

Supplementary Material

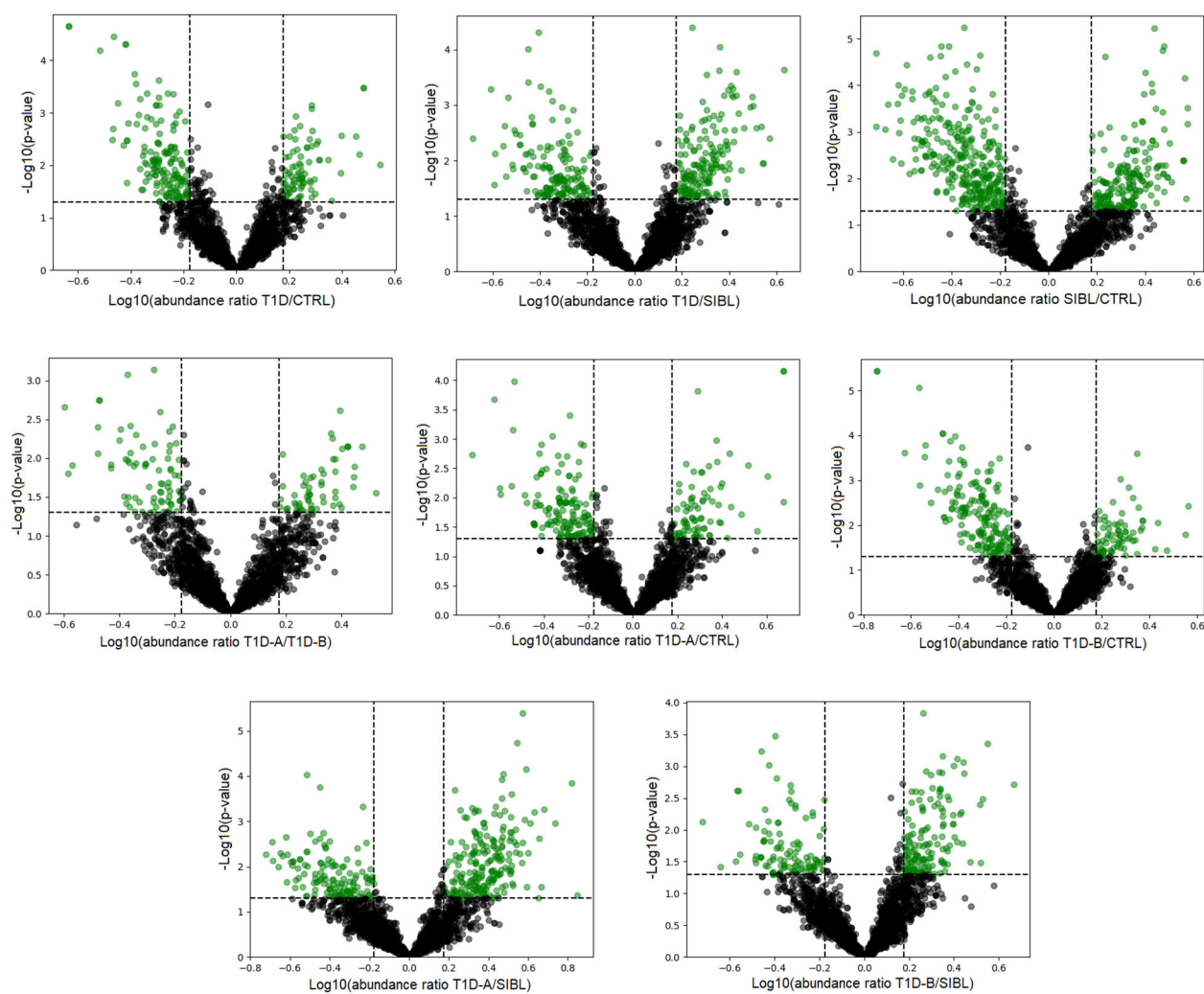


Figure S1. Volcano plots showing significant metaprotein dysregulation in all the discussed t-tests after label-free quantitative analysis. Significant values (spots marked in green) were selected by setting a p-value threshold <0.05 and considering abundance ratios >1.5 and <0.66 .

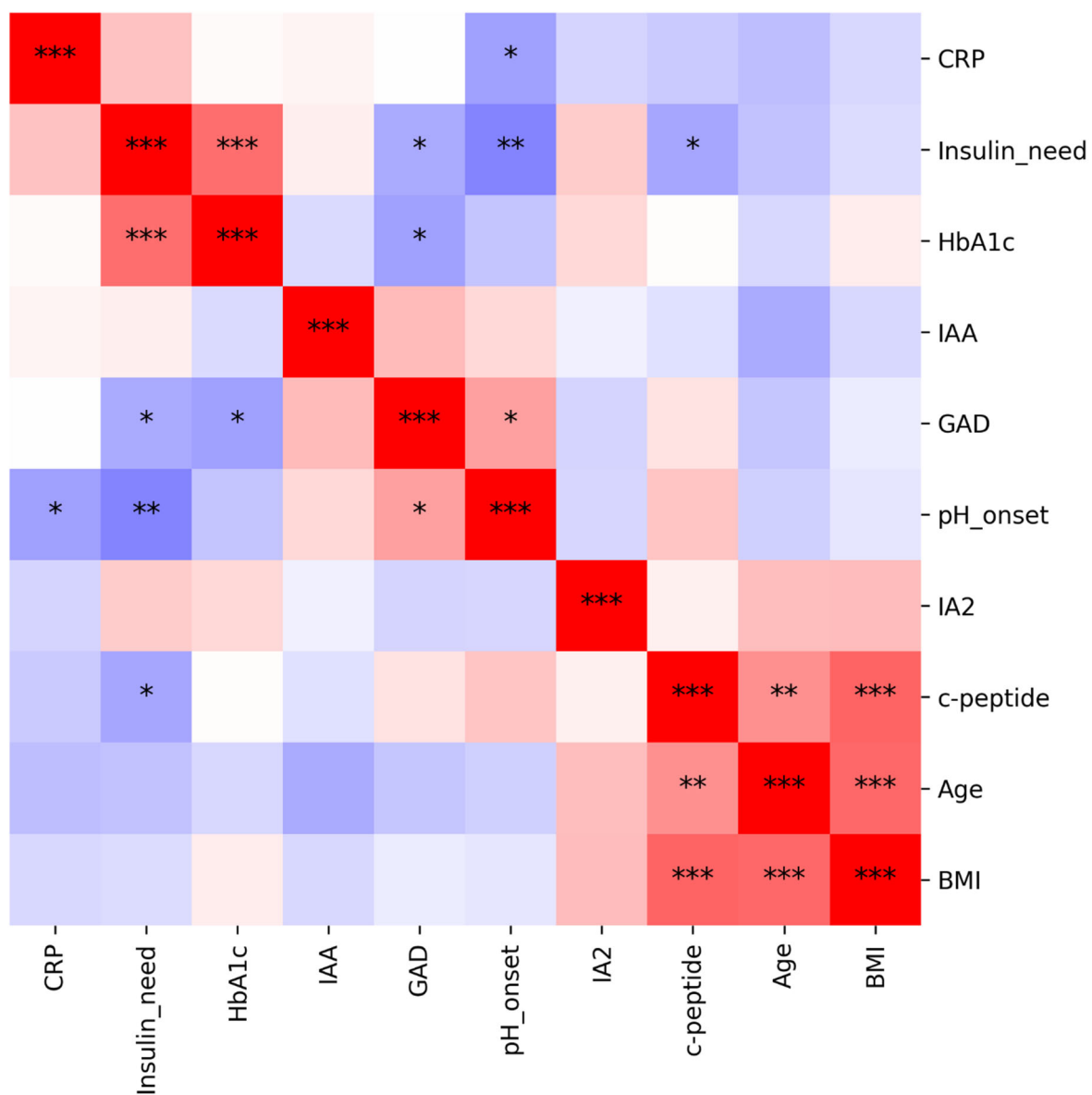


Figure S2. Correlation heatmap of clinical characteristics based on the Pearson's correlation metric.

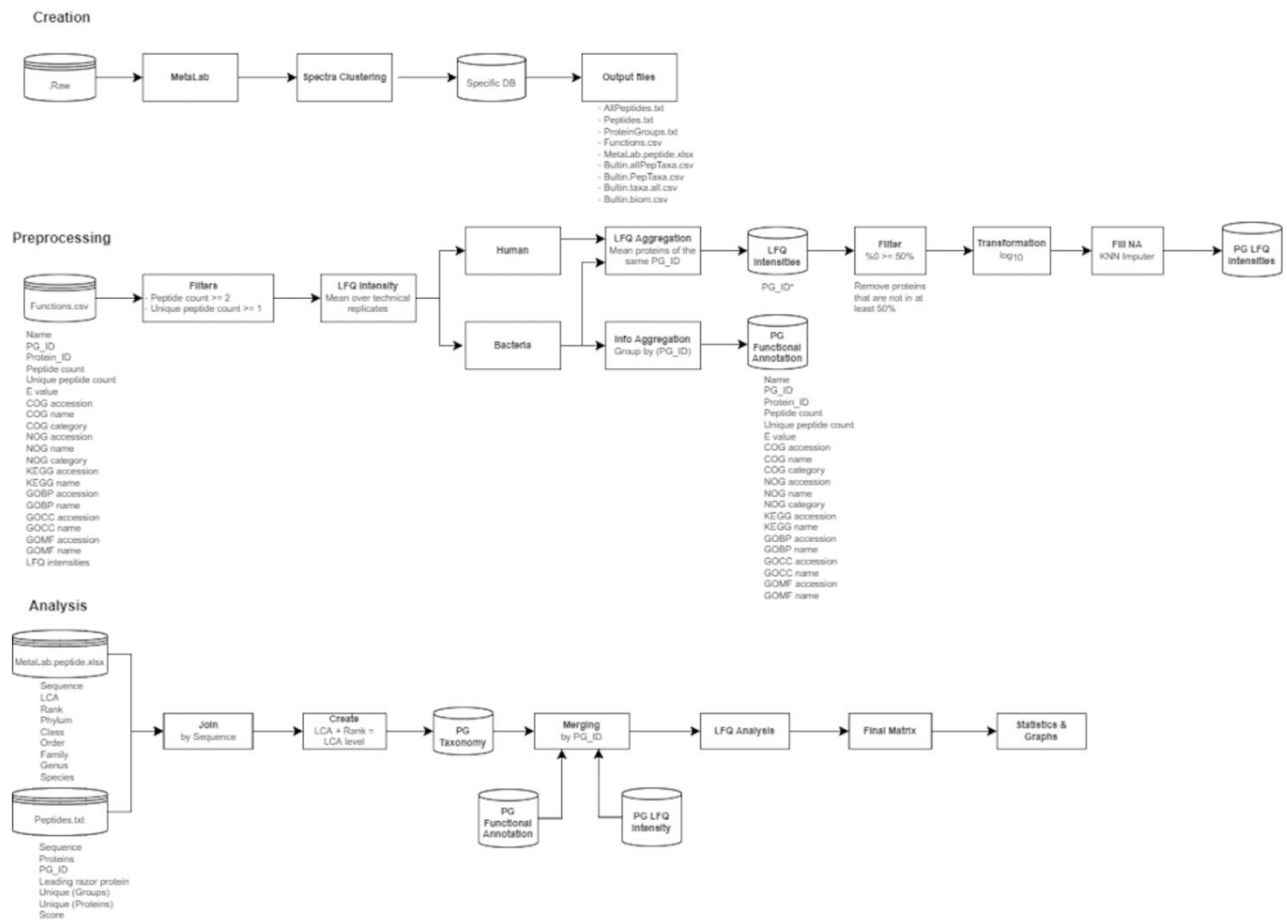


Figure S3. Scheme of the bioinformatic automated workflow for the metaproteomic analysis. Raw data were uploaded in MetaLab and the resulting output files were managed by an ad-hoc python script to obtain a single comprehensive matrix. Peptides and protein intensities were averaged from the results of the two technical replicates, and PG intensities for each individual were determined by averaging the intensities of all the proteins being part of the same PG. A log10 transformation was applied to the label-free quantitative (LFQ) intensities of the PGs and only those PGs with valid LFQ intensity values in $\geq 50\%$ of the whole sample set were used for statistical analysis. Finally, all missing values were imputed by a KNN imputer using a neighborhood of $\sqrt{\text{patients}}$. The resulting filtered matrices were merged together to retrieve a single comprehensive table where taxonomy (LCA rank and the taxon assignment at all levels), functional annotations and LFQ intensities were associated with each PG.