**ALPHA BETA diversity results**

This folder shows the results of the Alpha/Beta analysis. In this folders there is the fractional microbial abundances, for each combination of cohort/treatment and the Alpha/Beta diversity computed. For the former, we calculate, for each combination, the fraction abundances grouped per Class/Family and Phylum, while in the latter, Alpha-Beta diversity calculation, we can distinguish 3 **different type of analysis**

1. *Cohort\_based\_Alpha\_Beta\_Diversity*; Here, I just divide between **responder and non-responders** in a particular cohort. Based on that I calculated Alpha and Beta diversity.
2. *TNF\_selected\_microbes\_per\_list;* For patients receiving the TNF treatment, we focused on the microbes that are part of the selected genera (called ***Selected***) and the ones that are not part of the list (called ***Non selected***), known to have a relation with TNF, and the microbes not part of it. Hence in those folders, we created two different groups.

*Fusicatenibacter*

*Roseburia*

*Dorea*

*Blautia*

*Intestinimonas*

*Adlercreutzia*

*Clostridium\_IV*

*Parabacteroides*

*Collinsella*

*Anaerostipes*

*Subdoligranulum*

*Butyricicoccus*

*Oscillibacter*

*Akkermansia*

*Alistipes*

*Ruminococcus2*

*Ruminococcus*

*Veillonella*

*Barnesiella*

*Bifidobacterium*

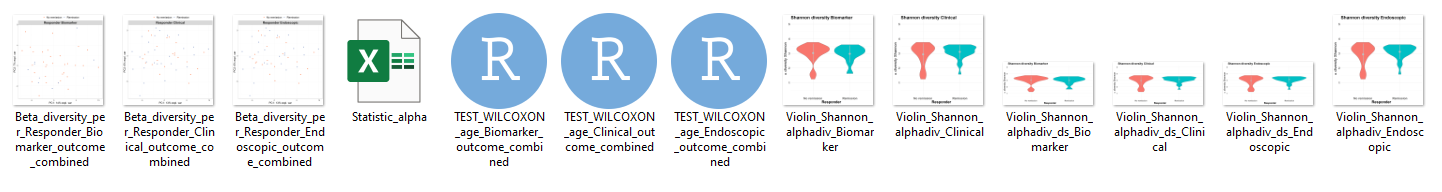
For each cohort, we did two sets of analysis; a) The folder with suffix **un\_name**, where the microbes **NOT mapped** to any genera **ARE present** in the group of **“*Non selected*”** microbes. B) the folder with suffix \_**canc**, the unnamed microbes (the microbes not mapped to any genera) are **OMITTED.**

1. *TNF\_selected\_microbes\_per\_list\_responders\_vs\_non\_responders* ; FOLDERS with **\_divider\_resonder** at the end of the folder

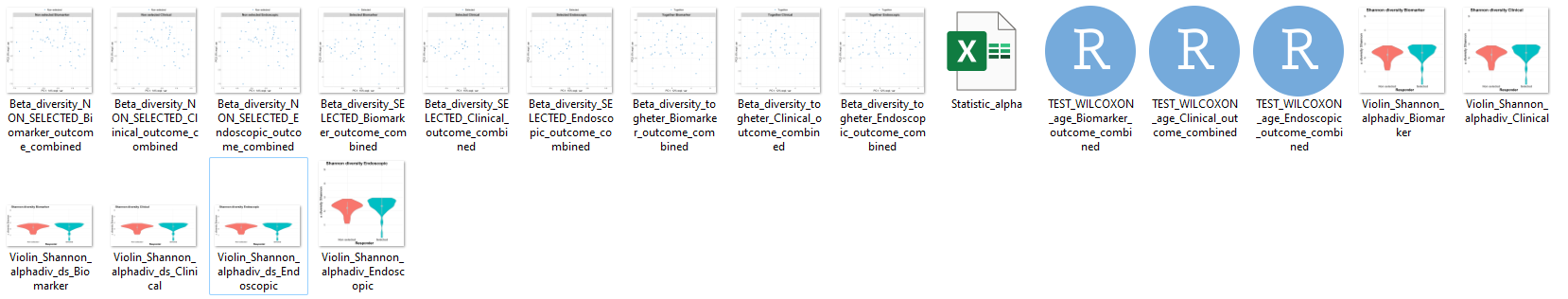
In those folders, I have combined the ***Remission*** vs ***Non remission*** and the ***Selected*** vs ***Non selected*** microbes.

In particular, FOCUSING only on i.e., ***Selected*** microbes, I have then divided again the individual into ***Remission* vs *Non remission*** and computed the alpha and beta diversity segregating (subset) to the **SELECTED** microbes. AS for (2), here also there are 2 folders, one with OMITTED (**canc**) and one with the UNNAMED microbes (**no\_canc\_un\_name**)

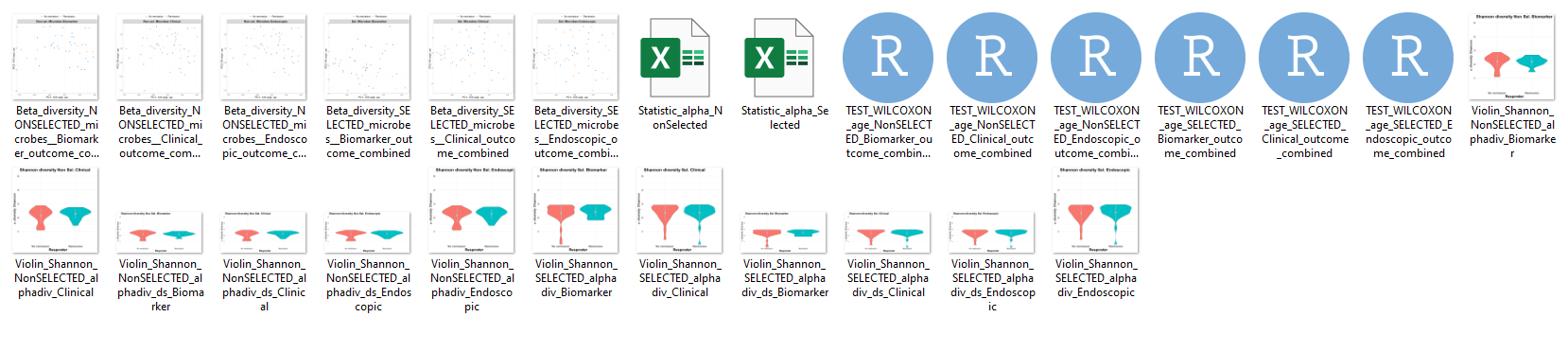
Structure of the folders, with an example in each folder:

1. 

* The first 3 files are the Beta diversity with different type of Outcome (Clinical; Biomarker, Endoscopic=.
* Then we have the statistic alpha, in which statistics of the violin plot for alpha diversity are (in different sheets) reported (i.e., median, IQR, … )
* Then the .rds files are, for the different type of outcomes, if there is a statistical significant difference between ***Remission*** and ***Non remission*** per the alpha diversity.
* THEN we have the ALPHA diversity plots
  + Each plot (for each outcome) is repeated TWICE, in different shapes, to choose the more adapt for the manuscript.
* N.b., individual that have missing value for a certain outcome are omitted for the calculation in that specific outcome.

1. 

* The first 9 plots are the BETA diversity.
  + N.b., all the analysis here are done once per every **Outcome**, but there is no need, since we do not use the outcome, hence they are exactly the same
  + There are 3 categories: ***Non selected***, i.e., BETA diversity computed only on the non-selected microbes, ***Selected***, i.e., BETA diversity computed only on the ***Selected*** one, and Together on the full microbes.
  + Of course this is less informative than in cases 1 and 3
* Statistic alpha as before,
* Test WILCOXON as before, on the alpha diversity, but here it is *paired* test and the two groups are the same individuals but with ***Selected*** vs ***Non selected*** microbes.
  + It is still interesting to think about the information it yields: i.e., it is clear that violin plots calculated on different microbes is different
* The alpha\_diversity violin plot graph as in (1), but the two groups are the same individuals with different microbes.

1. 

* BETA Diversity has 6 graphs: one for each outcome, and the comparison is between ***Remission*** vs ***Non remission*** on the subset of ***Non selected*** microbes (first 3) or ***Selected*** microbes (last 3).
* Statistic alpha is computing the statistic of the violin plot when divided by the ***Remission*** vs ***Non remission*** on the subset of the ***Non selected*** or the ***Selected*** microbes
* The test is a Wilcoxon *NON*-paired test to see if there are significant difference between ***Remission*** vs ***Non remission*** on the subset of ***Selected*** or ***Non selected*** per each outcome.
* ALPHA diversity graphs are specular to (1), but here they are divided and done only on the ***Selected*** or ***Non selected*** subset of microbes.

Out of those 3 types of analysis we found a significant Alpha diversity in the 1st and 2nd types, and in particular in those cohorts, where in yellow highlighted the ones of the 2nd type:

* CD\_TNF\_14\_selected\_microbes\_canc
* CD\_UC\_VDZ\_14
* CD\_TNF\_24\_selected\_microbes\_canc
* CD\_UC\_VDZ\_w0
* CD\_TNF\_24\_selected\_microbes\_no\_canc\_un\_name
* CD\_UC\_TNF\_14\_selected\_microbes\_canc
* CD\_UST\_w0
* CD\_UC\_TNF\_24\_selected\_microbes\_canc
* CD\_VDZ\_14
* CD\_UC\_TNF\_24\_selected\_microbes\_no\_canc\_un\_name
* CD\_VDZ\_w0
* UC\_TNF\_14\_selected\_microbes\_canc