Differential methylation analysis using MetKMR

Ruth Barral Arca , Msc in Omic Data Analysis

24 April 2019

Here we will use data of Hungtinton disease (HD) patients and controls fibroblast cells from the study (GSE67733): *Jia H, Morris CD, Williams RM, Loring JF et al. HDAC inhibition imparts beneficial transgenerational effects in Huntington's disease mice via altered DNA and histone methylation. Proc Natl Acad Sci U S A 2015 Jan 6;112(1):E56-64. PMID: 25535382*

Here we will study methilation patterns of patients with Hungtion Disease focusing on the chromosome 4 where the gen responsible of the disease HTT is located using kernel regression .

In this example the outcome variable it's dicotomous : healthy control (normal) vs Hungtinton (HD). But please note that MetKMR also allow the user to study continuous outcomes such as the weight or the gene expression level

```
#set the enviroment
library(minfi)
library(MetKMR)
library(IlluminaHumanMethylation450kanno.ilmn12.hg19)
library(doParallel)
library(rtracklayer)
library(biomaRt)
library(dbplyr)
registerDoParallel(cores = 4)
setwd("/home/ruth/Dropbox/TFM_RUTH/first_approach")

#load the data
phenoData<-c("normal","HD","normal","HD","normal","HD", "normal","HD","normal","HD")
load("betas.rda")
load("annotation2.rda")
#the annotation MUST have this format
head(annotation2)
```

```
##   row      pos       site  chr              gene
## 1   1  9363356 cg00050873 chrY    TSPY4;FAM197Y2
## 2   2 21239348 cg00212031 chrY            TTTY14
## 3   3 15815688 cg00214611 chrY     TMSB4Y;TMSB4Y
## 4   4  6778695 cg01707559 chrY TBL1Y;TBL1Y;TBL1Y
## 5   5 15815552 cg02004872 chrY     TMSB4Y;TMSB4Y
## 6   6  9194502 cg02011394 chrY             TSPY4
```

MetKMR

Assessment of DNA methylation changes in human fibroblasts from normal controls and patients with Huntington's disease (HD)

For illustrating purpouses we choose a window size of 9 and the wmethod default , so we will test only a reduced number of windows and therefore this tutorial script will run fast in the majority of laptops.

For example another interesting option, but more computationally demanding, would be to use the wmethod "location_fixed" and choose wsize of roughly 1000bp and a gap of 500 . As 1089 bp has been described as

the average lenght of human CpG islands by *"Han, L., Su, B., Li, W. H., & Zhao, Z. (2008). CpG island density and its correlations with genomic features in mammalian genomes. Genome biology, 9(5), R79."*

We encourage the user to "play" with the different windows options.

##DMRs Stimation

```
analysis <- new("MetRKAT",
                data = betas,
                annotation = annotation2,
                distmethod =  c("euclidean"),
                wsize = 9, gap = 0,
                max.na = 0.3,wmethod = "default")
analysis <- toSQLite(analysis, "hungtinton.sqlite")
analysis@intervals <- createIntervals(analysis)

y<-replace(phenoData,phenoData=="HD",1)
y<-replace(y,phenoData=="normal",0)
analysis@results <- applyRKAT(analysis, y = y)


results_df <- as.data.frame(analysis@results)
head(results_df)
```

```
##   first_row last_row  start    end  chr         pval    kernel omnibus
## 1         1        8  15865 566570 chr1 2.134266e-05 euclidean      NA
## 2         9       17 566731 663153 chr1 1.681856e-04 euclidean      NA
## 3        18       26 665298 763127 chr1 3.410678e-01 euclidean      NA
## 4        27       35 765632 805541 chr1 7.661683e-06 euclidean      NA
## 5        36       44 805554 853925 chr1 0.000000e+00 euclidean      NA
## 6        45       53 854766 855445 chr1 0.000000e+00 euclidean      NA
```

We keep only those DMRs(differentially methilated regions) that are statistically significant and that are located within the Chromosome 4 (where the HTT is located)

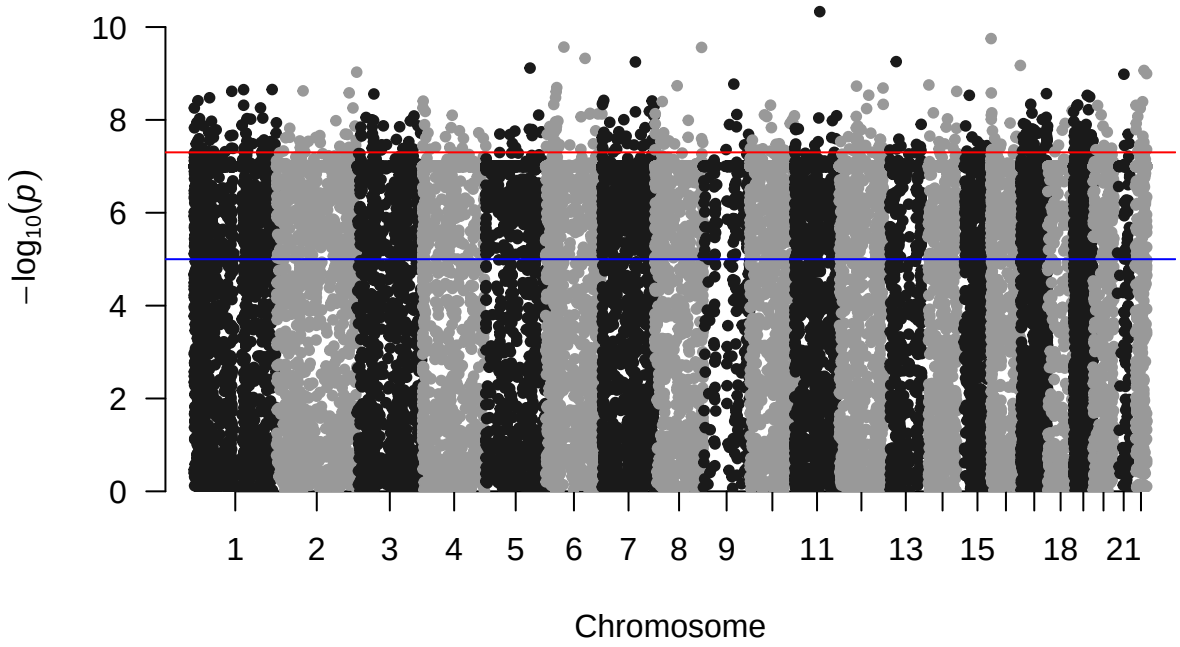
```
#we are interested in chr4 because hungtine gene
results_df <- results_df[results_df$chr == 'chr4', ]
filtered_results <- results_df[results_df$pval<= 0.05, ]
head(filtered_results)
```

```
##       first_row last_row  start    end  chr         pval    kernel omnibus
## 26136    235138   235145  53010  53270 chr4 2.854213e-06 euclidean      NA
## 26137    235146   235154  53362 107725 chr4 0.000000e+00 euclidean      NA
## 26138    235155   235163 107897 125112 chr4 0.000000e+00 euclidean      NA
## 26139    235164   235172 125504 206339 chr4 0.000000e+00 euclidean      NA
## 26140    235173   235181 206442 290582 chr4 0.000000e+00 euclidean      NA
## 26141    235182   235190 291309 331338 chr4 0.000000e+00 euclidean      NA
```

##Plots

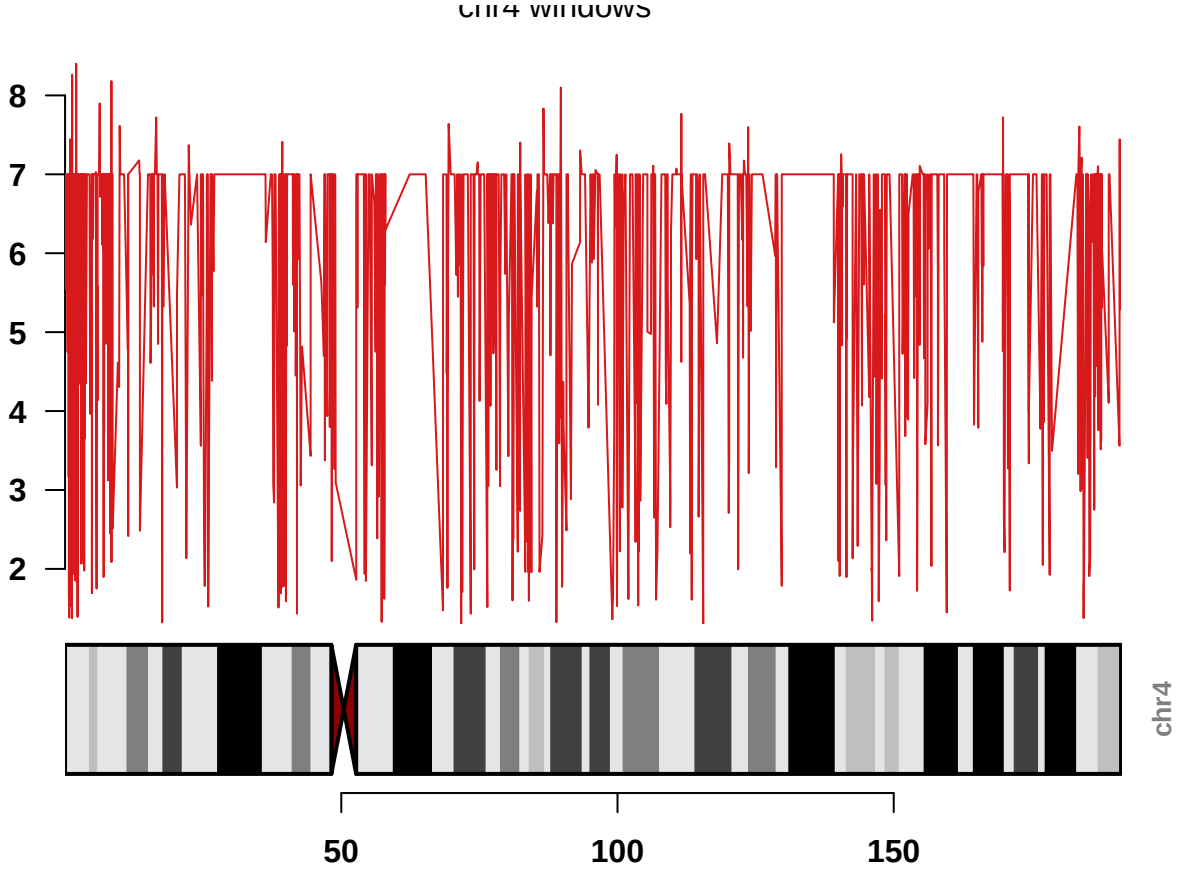
####First we will have a look at all the chromosomes using a Manhattan Plot

```
plotManhattan(analysis, pvals ='euclidean')
```



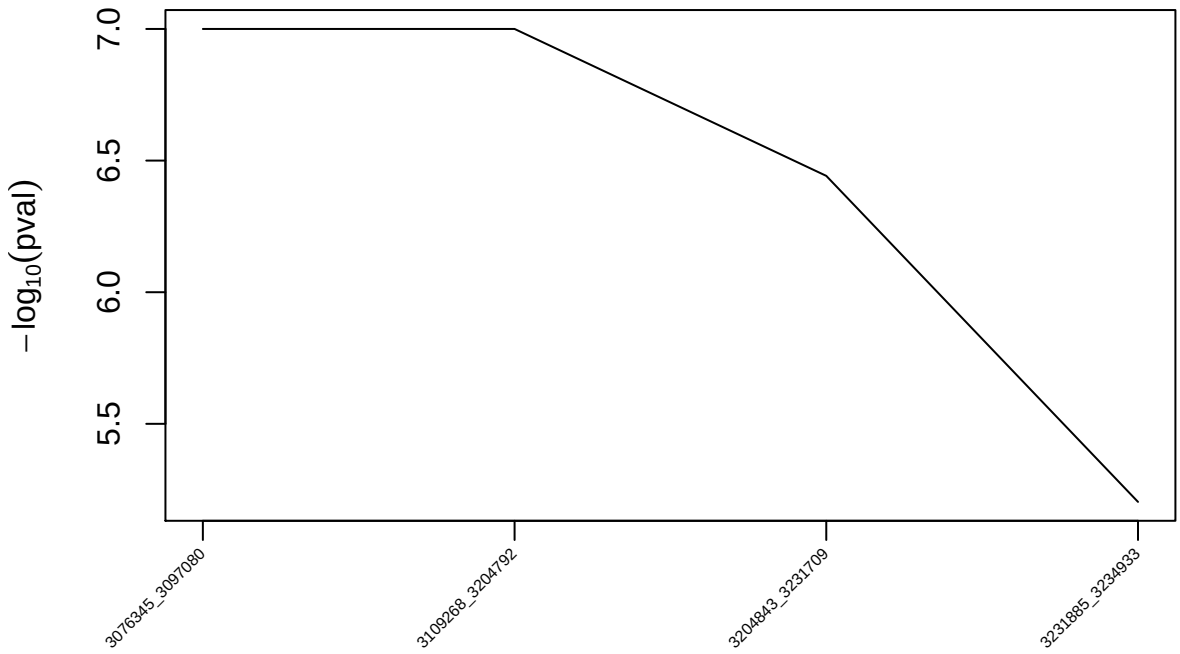
###Then we will focus on Chromosome 4 plotting an ideogram

```
plotChromosome(analysis, chrom = 'chr4', pvals = 'euclidean',cutoff = 0.05)
```



###Finally we will focus on the HTT gene in Chromosome 4 and plot its kernel windows

```
plotWindows(analysis, chrom = 'chr4', pvals = 'euclidean',cutoff = 0.05,
            startpos = 3041422,endpos = 3243960)
```



We take all the significant windows and convert them into genes using Biomart

```
# Map positions to genes
analysis@annotation<-as.data.frame(analysis@annotation)
DMG_symbol<-analysis@annotation[filtered_results$first_row, 'gene']
```

##Final results

We select those DMRs within the HTT gene between the all the differentially expressed DMRs

```
#Hungtin
annotated_result<-filtered_results
annotated_result$gene<-DMG_symbol
head(annotated_result)
```

```
##       first_row last_row  start    end  chr           pval    kernel
## 26136    235138   235145  53010  53270 chr4 0.000002854213 euclidean
## 26137    235146   235154  53362 107725 chr4 0.000000000000 euclidean
## 26138    235155   235163 107897 125112 chr4 0.000000000000 euclidean
## 26139    235164   235172 125504 206339 chr4 0.000000000000 euclidean
## 26140    235173   235181 206442 290582 chr4 0.000000000000 euclidean
## 26141    235182   235190 291309 331338 chr4 0.000000000000 euclidean
##       omnibus                        gene
## 26136      NA                ZNF595;ZNF718
## 26137      NA ZNF718;ZNF718;ZNF595;ZNF595
## 26138      NA                      ZNF718
## 26139      NA                      ZNF718
## 26140      NA                     ZNF876P
## 26141      NA                      ZNF732
```

```
HTT<-annotated_result[grep("HTT",annotated_result$gene),];HTT
```

```
##       first_row last_row   start     end  chr            pval    kernel
## 26336    236937   236945 3076345 3097080 chr4 0.0000000000000 euclidean
## 26337    236946   236954 3109268 3204792 chr4 0.0000000000000 euclidean
## 26338    236955   236963 3204843 3231709 chr4 0.0000003614419 euclidean
## 26339    236964   236972 3231885 3234933 chr4 0.0000062579278 euclidean
## 26340    236973   236981 3235000 3249975 chr4 0.0000000000000 euclidean
##       omnibus gene
## 26336      NA  HTT
## 26337      NA  HTT
## 26338      NA  HTT
## 26339      NA  HTT
## 26340      NA  HTT
```

```
HTT
```

```
##       first_row last_row   start     end  chr            pval    kernel
## 26336    236937   236945 3076345 3097080 chr4 0.0000000000000 euclidean
## 26337    236946   236954 3109268 3204792 chr4 0.0000000000000 euclidean
## 26338    236955   236963 3204843 3231709 chr4 0.0000003614419 euclidean
## 26339    236964   236972 3231885 3234933 chr4 0.0000062579278 euclidean
## 26340    236973   236981 3235000 3249975 chr4 0.0000000000000 euclidean
##       omnibus gene
## 26336      NA  HTT
## 26337      NA  HTT
## 26338      NA  HTT
## 26339      NA  HTT
## 26340      NA  HTT
```

Result: We observed differentially methylation patterns of the DMRs within the HTT gene between cases and controls. Which suggest that despite of being a mendelian dominant disease the methylation modifications also play a rol in hungtinton disease.