BINF-F402, Assignment 1: DNA methylation analysis

Matthieu Defrance

November 2018

Practical information

Your project (that accounts for the global evaluation) should be sent to matthieu.dc.defrance@ulb.ac.be as a PDF document containing the detailed results and figures. The document should also include a significant part of the R code that has been used to perform the analysis. The overall clarity of the document and the code will be assessed. The deadline is fixed to the **21th November**.

Objective

The objective of this project is to study the DNA methylation in the context of a simple experiment were a simultaneous KO of DNMT1 and DNMT3B has been performed on a human cell line. The DNA methylation profiling of three biological replicates in each condition (the control and the Double KO) has been performed using the Illumina Infinium methylation platform.

Data acquisition

From `Assignment-1` directory (https://tinyurl.com/y9xdtcwc) download the following file:

- `Signal_AB.txt.gz` containing the raw values obtained with the Infinium methylation technology. This file containts the Signal_A (U) and Signal_B (M) for each sample. The first three samples (WT1, WT2, WT3) correspond to control data and last three samples (KO1, KO2, KO3) to the KO data.
- `annotation.csv.gz` containing a simplified genomic annotation for the Infinium data.

ß-value distribution

In this first part, you will have to compute and analyze the ß-value distribution. Perform the following sequence of analysis:

1. Compute the ß-value using the Signal_A and Signal_B data (Signal_B corresponds to the Methylated signal (M) and Signal_A to the Unmethylated signal (U)).
2. Compute and plot the ß-value distribution for each sample independently. Describe the obtained results.

Global overview

In order to better highlight the local methylation level changes, it is useful to relate the probes to their genomic location and the associated genes.

3. Using the annotation provided in `annotation.csv.gz` compute the methylation level distribution for chromosome 6, 7 and 21 in both condition and plot it using a boxplot.

4. Compute the average methylation for the control and the case for the follwing genes: `MED14`, `CLK2` and `TLR5` and plot it using a barplot. How many probes are related to gene `CLK2`?

Differential analysis

In this third step, you will perform a differential methylation analysis.

5. Compute the delta-ß (ß-value difference between case mean and control mean) for all the probes. Using the t-test (t.test) compute the Pvalue of the differential methylation (case vs control). Plot and discuss the distribution of the delta-ß.
6. Draw a volcano plot with the delta-ß on the x-axis and the -log10(Pvalue) on the y-axis.

Random control (Bonus)

7. Perform a sequence of random controls by shuffling the data and redo the differential analysis. What are your conclusions?