Gut microbiota metabolic reconstruction to assess severity of type 2 diabetes in obesity

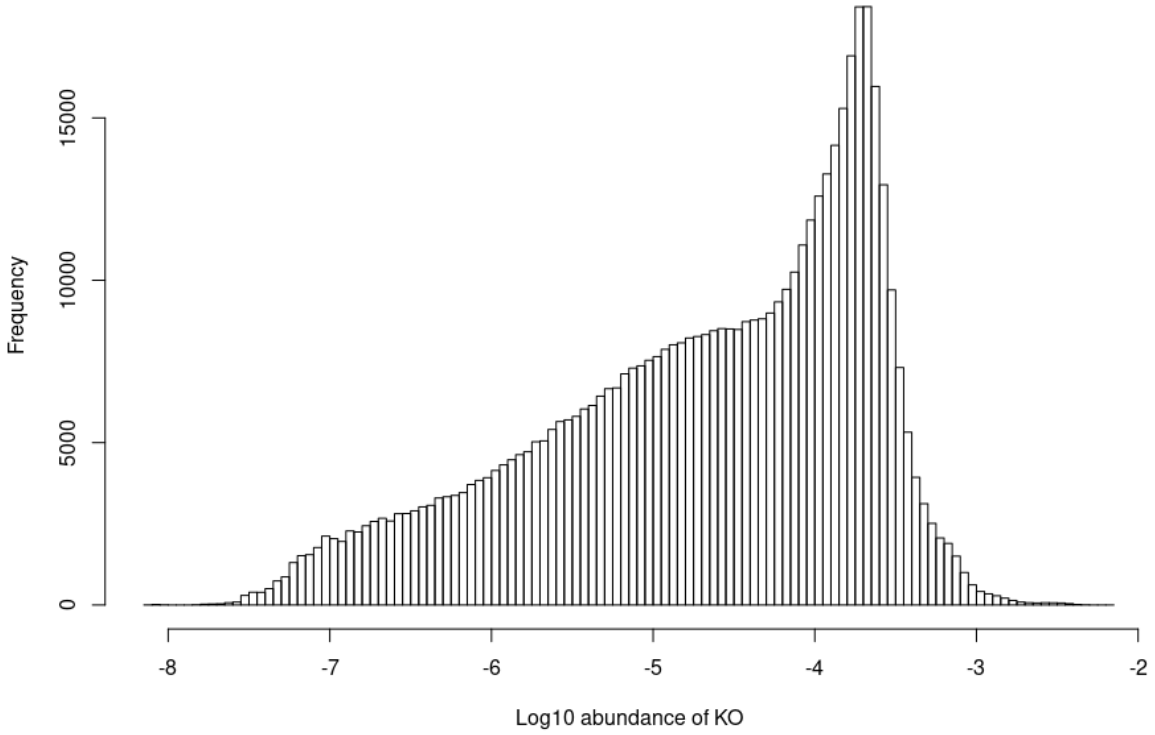

Figure S1 - Log10 abundance of KO histogram for bariatric cohort.

KO abundance threshold

In order to build each gut microbiota metabolic network, we need a list of KO (KEGG Orthology) genes per patient. Our initial data is a KO abundance per patient table, obtained through the fecal metagenomics and consequential mapping to a gene catalog[2]. The list of KOs can only be obtained through an abundance threshold that will binarise the table into genes to be taken into account and genes to be removed for each patient, so as to reduce data noise and increase differences between patients.

Figure S1 shows the histogram of different KO gene abundances, for all patients in the bariatric cohort. This figure allowed us to evaluate which threshold values to check.

We tested a variety of network topological properties for networks at different threshold values, and analysed partial correlations between the topological properties and clinical phenotypes related to diabetes, corrected by age, as is shown in figure S2. We tested threshold values 0, 1e-07, 5e-07, 1e-06, 3e-06, 1e-05, and 1e-04. Higher values than 1e-04 modified too significantly the networks, as there is a great number of genes with this abundance (see figure S1). The best correlations seemed to be at the 1e-05/1e-04 mark, so we further studied in-between values 2e-05, 3e-05, 5e-05, 6e-05, 7e-05, which are shortly after the peak in the histogram in figure S1. Strongest and most numerous correlations seemed to be at the 3e-05 threshold for various diabetes-related phenotypes, therefore being the chosen threshold for the rest of the study.

These network properties were only used in this study for the choice of threshold, so as to choose a threshold that provides networks that contain information, but in the body of our study we used another property not listed here, the scope, so as to avoid bias.

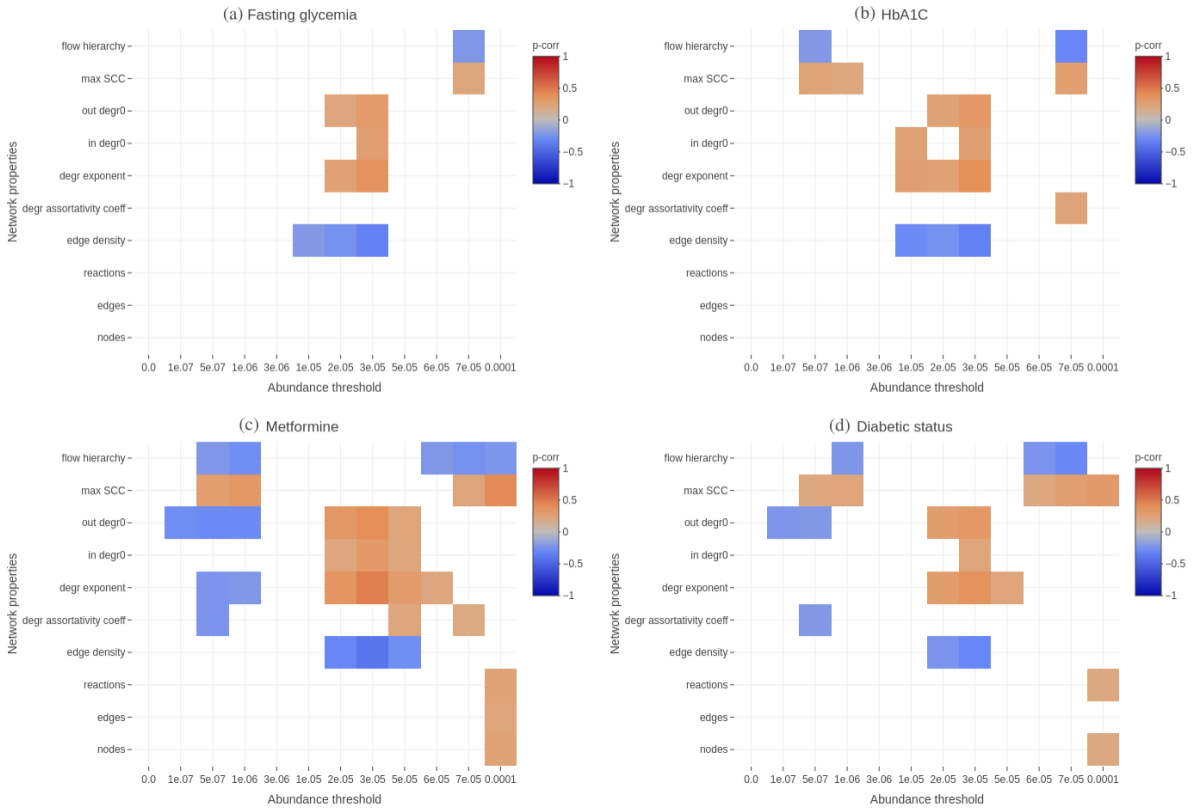

Figure S2 - Partial correlation between various network properties and various diabetic-related variables per abundance threshold, corrected by age, for the bariatric cohort. (a) Against fasting glycemia. (b) Against HbA1c levels. (c) Against metformin intake. (d) Against diabetic status. Non significant correlations are not shown (significance p-value < 0.05). Network topological properties: *nodes* - number of nodes (N), *edges* - number of edges (E), *reactions* - number of reactions, *edge density* - E/N, *degr assortativity coeff* - degree assortativity coefficient (similarity of connections in the network in regard to node degree), *degr exponent* - maximum likelihood estimate of the degree exponent assuming a power law degree distribution[1], *in degr0* - proportion of nodes with an in-degree of 0, *out degr0* - proportion of nodes with an out-degree of 0, *max SCC* - proportion of the size of the largest strongly connected component (SCC), *flow hierarchy* - proportion of nodes in the largest weakly connected component that do not belong to the largest SCC. “1e.0X” values correspond to 1e-0X values.