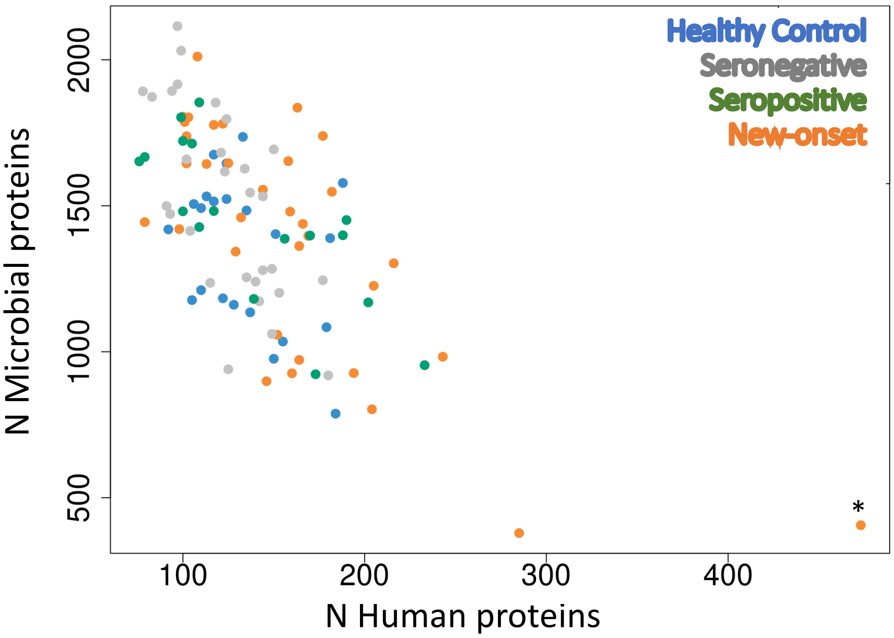
Supplementary Figure S1.

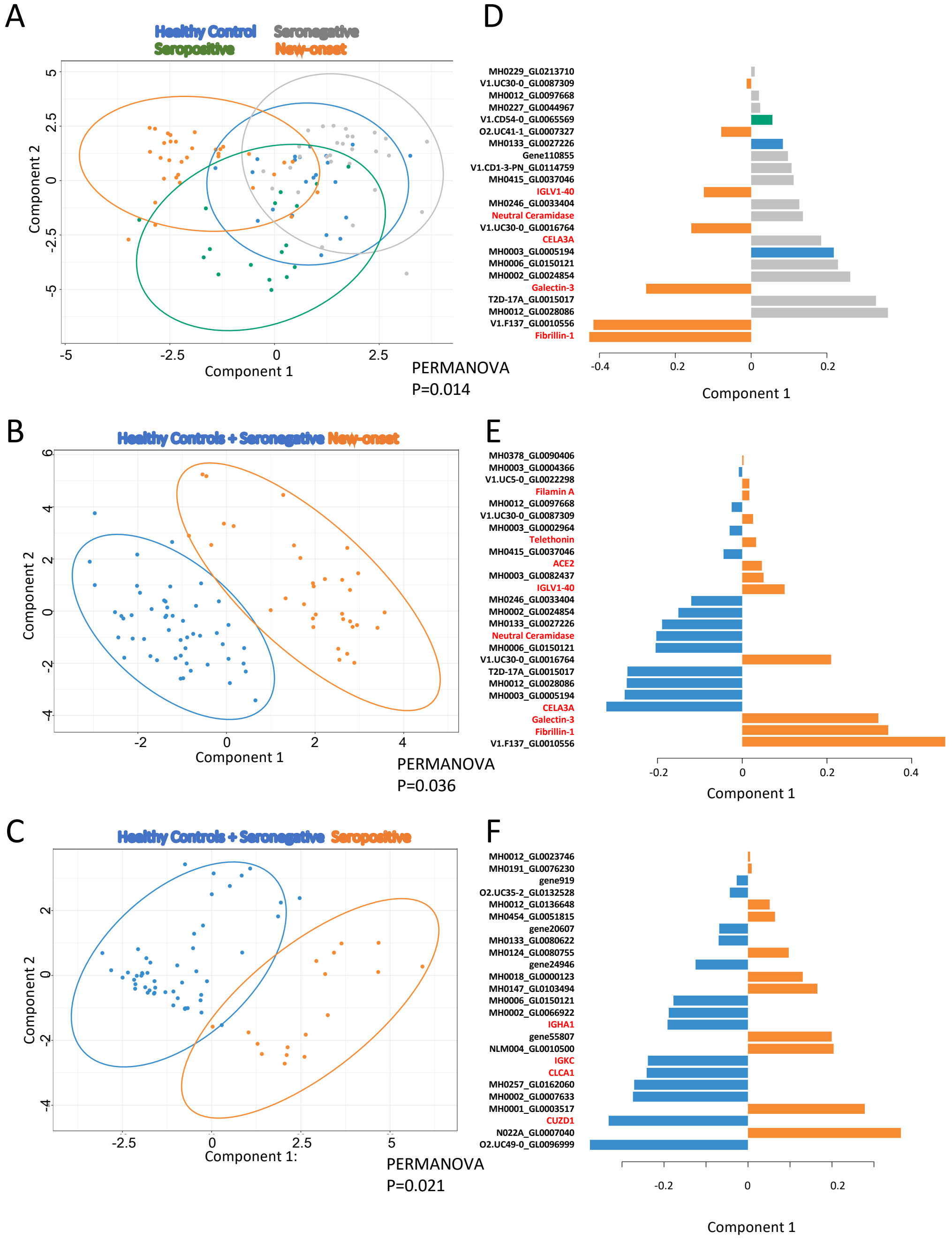
Variation in number of identified human and microbial proteins. Number of identified microbial and human proteins identified in each sample. Note outlier sample in bottom right corner was excluded from further analysis. *Outlier was excluded from further analysis.



SUPPLEMENTARY DATA

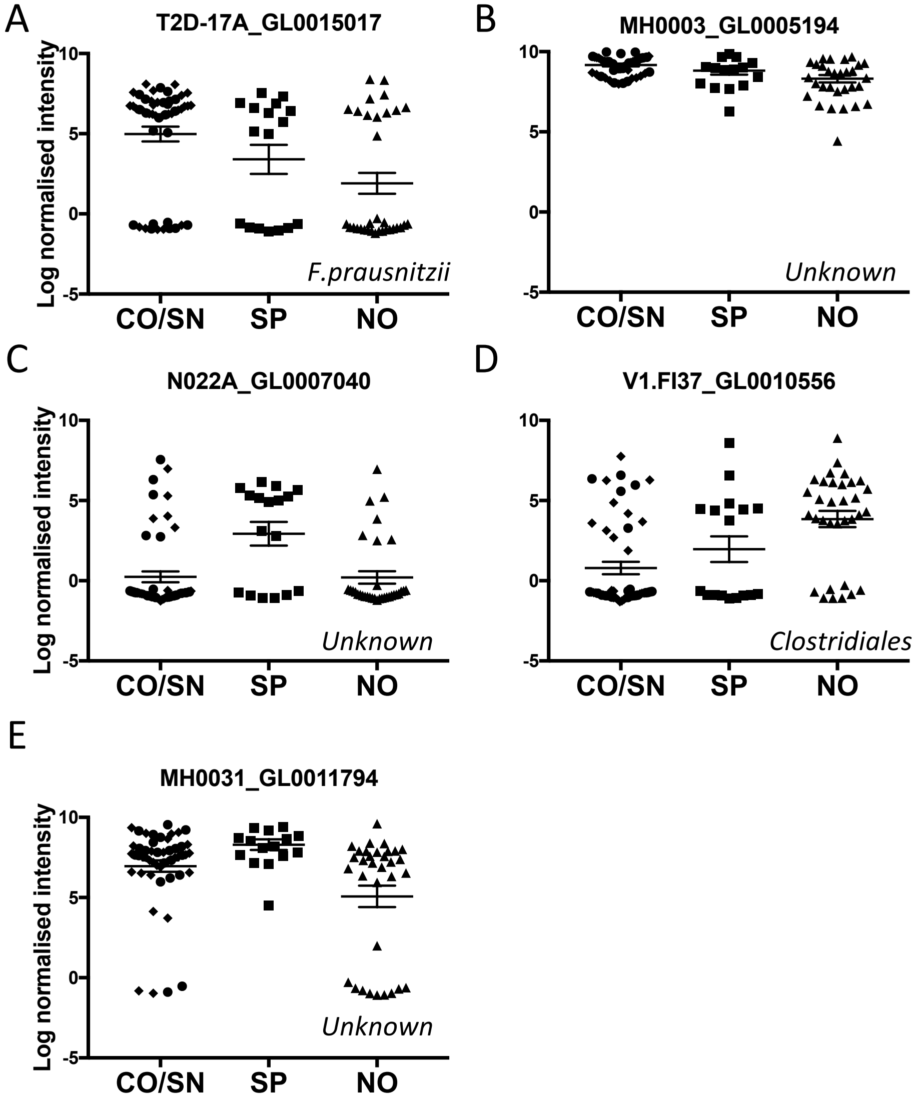
Supplementary Figure S2.

sPLS-DA multi-variate analysis of stool proteins discriminates new-onset and seropositive individuals from healthy controls and seronegative individuals. (A) Comparison of 4 individual groups. (B) New-onset and combined control groups. (C) Seropositive and combined control groups. (D-E) Component 1 contributing variables for each plot. Proteins in red text: human, black text : microbial. 158742018-stool1_revised_scaffold9355_1_gene110855 abbreviated as gene110855, 763577454-stool2_revised_C1073689_1_gene919 abbreviated as gene919, 160421117 stool1_revised_scaffold9704_1_gene20607 abbreviated as gene20607, 763536994 stool1_revised_scaffold21106_2_gene24946 abbreviated as gene24946 and 404239096 stool1_revised_scaffold7674_1_gene55807 abbreviated as gene55807.



Supplementary Figure S3.

Individual human proteins that significantly differ between new-onset, seropositive and control seronegative subject groups. (A-G) Log normalized intensity of significant proteins comparing new-onset (NO), seropositive (SP) and combined healthy control and seronegative (CO/SN) subject groups by one-way ANOVA (adj $P<0.05$). Mean +/-SEM are shown. Circles: healthy controls, diamonds: seronegative individuals.

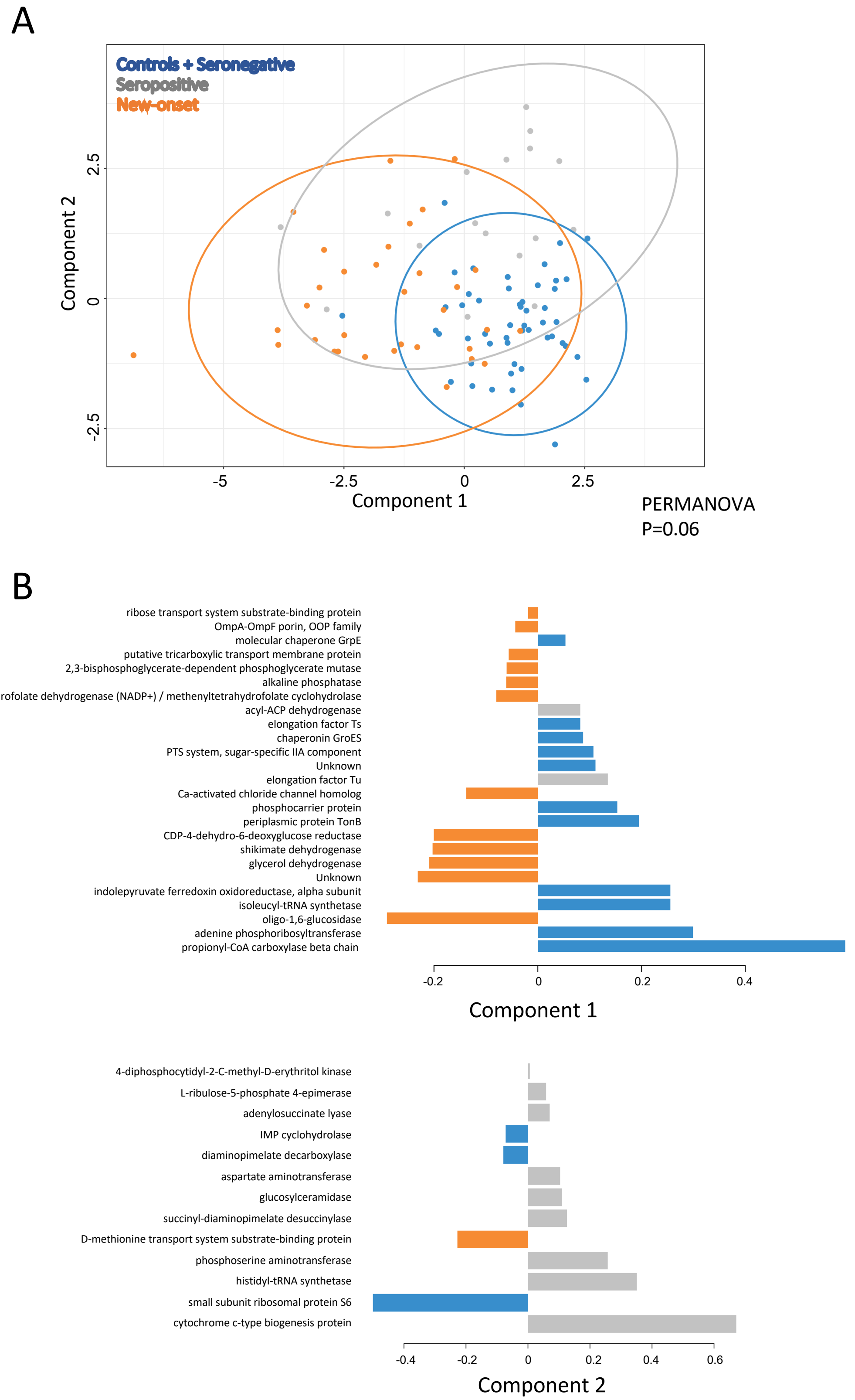


SUPPLEMENTARY DATA

Supplementary Figure S4.

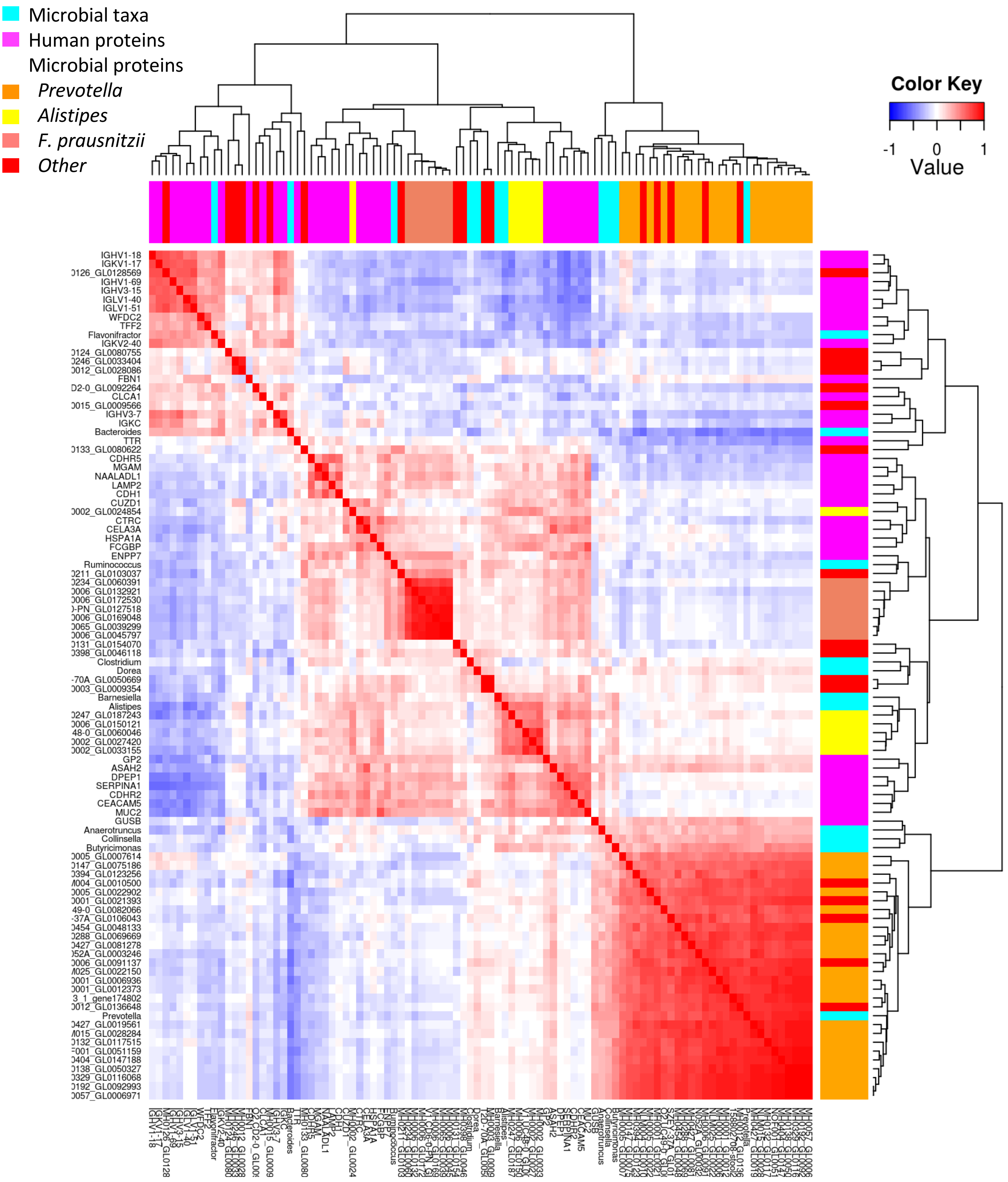
sPLS-DA multi-variate analysis of KEGG orthologues.

(A) sPLS-DA analysis of microbial KEGG orthologues, P=0.06. (B) Component 1 and 2 contributing variables.



Supplementary Figure S5.

Correlations between variables selected in the DIABLO models. Pearson's correlations were calculated between each of the variables selected in the DIABLO models depicted in Fig. 2A-B and visualized in a heat-map with hierarchical clustering analysis.



Supplementary Figure S6.

Correlation between age and multivariate biomarker signature Pearson's correlation between age and probability of being classified as NOD for (A) all subjects and (B) only subjects <25 years of age. C: Probability of being classified as NO for subjects <25 years. Probabilities were determined using a linear model developed on all CO/SN and NO patients. The ten proteins selected for this model included MH0003_GL0005194, T2D.17A_GL0015017, V1.FI37_GL0010556, Galectin-3, MH0133_GL0027226, Fibrillin-1, CELA3A, Neutral Ceramidase, MH0012_GL0028086, and V1.UC30.0_GL0016764. Blue dots control and seronegative, grey dots seropositive and orange dots new-onset subjects.