

Group work 3 - part a

The file [Genus_otu_table.txt](#) is a tab delimited file that you can download from moodle. It contains raw count data of 16S sequencing of fecal samples from healthy donors and patients affected by *Clostridium difficile*, summarized at genus level. You are asked to analyze the data.

Exercise 1)

Choose which type(s) of data transformation (clr, alr), normalization(GMPR,CSS,TMM), imputation (mbimpute) you want to perform

NOTE (these aspects will be evaluated):

- You have to comment your code (I need to understand what you are doing)
- You have to explain your choices. For example, you might choose to alr transform and impute, but not normalize the data and you have to explain the reasons of your choices (in case include plots in your code)
- Submit your code as an Rmarkdown (.rmd) file

Group work 3 - part b

The file [Genus_otu_table.txt](#) is a tab delimited file that you can download from moodle. It contains raw count data of 16S sequencing of fecal samples from healthy donors and patients affected by *Clostridium difficile*, summarized at genus level. You are asked to analyze the data.

Exercise 2)

Run the DA analysis using Aldex and Ancom (as seen in the handson) and compare the results

NOTE (these aspects will be evaluated):

- You have to comment your code (I need to understand what you are doing)
- You have to explain your choices. For example, you might choose to alr transform and impute, but not normalize the data and you have to explain the reasons of your choices (in case include plots in your code)
- Submit your code as an Rmarkdown (.rmd) file

Group work 3 – part c

The file [Genus_otu_table.txt](#) is a tab delimited file that you can download from moodle. It contains raw count data of 16S sequencing of fecal samples from healthy donors and patients affected by *Clostridium difficile*, summarized at genus level. You are asked to analyze the data.

Exercise 3)

Based on the results obtained in exercise 2, select a set of DA taxa and, based only on these taxa, run NMDS, t-sne and UMAP to project the healthy donors and the patients on 2 dimensions. Are these plots equal to the ones you obtain using all taxa instead of using only the DA?

NOTE (these aspects will be evaluated):

- You have to comment your code (I need to understand what you are doing)
- You have to explain your choices. For example, you might choose to `alr` transform and impute, but not normalize the data and you have to explain the reasons of your choices (in case include plots in your code)
- Submit your code as an Rmarkdown (.rmd) file

I will test your solution

Only one student from each group has to submit the code on behalf of his/her mates as representative of the group.

I will test your code with different inputs... this is not a programming course so **a valid submission means that your code must be submitted fully working**... I am not going to correct coding and programming bugs...

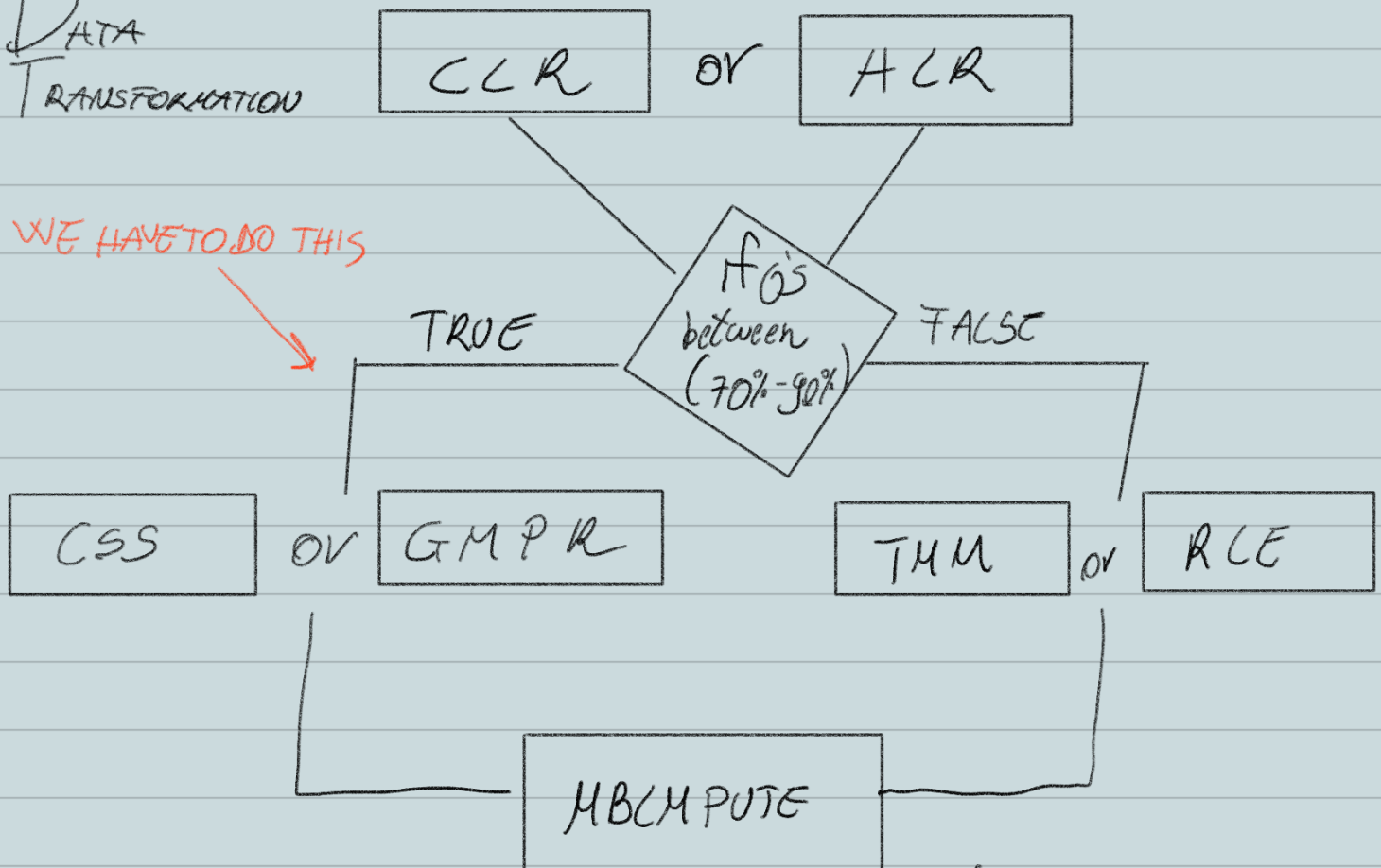
Therefore, before submitting it, try running your code placing the input files in a directory different from the working directory and cleaning the workspace before running your functions.

To assign a grade I will evaluate if the analysis is correct and give appropriate and reasonable results, and if the code is well commented and your choices well motivated

PART A

First thing that we are going to perform is the Juba Transform and the count of overall zeros, to have a better understanding of how many 0's we have to take in account. She wrote that the general number is around 70% - 90%.

DATA TRANSFORMATION




Some one have implemented it?

PART B

Implemented have just to use

<https://bioconductor.org/packages/release/bioc/html/ALDEx2.html>

with examples



[http://www.bioconductor.org/packages/release/bioc/
vignettes/ANCOMBC/inst/doc/ANCOMBC.html](http://www.bioconductor.org/packages/release/bioc/vignettes/ANCOMBC/inst/doc/ANCOMBC.html)