[francesco.ravaioli2@unibo.it](mailto:francesco.ravaioli2@unibo.it)

Instructions:

* For each step of the analytical pipeline provide the R code you used
* In some steps, group specific instructions are given; refer to the tables below for assignments

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Members | | | | | | | |
| 1 | Amir Aynede | Kian Keshani | Bianca Gustini | Yuri Dobrokhotov | Priscilla Castro Vargas | Favour Osese Otoijamun |  |  |
| 2 | Sareh Salehi | Azam Bakhshandeh | Shaghayegh Shirani | delnia khezragha | Maryam Kazemi | Milad Arvand |  |  |
| 3 | Andrea Arriola Gamboa | Kristina Djikic | Nina Talajic | Rebecca Barbera | Aniello Di Vaio | Domenico Zianni |  |  |
| 4 | Andrea Pusiol | Aurora Mazzoni | Martina Castellucci | Alessia Corica | Sofia Natale | Bianca Mastroddi | Perla Lucaboni |  |
| 5 | Luca Cagnini | Marco Cuscunà | Marco Centenaro | Marina Mariano | Massimo Lanari | Michele Carbonieri |  |  |
| 6 | Nikita Leino | Vanessa El Debs |  |  |  |  |  |  |
| 7 | Negin Nillforoosh | Mahan balooei | Elif Güler | Mustafa Barkın Kemeç | Simay Erol | Eyip Sinay Dalmaz | Kimia Kanouni |  |
| 8 | Mohsin | Jingyu Zeng | Rajab | Yusra | Hasan | Junlong Chen | Jia Guangming |  |
| 9 | Deniz Ertuğrul | Başak Akkoyun | Kağan Sağlam | Nadir Mammadov | Betül Yalçın |  |  |  |
| 10 | Anna Rossi | Martina Marotta | Teresa Gianni | Giacomo Timelli | Andrea Lenti | Valerio Piersanti | Enrico Gallus | Amedeo Antoci |

**Step 3**

|  |  |
| --- | --- |
| Group ID | Address |
| 1 | 18756452 |
| 2 | 71773431 |
| 3 | 31763489 |
| 4 | 44666390 |
| 5 | 13673406 |
| 6 | 71636344 |
| 7 | 40796508 |
| 8 | 13673406 |
| 9 | 71636344 |
| 10 | 39802405 |

**Step 5**

|  |  |
| --- | --- |
| Group ID | detPvalue threshold |
| 1 | 0.05 |
| 2 | 0.01 |
| 3 | 0.05 |
| 4 | 0.01 |
| 5 | 0.05 |
| 6 | 0.01 |
| 7 | 0.05 |
| 8 | 0.01 |
| 9 | 0.05 |
| 10 | 0.01 |

**Step 7**

|  |  |
| --- | --- |
| Group ID | Normalization |
| 1 | preprocessNoob |
| 2 | preprocessSWAN |
| 3 | preprocessQuantile |
| 4 | preprocessFunnorm |
| 5 | preprocessNoob |
| 6 | preprocessSWAN |
| 7 | preprocessQuantile |
| 8 | preprocessFunnorm |
| 9 | preprocessNoob |
| 10 | preprocessSWAN |

**Pipeline steps**

1. Load raw data with minfi and create an object called RGset storing the RGChannelSet object
2. Create the dataframes Red and Green to store the red and green fluorescences respectively
3. Fill the following table: what are the Red and Green fluorescences for the address assigned to your group? **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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample** | **Red fluor** | **Green fluor** | **Type** | **Color** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

1. Create the object MSet.raw
2. 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 group?

|  |  |
| --- | --- |
| **Sample** | **n° Failed positions** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

1. Calculate raw beta and M values and plot the densities of mean methylation values, dividing the samples in WT and MUT (*suggestion: subset the beta and M values matrixes in order to retain WT or MUT subjects and apply the function mean to the 2 subsets*). Do you see any difference between the two groups?
2. Normalize the data using the function assigned to each group 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 about the changes you observe. *Optional: do you think that the normalization approach that you used is appropriate considering this specific dataset? Try to color the boxplots according to the group (WT and MUT) and check whether the distribution of methylation values is different between the two groups, before and after normalization …*
3. Perform a PCA on the matrix of normalized beta values generated in step 7, after normalization*.* Comment the plot (Do you see any outlier? Do the samples divide according to the group? Do they divide according to the sex of the samples? Do they divide according to the batch, that is the column Sentrix\_ID?).