Pipeline for methylation assay

Stefano Roncelli

14/06/20

Student N° 26 Address 10633381 p-value threshold: 0.01 normalization preprocessSWAN Mann_whitney test

Load raw data with minfi and create an object called RGset storing the RGChannelSet object

```
setwd('.')
suppressMessages(library(minfi))
baseDir <- ('./Input_data')
targets <- read.metharray.sheet(baseDir)
```

```
## [1] "./Input_data/Samplesheet_report_2020.csv"
```

```
RGset <- read.metharray.exp(targets = targets)
save(RGset, file = "RGset.RData")
RGset
```

```
## class: RGChannelSet
## dim: 622399 8
## metadata(0):
## assays(2): Green Red
## rownames(622399): 10600313 10600322 ... 74810490 74810492
## rowData names(0):
## colnames(8): 5775278051_R01C01 5775278051_R04C02 ... 5930514035_R04C02
##   5930514035_R06C02
## colData names(7): Sample_Name Group ... Basename filenames
## Annotation
##   array: IlluminaHumanMethylation450k
##   annotation: ilmn12.hg19
```

Create the dataframes Red and Green to store the red and green fluorescences respectively

```
Red <- data.frame(getRed(RGset))
Green <- data.frame(getGreen(RGset))
```

Fill the following table: what are the Red and Green fluorescences for the address assigned to you? Optional: check in the manifest file if the address corresponds to a Type I or a Type II probe and, in case of Type I probe, report its color.

```
probes_I <- getProbeInfo(RGset, type = 'I')
probes_II <- getProbeInfo(RGset, type = 'II')
probes_I[probes_I$AddressA == 10633381,]
```

```
## DataFrame with 0 rows and 8 columns
```

```
probes_I[probes_I$AddressB == 10633381,]
```

```
## DataFrame with 1 row and 8 columns
##          Name    AddressA    AddressB       Color       NextBase
```

```
##   <character> <character> <character> <character> <DNAStringSet>
## 1  cg03868159    21656441    10633381         Red               A
##                  ProbeSeqA               ProbeSeqB      nCpG
##             <DNAStringSet>          <DNAStringSet> <integer>
## 1 CTAAACATCC...AACTATACCA CTAAACGTCC...AACTATACCG         2
```

```
probes_II[probes_II$AddressA == 10633381,]
```

```
## DataFrame with 0 rows and 4 columns
```

```
Red[rownames(Red) == '10633381',]
```

```
##          X5775278051_R01C01 X5775278051_R04C02 X5775278078_R02C01
## 10633381               1852               1694               1354
##          X5775278078_R05C01 X5775278078_R05C02 X5930514034_R01C02
## 10633381               1091               1131                796
##          X5930514035_R04C02 X5930514035_R06C02
## 10633381                894               1149
```

```
Green[rownames(Green) == '10633381',]
```

```
##          X5775278051_R01C01 X5775278051_R04C02 X5775278078_R02C01
## 10633381                458                631                358
##          X5775278078_R05C01 X5775278078_R05C02 X5930514034_R01C02
## 10633381                396                424                302
##          X5930514035_R04C02 X5930514035_R06C02
## 10633381                354                479
```

We can see it's a type I infinium with the Red channel.

Sample	Row	Column	Red Intensity	Green Intensity	Type	Color
5775278051	1	1	1852	458	I	Red
5775278051	4	2	1694	631	I	Red
5775278078	2	1	1354	358	I	Red
5775278078	5	1	1091	396	I	Red
5775278078	5	2	1131	424	I	Red
5930514034	1	2	796	302	I	Red
5930514035	4	2	894	354	I	Red
5930514035	6	2	1149	479	I	Red

Step 4 Create the object MSet.raw

Step 5 Perform the following quality checks and provide a brief comment to each step: - QCplot - check the intensity of negative controls using minfi - calculate detection pValues; for each sample, how many probes have a detection p-value higher than the threshold assigned to each student? Step 6 Calculate raw beta and M values and plot the densities of mean methylation values, dividing the samples in DS and WT (suggestion: subset the beta and M values matrixes in order to retain DS or WT subjects and apply the function mean to the 2 subsets).

Step 7 Normalize the data using the function assigned to each student and compare raw data and normalized data. Produce a plot with 6 panels in which, for both raw and normalized data, you show the density plots of beta mean values according to the chemistry of the probes, the density plot of beta standard deviation values according to the chemistry of the probes and the boxplot of beta values. Provide a short comment regarding the changes you observe.

Step 8 Perform a PCA on the beta matrix generated in step 7. Comment the plot.

Step 9 Using the matrix of normalized beta values generated in step 7, identify differentially methylated probes between group DS and group WT using the functions assigned to each student. Note; it can take several minutes; if you encounter any problem you can run the differential methylated analysis only on a subset of probes (for example those on chromosome 1, 18 and 21)

Step 10 Apply multiple test correction and set a significant threshold of 0.05. How many probes do you identify as differentially methylated considering nominal pValues? How many after Bonferroni correction? How many after BH correction?

Step 11 Produce an heatmap of the top 100 differentially mehtylated probes

Step 12 Produce a volcano plot and a Manhattan plot of the results of differential methylation analysis

Step 13 (optional) As DS is caused by the trisomy of chromosome 21, try also to plot the density of the methylation values of the probes mapping on chromosome 21. Do you see a very clear difference between the samples? How many differentially methylated probes do you find on chromosome 21?