

## **Project 2**

The data set (data\_project\_2.csv) is based on the study of de Vega et al (2017). This study aimed at identifying CpG probes that are differentially methylated between patients with Chronic Fatigue Syndrome (CFS) and healthy controls (HC). It concerns data from CpG probes located in the coding regions of ACE and ACE2 genes. The data dictionary is the following:

- Group – CFS versus HC
- cg05748796, ..., cg21094739 are the beta values for each probe evaluated

Before conducting any formal analysis, present and describe the data of each probe with appropriate statistics and plots. Derive the corresponding M values for each probe and conduct the same descriptive statistics. Do you think the normal distribution describes M values of each probe well? Use statistical tests to answer this question.

Perform different associations analyses using beta values and M values alternatively. That is, find which CpG probes that would appear to be differentially methylated between the two groups. Don't forget to apply an appropriate multiple testing correction to each analysis and justifying it is used. Which of your analyses is the most reliable statistically speaking?

Construct a diagnostic model for CFS using the M values. Evaluate the overall performance of your diagnostic model by calculating the ROC curve and AUC. Decide the optimal sensitivity and specificity for your diagnostic model. Justify your decision. Do you have evidence that this is a good model to be used in CFS diagnosis?

### **Important:**

Prepare a 15-minute presentation with your main findings. There will be a penalty of 0.5 points in your project grade if you exceed the time of your presentation. Upload your R script/R Markdown for code verification. Also upload your presentation as a pdf file. Failure to upload these files before classroom evaluation leads to a penalty of 0.5 points in your project grade.

### **Reference:**

de Vega WC, Herrera S, Vernon SD, McGowan PO. Epigenetic modifications and glucocorticoid sensitivity in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). BMC Med Genomics. 2017;10(1):11. doi:10.1186/s12920-017-0248-3