Mouse Infinium methylation data processing with RnBeads

Seung June Kim

Introduction

The goal is to display some key data frame structures and visuals that you can expect to get from running the code in RnBeads_with_RepeatMasker.R

Using RnBeads dev version (Müller et al. 2019).

Setup[1][2]

What should RepeatMasker table and $SAMPLE.csv file look like? `sample.table` was created solely for demonstration purposes; it's not needed to run the code.

```
head(tab.rmsk)
```

```
head(sample.table,14)
```

Interesting but confusing functions

`get.table()` returns a useful, unsorted data frame of statistics for each $ANNOTATION probe(rows) from `diffmeth.\$ANNOTATION` S4 object. `annotation()` returns row-matching annotation for each probe in tab.$ANNOTATION.

```
diffmeth.rmsk <- rnb.execute.computeDiffMeth(rnb.set, cmp.cols, region.types="rmsk")
tab.rmsk <- get.table(diffmeth.rmsk,comparison, region.type="rmsk",return.data.frame=T)
aa.rmsk <- annotation(rnb.set,type="rmsk")
```

```
head(tab.rmsk)
```

```
head(aa.rmsk)
```

These two data frames are bound together.

```
annotated.tab.rmsk <- data.frame(tab.rmsk,aa.rmsk, row.names=NULL)
head(annotated.tab.rmsk)
```

Order the rows by "combinedRank". The rownames of the `annotated.tab.\$ANNOTATION.order` data frame refers to the original row numbers of the annotated.tab.sites data frame, which is not ordered.

[1] Run rmarkdown::render("RnBeads_with_RepeatMasker.Rmd") in the same workspace as .R to avoid worrying about global environment.

[2] rnb.set object is not stored well in the global environment. If there are errors, re-run the script from top to bottom (remove output folders "result" and "resultFull"). There is a way to write theses objects to disk instead of RAM, but I didn't want to lose any disk space.

```
annotated.tab.rmsk.order <- annotated.tab.rmsk[order(annotated.tab.rmsk[,"combinedRank"]),]
head(annotated.tab.rmsk.order,3)
```

```
rownames(annotated.tab.rmsk.order)[1:10]
```

```
##  [1] "12842" "2839"  "22365" "24294" "6831"  "20269" "17342"
##  [8] "17343" "7604"  "8467"
```

`annotated.tab.\$ANNOTATION.order` data frame does not contain any information about beta values. These are extracted from rnb.set using `meth()`, annotation name, and row index from `annotated.tab.\$ANNOTATION.order`.

```
topDiffBeta.rmsk.full <- meth(rnb.set.full, type="rmsk")[as.integer(rownames(annotated.tab.rmsk.order)),]
```

Heatmap(s)

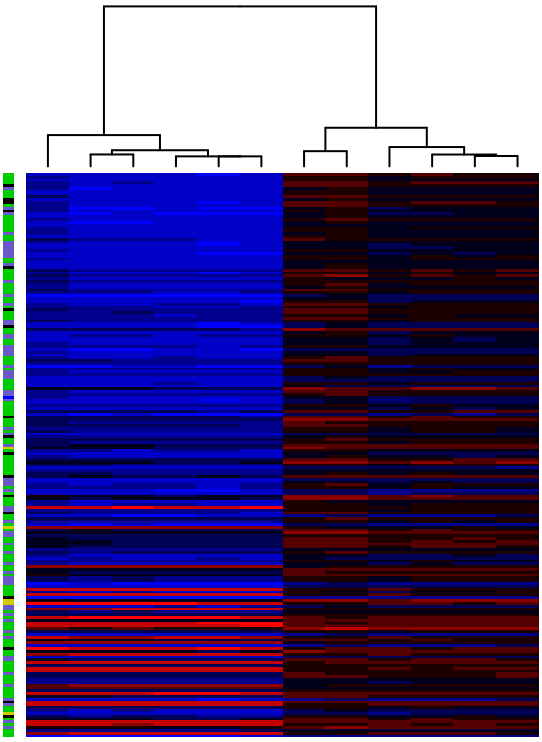
Differentially methylated probes annotated with RepeatMasker

```
colSide <- annotated.tab.rmsk.order$name

colSide <-gsub("LTR","slateblue3",colSide)
colSide <- gsub("LINE","green3",colSide)
colSide <- gsub("SINE","green3",colSide)
colSide <- gsub("DNA","black",colSide)
colSide <- gsub("Simple_repeat","yellow3",colSide)
colSide <- gsub("Other","blue",colSide)
colSide <- gsub("Unknown","blue",colSide)

par(font=2, font.axis=2, font.lab=2,cex.lab=1.5, cex.axis=1.5, lty=1)
heatmap.2(topDiffBeta.rmsk.full[1:200,c(1,5,9,2,6,10,3,7,11,4,8,12)],
          col=colfunc(10), scale="none", Rowv=F, Colv=T,
          cexCol=1, labCol=NA, labRow=NA,srtCol=0, adjCol=c(0.5,0),
          density.info = "none",trace="none", dendrogram = "column",
          symkey=FALSE,symbreaks=F,revC = FALSE, RowSideColors = colSide[1:200],
          breaks=c(0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1),
          lmat=rbind(c(5,0,4),c(3,1,2)),
          lhei=c(1,4),
          lwid=c(1.75,0.1, 3.5),margins=c(5,12),cexRow=1.2)
```

```
## Error in plot.new(): figure margins too large
```



Müller, Fabian, Michael Scherer, Yassen Assenov, Pavlo Lutsik, Jörn Walter, Thomas Lengauer, and Christoph Bock. 2019. "RnBeads 2.0: Comprehensive Analysis of DNA Methylation Data." *Genome Biology* 20 (1): 55. https://doi.org/10.1186/s13059-019-1664-9.