSet)
> sex=predSex(mdat)
> #Methylation age calculation
> mage=methyAge(beta)
> #ICC analysis, see manual
> #DMR analysis, see manual
```

5   Setting up the data

The first step is to import array raw data files (*.idat) to create an object of `rgDataSet`, which
is similar to `RGChannelSetExtended` in minfi package, but with probe annotation integrated
into data object, and having smaller data size. User can also use minfi package to read in idat files
and create `RGChannelSetExtended`. Most functions in ENmix support `RGChannelSetExtended`
or `RGChannelSet`.

```
> library(ENmix)
> rgSet <- readidat(path = "directory containing idat files",recursive = TR
```

When methylation IDAT raw data files are not available, such as in many publically available
datasets, users can use methylated (M) and unmethylated (U) intensity data to create an object of
`MethylSet`.

```
> M<-matrix_for_methylated_intensity
> U<-matrix_for_unmethylated_intensity
> pheno<-as.data.frame(cbind(colnames(M), colnames(M)))
> names(pheno)<-c("Basename","filenames")
> rownames(pheno)<-pheno$Basename
> pheno<-AnnotatedDataFrame(data=pheno)
> anno<-c("IlluminaHumanMethylation450k", "ilmn12.hg19")
> names(anno)<-c("array", "annotation")
> mdat<-MethylSet(Meth = M, Unmeth = U, annotation=anno,
+ phenoData=pheno)
```

6

6 Quality Control

6.1 Internal control probes

Illumina 450k chip incorporated 15 different types of internal control probes (total of 848 probes). The control plots generated using the data by Illumina GenomeStudio software are very useful to inspect experimental process and data quality. However, GenomeStudio only works on the windows operating system, it is time consuming to generate these plots for larger dataset, and there is also no option to save the plots into file. The function `plotCtrl` can generate similar plots for each type of control.

```
> plotCtrl(rgSet)
```

See Illumina Infinium HD Methylation Assay for detailed description on how to interpret these control figures. Here is a list of control types:

Control types	Number of probes
Sample-Independent Controls	
STAINING	4
EXTENSION	4
HYBRIDIZATION	3
TARGET REMOVAL	2
RESTORATION	1
Sample-Dependent Controls	
BISULFITE CONVERSION I	12
BISULFITE CONVERSION II	4
SPECIFICITY I	12
SPECIFICITY II	3
NON-POLYMORPHIC	4
NORM_A	32
NORM_C	61
NORM_G	32
NORM_T	61
NEGATIVE	613

These controls can also be plotted in user specified order to check how experimental factors affect methylation measures, such as batch, plate, array or array location.

```
> path <- file.path(find.package("minfiData"),"extdata")
> rgSet <- readidat(path = path,recursive = TRUE)
> sheet <- read.metharray.sheet(file.path(find.package("minfiData"),"extdat
+     pattern = "csv$")
> rownames(sheet)=paste(sheet$Slide,sheet$Array,sep="_")
```
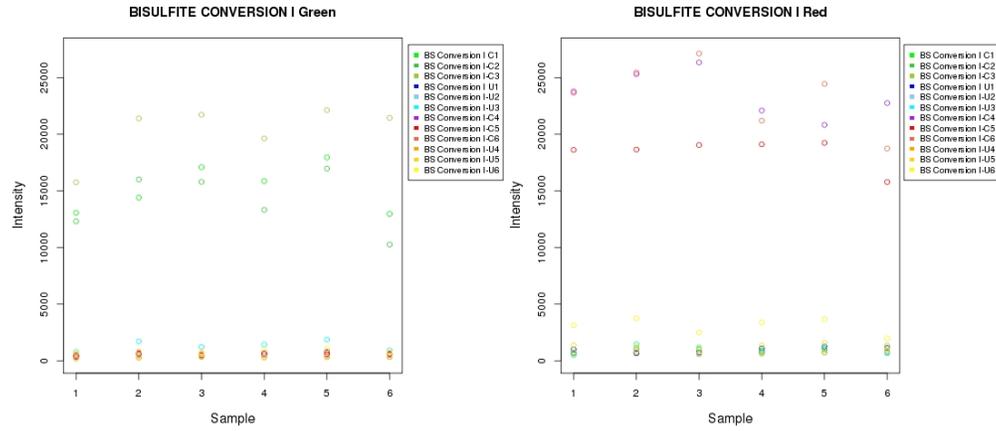
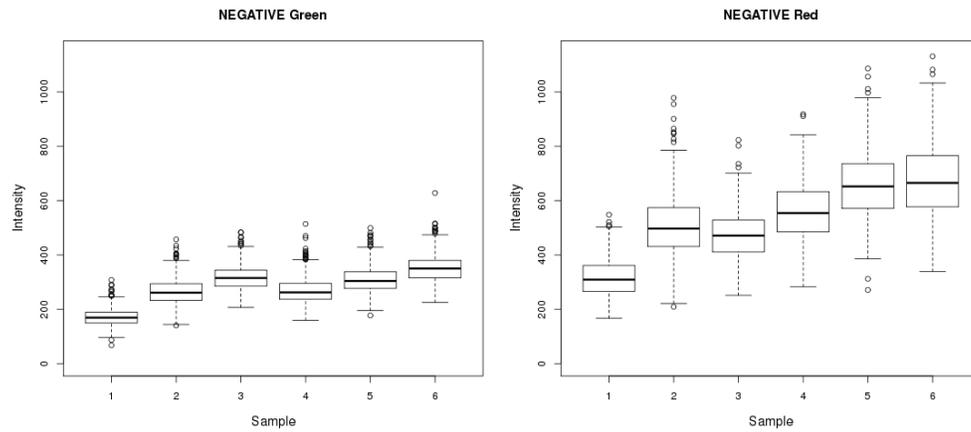Figure 1: Bisulfite conversion controls for type I probes
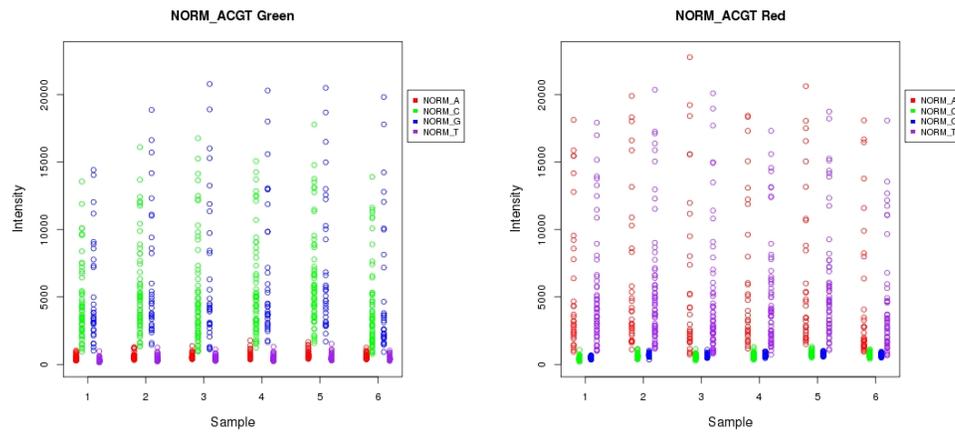

Figure 2: Negative control probes

Figure 3: NORM ACGT control probes

```
> colData(rgSet)=as(sheet[colnames(rgSet),],"DataFrame")
> #control plots
> IDorder=rownames(colData(rgSet))[order(colData(rgSet)$Slide,colData(rgSet)
> plotCtrl(rgSet,IDorder)
```

6.2 Data distribution

Methylation intensity or beta value distribution plots are very useful for data summary, visual inspection and identification of outlier samples. Density plot is routinely generated using R function `multidensity`. However, the function is computationally intensive, and can take several hours to produce density plots for a large methylation dataset. Furthermore, density plot is difficult to understand for many investigators, also as noted in the man page of `multidensity`, density plot may not be able to display data distribution accurately for some data because of the smooth function which may lead part of the distribution to be out of range, and may obscure important details in data distribution.

ENmix's frequency polygon plot provides a better alternative for inspection of data distribution. It can accurately reflect data distribution and, like histogram it is easy to understand. It is also much faster, and only take a few minutes to produce a distribution plot for >1000 samples.

```
> mraw <- getmeth(rgSet)
> #total intensity plot is userful for data quality inspection
> #and identification of outlier samples
> multifreqpoly(assays(mraw)$Meth+assays(mraw)$Unmeth,
+ xlab="Total intensity")
> #Compare frequency polygon plot and density plot
> beta<-getB(mraw)
> anno=rowData(mraw)
> beta1=beta[anno$Infinium_Design_Type=="I",]
> beta2=beta[anno$Infinium_Design_Type=="II",]
> library(geneplotter)
> jpeg("dist.jpg",height=900,width=600)
> par(mfrow=c(3,2))
> multidensity(beta,main="Multidensity")
> multifreqpoly(beta,main="Multifreqpoly",xlab="Beta value")
> multidensity(beta1,main="Multidensity: Infinium I")
> multifreqpoly(beta1,main="Multifreqpoly: Infinium I",
+ xlab="Beta value")
> multidensity(beta2,main="Multidensity: Infinium II")
> multifreqpoly(beta2,main="Multifreqpoly: Infinium II",
+ xlab="Beta value")
> dev.off()
```

See the following figures (Figure 4) generated from the above code. When type I and type II probes are plotted separately (Fig 4 bottom 4 panels) the difference in modes between type I and II probes can be appreciated. But when all probes are plotted together (Fig 4 top panels), the multidensity plot obscures these differences, while they remain readily apparent in the multifreqpoly plot. In addition, the multidensity plots appear to suggest that probes range in value from <0 to >1, whereas multifreqpoly correctly show the range from 0 to 1.

6.3   QC information, filtering of low quality samples and probes

Data quality measures, including detection P values, number of beads for each methylation read and average intensities for bisulfite conversion probes can be extracted using the function `QCinfo` from an object of `RGChannelSetExtended`. Based on default or user specified quality score thresholds, the `QCinfo` can also identify and export a list of low quality samples and CpG probes. Outlier samples based on total intensity or beta value distribution should be excluded before further analysis. Such samples have been difficult to identify, but by using the argument outlier=TRUE, these outlier samples will be identified automatically. Data quality score figures from `QCinfo` can be used to guide the selection of quality score thresholds. Low quality samples and probes can be filtered out using `QCfilter` or `preprocessENmix`.

```
> qc<-QCinfo(rgSet)
> #exclude before backgroud correction
> mdat<-preprocessENmix(rgSet, bgParaEst="oob", dyeCorr="RELIC",
+ QCinfo=qc, nCores=6)
> #Or exclude after background correction
> mdat <- QCfilter(mdat,qcinfo=qc,outlier=TRUE)
```

6.4   Filtering out outliers and low quality data values

Outlier and low quality data values can have a large impact on association statistical tests. Function `rm.outlier` can filter out these data points and replace them as missing values. Outliers are defined as values smaller than 3 times IQR from the lower quartile or larger than 3 times IQR from the upper quartile. Some statistical methods do not allow missing values, argument *impute=TRUE* in the function can be specified to impute missing data using k-nearest neighbors method.

```
> #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
```
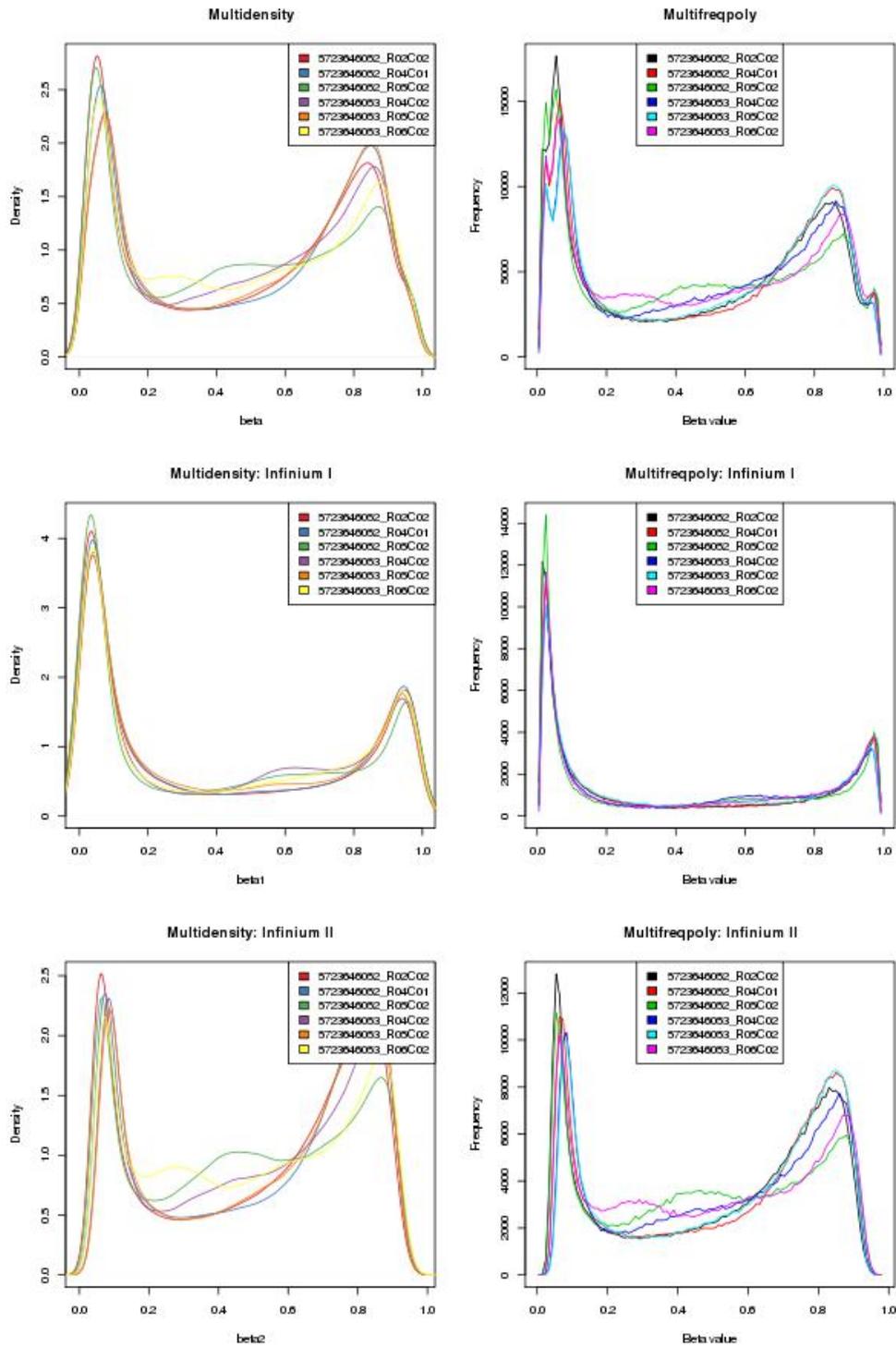
Figure 4: Methylation beta value distribution plots for all probes (top 2 panels) and for type I (middle panels) and II (bottom panels) probes separately. The smoothing function in multidensity plots (panels on left) results in misleading range and mode information which are more accurately depicted in the multifreqpoly plots (panels on right)

```
> # with too many missing values, and then do imputation
> b3=rm.outlier(beta,qcscore=qcscore,rmcr=TRUE,impute=TRUE)
```

7 Background correction and dye-bias adjustment

Function `preprocessENmix` incorporates a model based background correction method *ENmix*, which models methylation signal intensities with a flexible exponential-normal mixture distribution, together with a truncated normal distribution to model background noise. Users can also specify a list of poor performance CpGs to be excluded before background correction using argument *exCpG*. Argument dyeCorr can be used to specify a method for dye-bias correction, the default is RELIC.

See the following papers for the detailed description of related methods:

Zongli Xu, et. al. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip, Nucleic Acids Research, 2015

Xu Z, Langie SA, De Boever P, Taylor JA, Niu L. RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip. BMC Genomics. 2017

If argument QCinfo is specified, the low quality samples and probes identified by function `QCinfo` will be excluded before ENmix background correction. Using argument *exSample* and *exCpG*, User can also specify a list of samples or probes to be excluded before background correction.

```
> qc=QCinfo(rgSet)
> mdat<-preprocessENmix(rgSet, bgParaEst="oob", dyeCorr="RELIC",
+ QCinfo=qc, exCpG=NULL, nCores=6)
```

8 Inter-array normalization

Function `norm.quantile` can be used to perform quantile normalization on methylation intensity values.

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

9 Probe type bias adjustment

The majority of probes on Illumina 450K and EPIC BeadChips are type II probes. Although type II probes facilitate increased array genome coverage, they were shown to have decreased dynamic range and reproducibility compared to type I probes. Taking advantage of the high spatial correlation of DNA methylation levels along the human genome, The RCP (Regression on Correlated

Probes) method utilizes nearby (<25 bp) type I and II probe pairs to derive the quantitative relationship between probe types and then recalibrates type II probe measurements using type I probes as referents.

```
> beta<-rcp(mdat)
```

See the following publication for the detailed description of the method:

Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics, 2016

10   Batch effect correction

Function `ctrlsva` can be used to estimate surrogate variables for batch effect and unknown experimental confounders from intensity data for non-negative internal control probes. These variables can then be adjusted as covariables in downstream association analysis to remove unwanted data variation.

```
> sva<-ctrlsva(rgSet)
```

11   Principal component regression analysis plot

First, principal component analysis will be performed in standardized beta value matrix (standardized for each 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, such as batch or environmental variables. Regression P values will be plotted to explore methylation data variance structure and to identify possible confounding variables to guide association statistical analysis.

```
> cov<-data.frame(group=colData(mdat)$Sample_Group,
+     slide=factor(colData(mdat)$Slide))
> pcrplot(beta, cov, npc=6)
```

12   Multimodal CpGs

Function `nmode.mc` uses an empirical approach to identify multimodal distributed CpGs (SNP like probes). When measured in a population of people the majority of CpGs on the Illumina HumanMethylation450 BeadChip have unimodal distributions of DNA methylation values with relatively small between-person variation. However, some CpGs (typically around 10,000 in 450k array often seemingly the result of SNPs in the probe region) may have multimodal distributions of
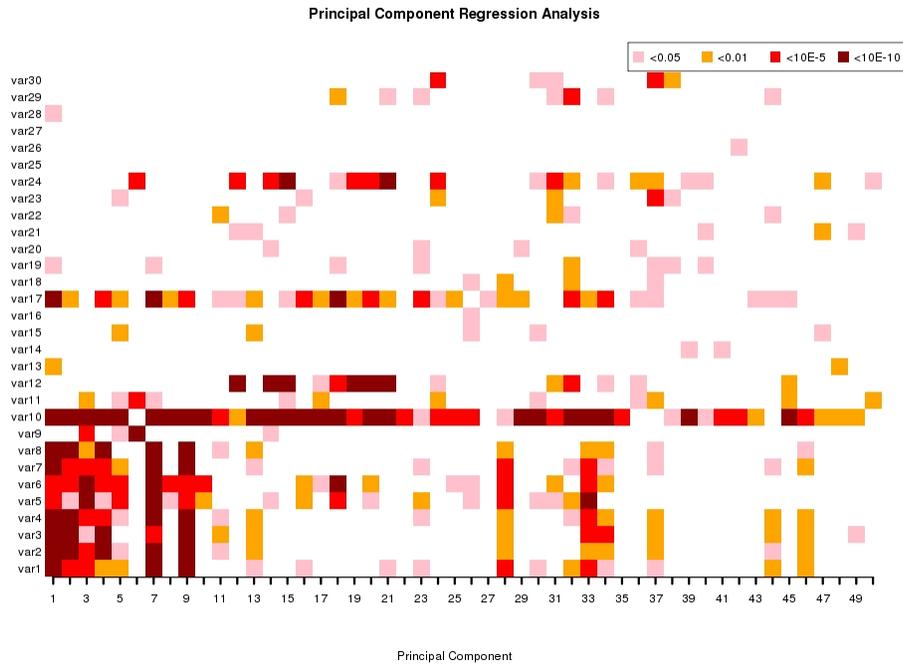
Figure 5: Example principal component regression p value plot of raw data generated using 450K methylation data from a published study

methylation values with sizeable differences between modes and large between-person variation. These multimodal distributed data are usually caused by SNP effect, problematic probe design or other unknown artifacts instead of actual methylation level and thus should be excluded from DNA methylation analysis. Researchers have often excluded CpGs based on SNP annotation information. However, because SNP annotation always depends on population origin, we found that this approach alone may exclude many well-distributed (unimodal) CpGs, while still failing to identify other multi-modal CpGs. We developed an empirical approach to identify CpGs that are obviously not uni-modally distributed, so that researchers can make more informed decisions about whether to exclude them in their particular study populations and analyses.

See online supplementary materials of the following paper for an evaluation of the method using published EWAS data.

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

```
> nmode<- nmode.mc(beta, minN = 3, modedist=0.2, nCores = 5)
```

14

13 Compatibility with other related R packages

The ENmix uses the data structure provided by R minfi packages as input and output, and thus is fully compatible with the minfi package. The same data structures were also used by several other R packages, such as ChAMP and wateRmelon, so the output from ENmix functions can be easily utilized in these packages for further analysis. Here are some examples:

Example 1: mixed use of minfi and ENmix functions

```
> library(ENmix)
> #minfi functions to read in data
> sheet <- read.metharray.sheet(file.path(find.package("minfiData"),
+  "extdata"), pattern = "csv$")
> rgSet <- read.metharray.exp(targets = sheet, extended = TRUE)
> #ENmix function for control plot
> plotCtrl(rgSet)
> #minfi functions to extract methylation and annotation data
> mraw <- preprocessRaw(rgSet)
> beta<-getB(mraw, "Illumina")
> anno=getAnnotation(rgSet)
> #ENmix function for fast and accurate distribution plot
> multifreqpoly(beta,main="Data distribution")
> multifreqpoly(beta[anno$Type=="I",],main="Data distribution, type I")
> multifreqpoly(beta[anno$Type=="II",],main="Data distribution, type II")
> #ENmix background correction
> mset<-preprocessENmix(rgSet, bgParaEst="oob", dyeCorr="RELIC", nCores=6)
> #minfi functions for further preprocessing and analysis
> gmSet <- preprocessQuantile(mset)
> bumps <- bumphunter(gmSet, design = model.matrix(~ gmSet$status), B = 0,
+ type = "Beta", cutoff = 0.25)
```

Example 2: add ENmix background correction step into ChAMP pipeline

```
> library(ENmix)
> library(ChAMP)
> testDir=system.file("extdata",package="ChAMPdata")
> myLoad=champ.load(directory=testDir)
> #ENmix background correction
> mset<-preprocessENmix(myLoad$rgSet,bgParaEst="oob", nCores=6)
> #remove probes filtered by champ.load()
> mset=mset[rownames(myLoad$beta),]
> #update myLoad object with background corrected intensity data
> myLoad$mset=mset
> myLoad$beta=getB(mset)
```

```
> myLoad$intensity=getMeth(mset)+getUnmeth(mset)
> #continue ChAMP pipeline
> myNorm=champ.norm()
```

14   SessionInfo

- R version 4.0.0 (2020-04-24), `x86_64-pc-linux-gnu`

- Locale: `LC_CTYPE=en_US.UTF-8`, `LC_NUMERIC=C`, `LC_TIME=en_US.UTF-8`,
  `LC_COLLATE=C`, `LC_MONETARY=en_US.UTF-8`, `LC_MESSAGES=en_US.UTF-8`,
  `LC_PAPER=en_US.UTF-8`, `LC_NAME=C`, `LC_ADDRESS=C`, `LC_TELEPHONE=C`,
  `LC_MEASUREMENT=en_US.UTF-8`, `LC_IDENTIFICATION=C`

- Running under: `Ubuntu 18.04.4 LTS`

- Matrix products: default

- BLAS: `/home/biocbuild/bbs-3.11-bioc/R/lib/libRblas.so`

- LAPACK: `/home/biocbuild/bbs-3.11-bioc/R/lib/libRlapack.so`

- Base packages: base, datasets, grDevices, graphics, methods, stats, utils

- Loaded via a namespace (and not attached): compiler 4.0.0, tools 4.0.0

15   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

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

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.

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

Illumina Inc., Infinium HD Assay Methylation Protocol Guide, Illumina, Inc. San Diego, CA.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD and Irizarry RA (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. Bioinformatics, 30(10), pp. 13631369.

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.

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

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