

minfi methylation pipeline

Mikhail Dozmorov

2021-04-12

minfi - Analyze Illumina Infinium DNA methylation arrays

- Reads Illumina's 450k array raw data (IDAT files) into R
- Performs QC and normalization
- Identifies differential methylation positions (DMP)

```
source("https://bioconductor.org/biocLite.R")  
biocLite("minfi")  
biocLite("minfiData")
```

```
library(minfi)  
library(minfiData)
```

<https://bioconductor.org/packages/release/bioc/html/minfi.html>

Methylation data

```
baseDir <- system.file("extdata", package = "minfiData")  
list.files(baseDir)
```

```
## [1] "5723646052"      "5723646053"      "SampleSheet.csv"
```

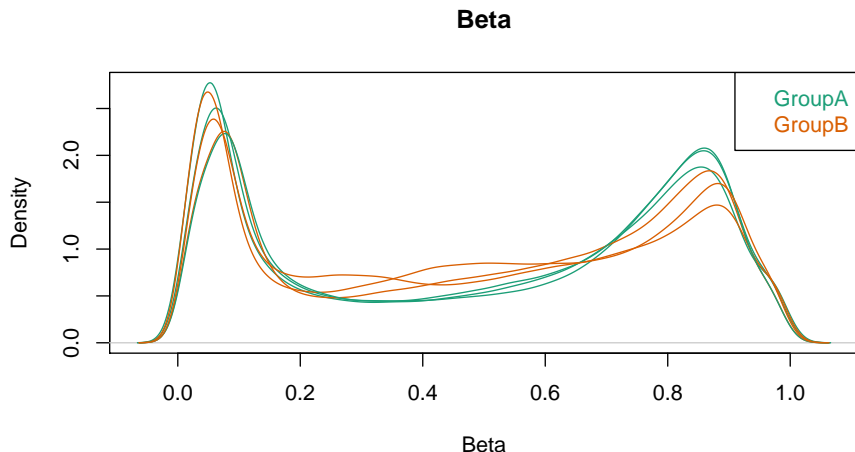
```
targets <- read.metharray.sheet(baseDir)
```

```
## [1] "/Users/mdozmorov/Library/R/4.0/library/minfiData/extda
```

```
RGset <- read.metharray.exp(targets = targets)  
pd <- pData(RGset) ## phenotypic data
```

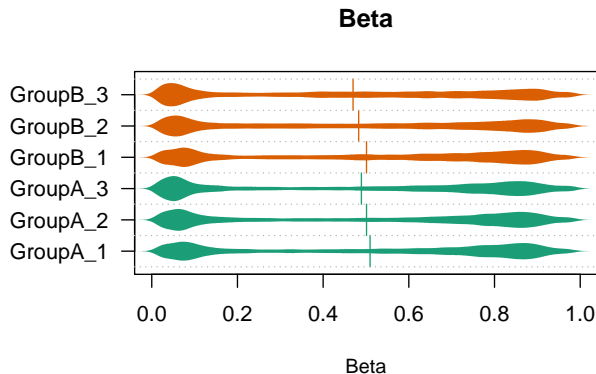
QC: Beta values are expected to cluster around 0/1.

```
densityPlot(RGset, sampGroups = pd$Sample_Group, main = "Beta")
```



QC: Beta values are expected to cluster around 0/1.

```
par(oma=c(2,10,1,1))  
densityBeanPlot(RGset, sampGroups = pd$Sample_Group, sampNames = ...)
```



Normalization

Different methods for normalization have been proposed and still being developed

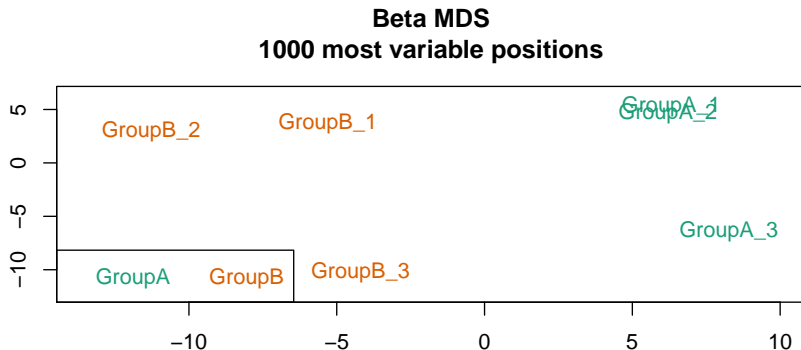
- Dye-bias adjustment
- Probe type I and II adjustment

```
MSet.norm <- preprocessIllumina(RGset, bg.correct = TRUE,  
                                normalize = "controls",  
                                reference = 2)
```

Yousefi P. et. al. "Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies" Epigenetics 2013 <http://www.tandfonline.com/doi/abs/10.4161/epi.26037>

Multi-dimensional scaling (MDS) plot

```
mdsPlot(MSet.norm, numPositions = 1000, sampGroups = pd$Sample
```



Similar to PCA, useful to identify outlier samples.

Getting M-values

```
# A small subset to speed up the demo:  
mset <- MSet.norm[1:20000,]  
# Getting the M values:  
M <- getM(mset, type = "beta", betaThreshold = 0.001)
```

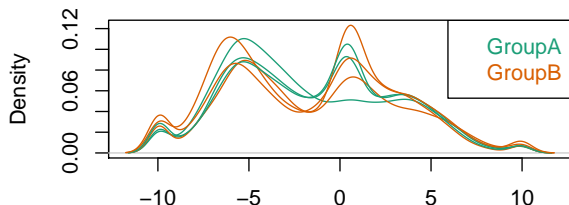

QC: M values should show the level of methylation centered around 0

Beta values ≤ 0.001 , or ≥ 0.999 are truncated to avoid numerical issues.

Look at the density distribution

```
par(oma=c(2,10,1,1))
```

```
densityPlot(M, sampGroups = pd$Sample_Group, sampNames = pd$Sa
```



Differentially methylated positions

```
dmp <- dmpFinder(M, pheno=pd$Sample_Group, type="categorical")  
head(dmp)
```

##		intercept	f	pval	qval
##	cg10805483	-9.964341	1706.1212	2.053224e-06	0.02639720
##	cg20386875	-5.434480	1445.1107	2.859882e-06	0.02639720
##	cg07155336	-5.799521	550.9746	1.952772e-05	0.05148498
##	cg13059719	-2.505878	549.6611	1.962059e-05	0.05148498
##	cg08343042	-3.565042	506.2230	2.310839e-05	0.05148498
##	cg23098069	1.532107	497.6219	2.390872e-05	0.05148498

Rows ordered by p-value.

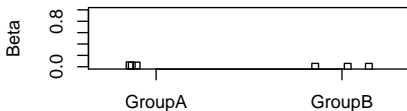
Plotting methylation levels

```
cpgs <- rownames(dmp)[1:4]
par(mfrow=c(2,2))
plotCpg(mset, cpg=cpgs, pheno=pd$Sample_Group)
```

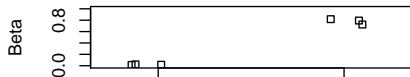
cg10805483



cg20386875



cg07155336



cg13059719

