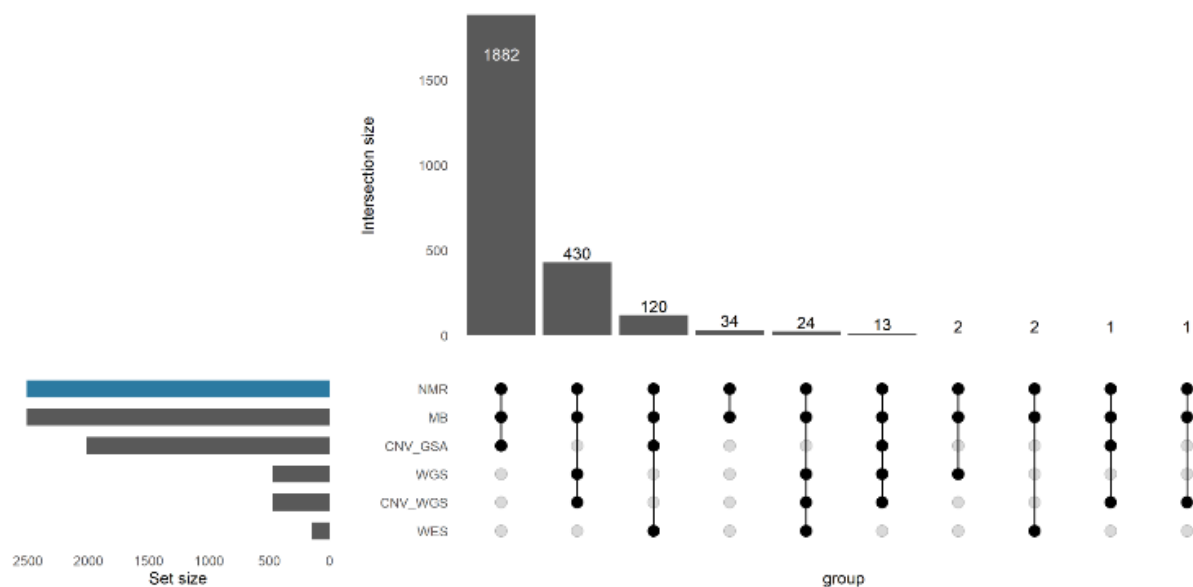


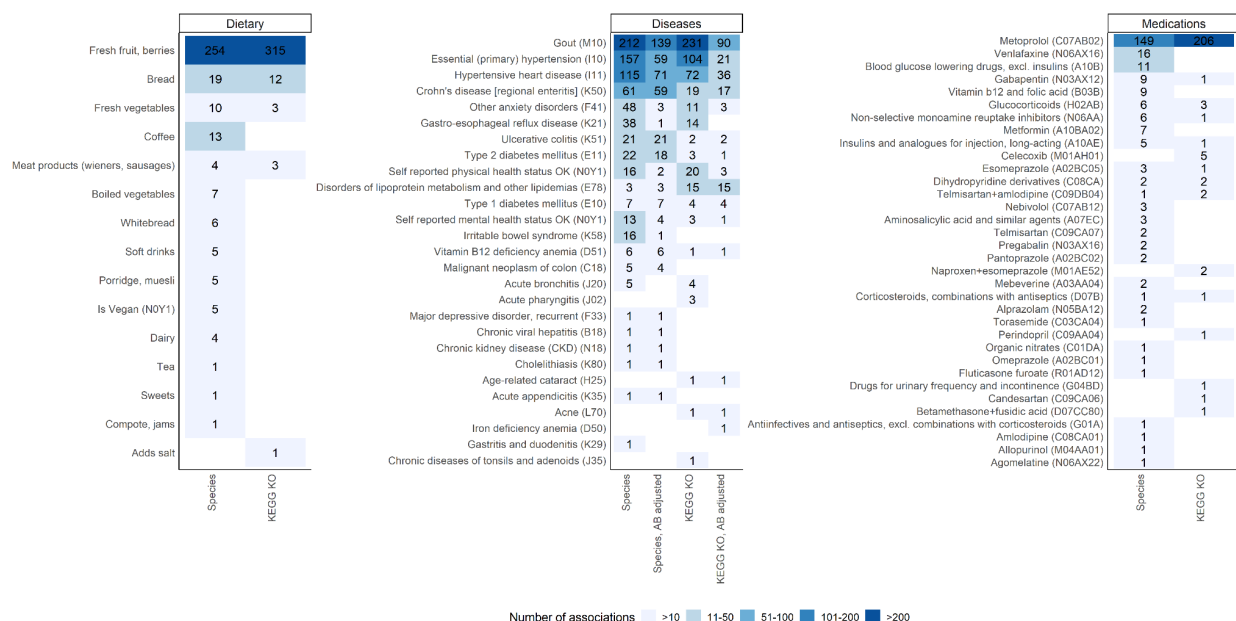
SUPPLEMENTARY INFORMATION

Gut metagenome associations with extensive digital health data in a volunteer-based Estonian microbiome cohort

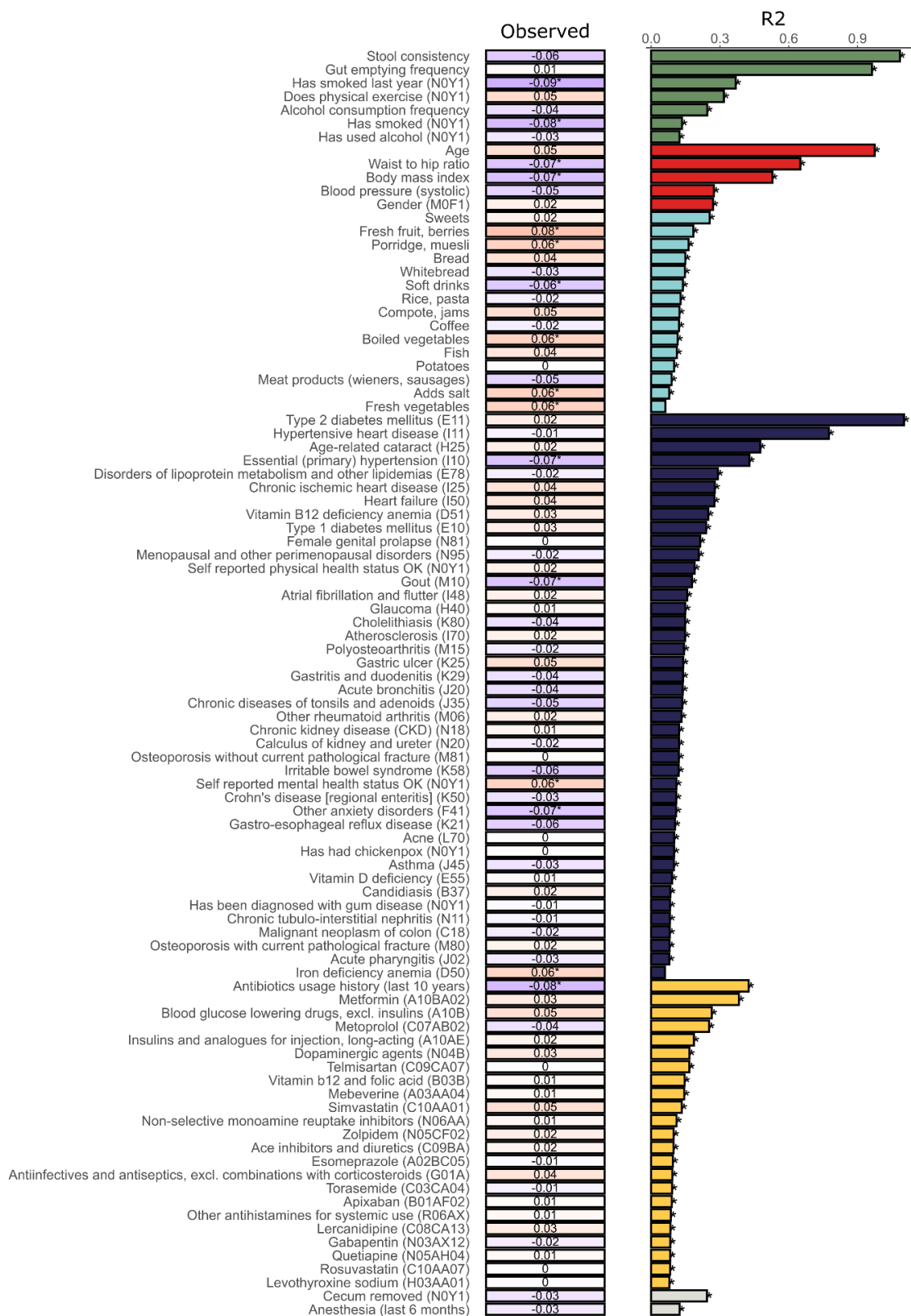
Oliver Aasmets¹, Kertu Liis Krigul¹, Kreete Lüll¹, Andres Metspalu^{1,2,3} and Elin Org^{1*}



Supplementary Fig. 1. Omics data overlap in the Estonian Microbiome cohort. NMR metabolite data (highlighted in blue) will be available in 2022 from the same time points as MB collection. NMR, nuclear magnetic resonance; MB, microbiome; WES, whole-exome sequencing; WGS, whole-genome sequencing; CNV_GSA, copy number variation data from genotyped individuals; CNV_WGS, copy number variation data from individuals with whole-genome sequencing.



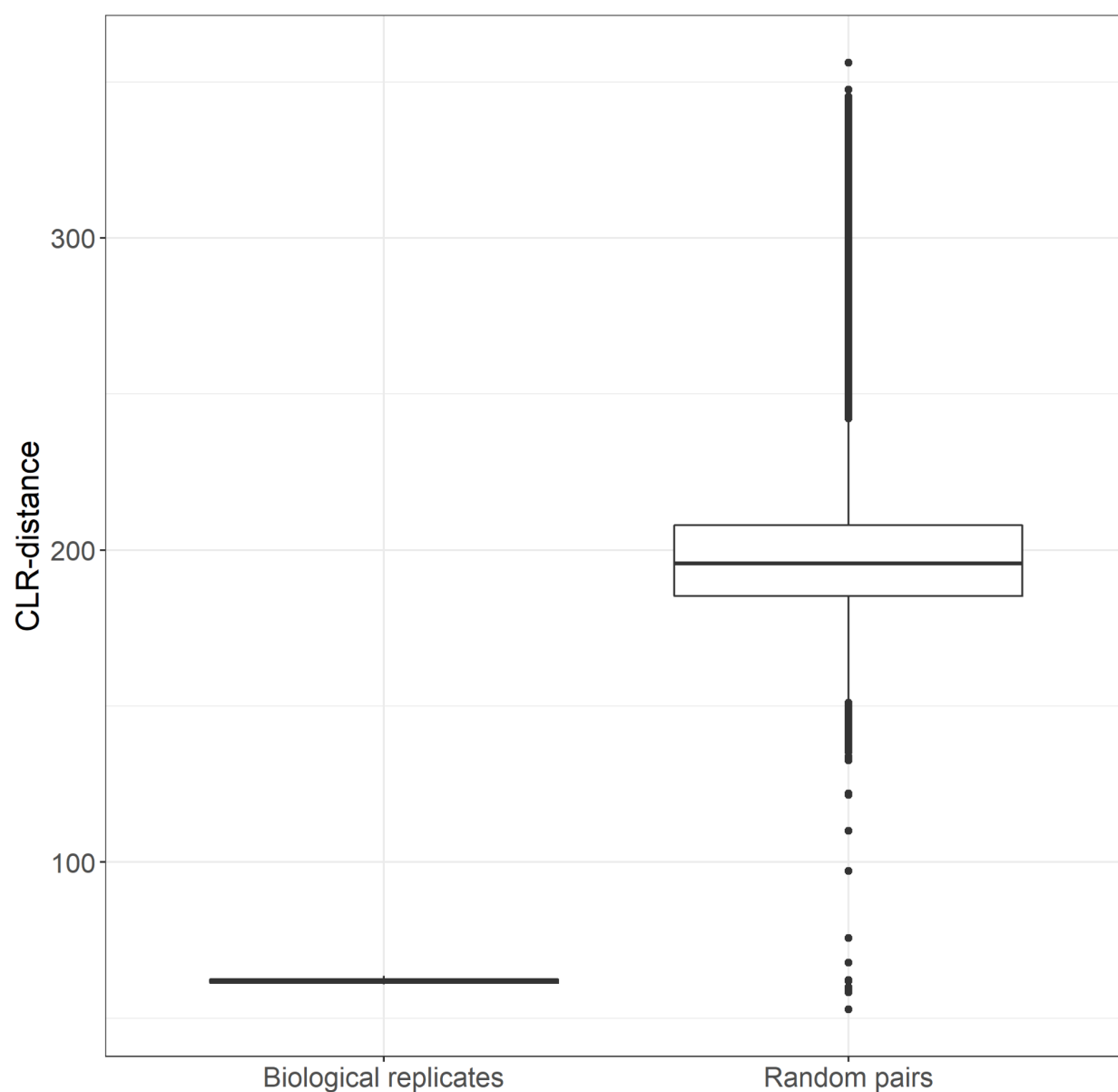
Supplementary Fig. 2. Summary of associations found by metagenome association analysis with microbial species and KOs. Abbreviations in the y-axis labels are the international classification of diseases-10 and anatomical therapeutic chemical codes for diagnoses and medications, respectively. AB, antibiotics; KEGG KO, Kyoto Encyclopedia of Genes and Genomes Orthology.



■ Dietary
 ■ Intrinsic
 ■ Other
 ■ Diseases
 ■ Medications
 ■ Procedures

Supplementary Fig. 3. Statistically significant associations with functional (KOs) alpha and beta diversity. The bar plot indicates the explained variance in the interindividual variation of microbial composition obtained by the permutational analysis of variance (based on the Euclidean distance on the CLR-transformed data). The heatmap shows the Spearman correlation coefficients of each factor with the observed KO richness. Blue indicates negative correlation, and red indicates positive correlation. Asterisks indicate associations with false discovery rate < 0.05 . Abbreviations in the y-axis correspond to the international classification of diseases-10 codes for diagnoses and anatomical therapeutic chemical codes for medications.

Supplementary Fig. 4. Species co-occurrence networks depicting microbial associations with complex diseases before and after correction for long-term antibiotic (AB) usage. Each species (node) is colored according to its association with disease (blue corresponds to a negative association, and red corresponds to a positive association). Only statistically significant associations are shown (false discovery rate < 0.05). Abbreviations on top of the figures correspond to the international classification of diseases-10 codes for diagnoses.



Supplementary Fig. 5. Similarity of the microbiome composition between biological replicates ($n = 5$) and pairs of unrelated samples ($n = 2509$) based on the Euclidean distance on the center log ratio (CLR)-transformed species-level microbiome profile. The central line, box and whiskers in boxplots represent the median, interquartile range (IQR), and 1.5 times the IQR, respectively.