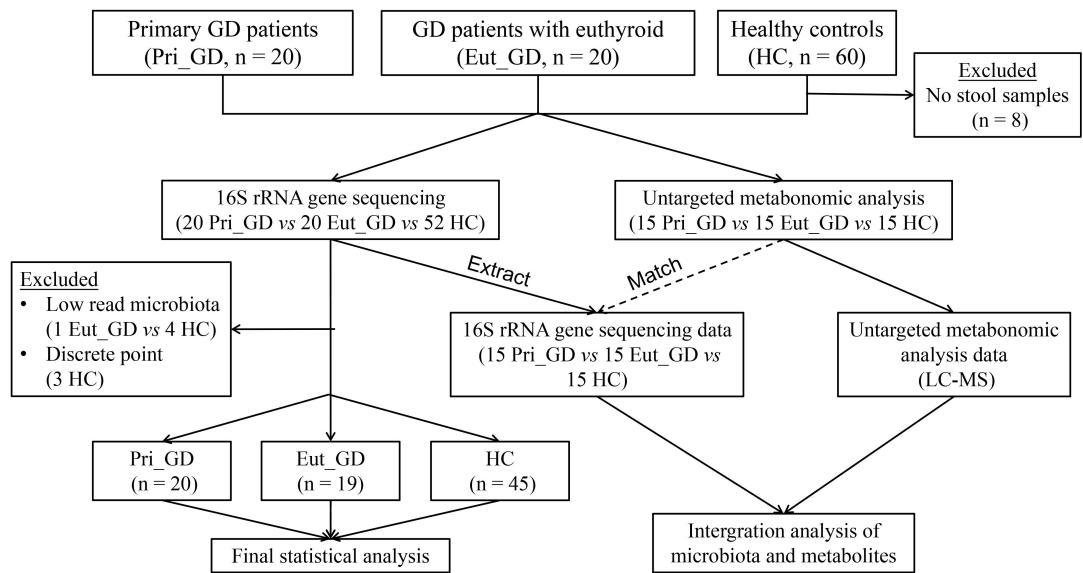


Supplementary Figures

Figure S1: Flowchart explaining recruitment of subjects



In this study, 84 samples for microbiome analysis, and 45 samples were randomly selected from 84 samples for metabolic analysis. Lastly, microbial data of 45 samples were extracted and integrated with metabolomic data.

Figure S2: The process of building disease prediction model based on random forest method

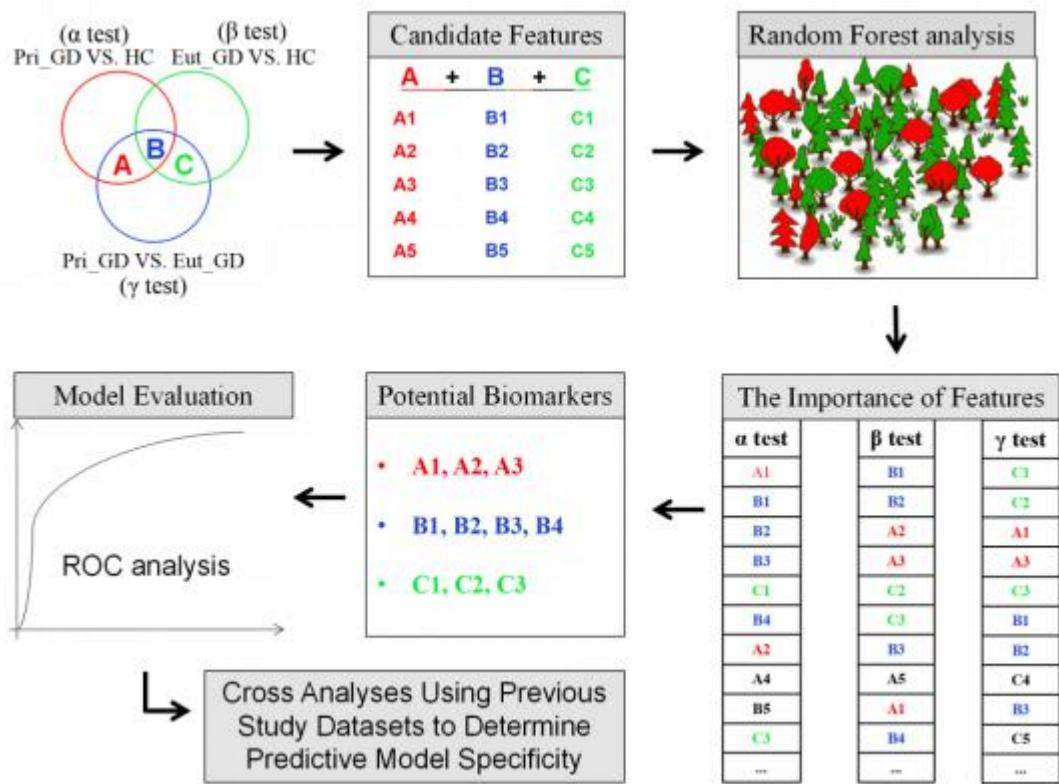
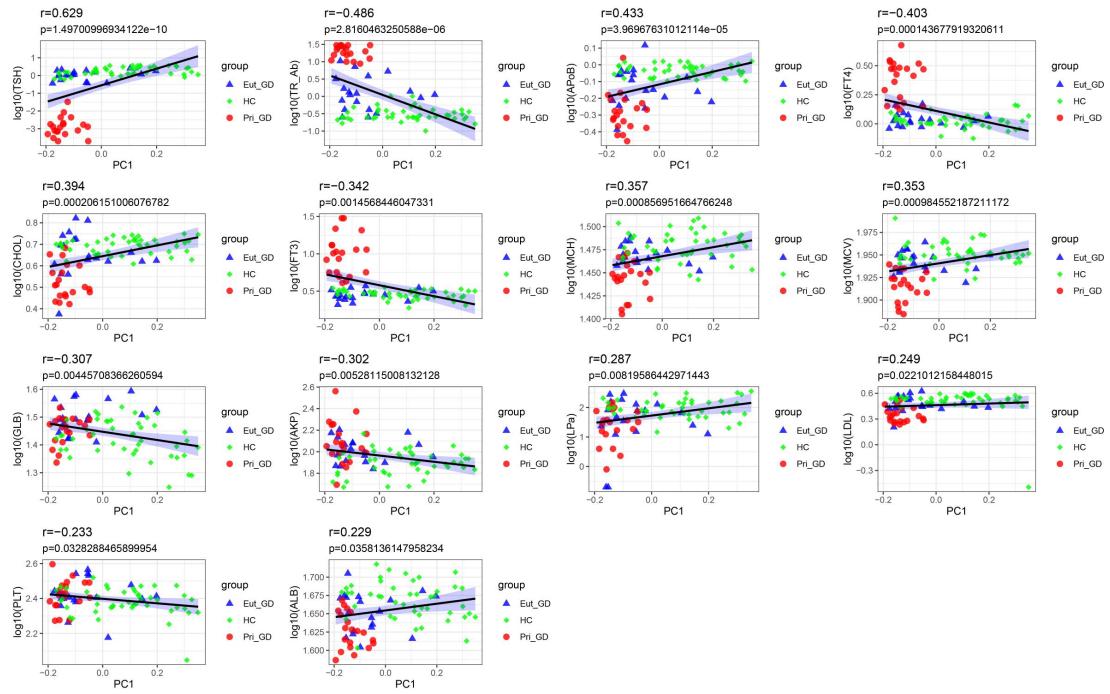
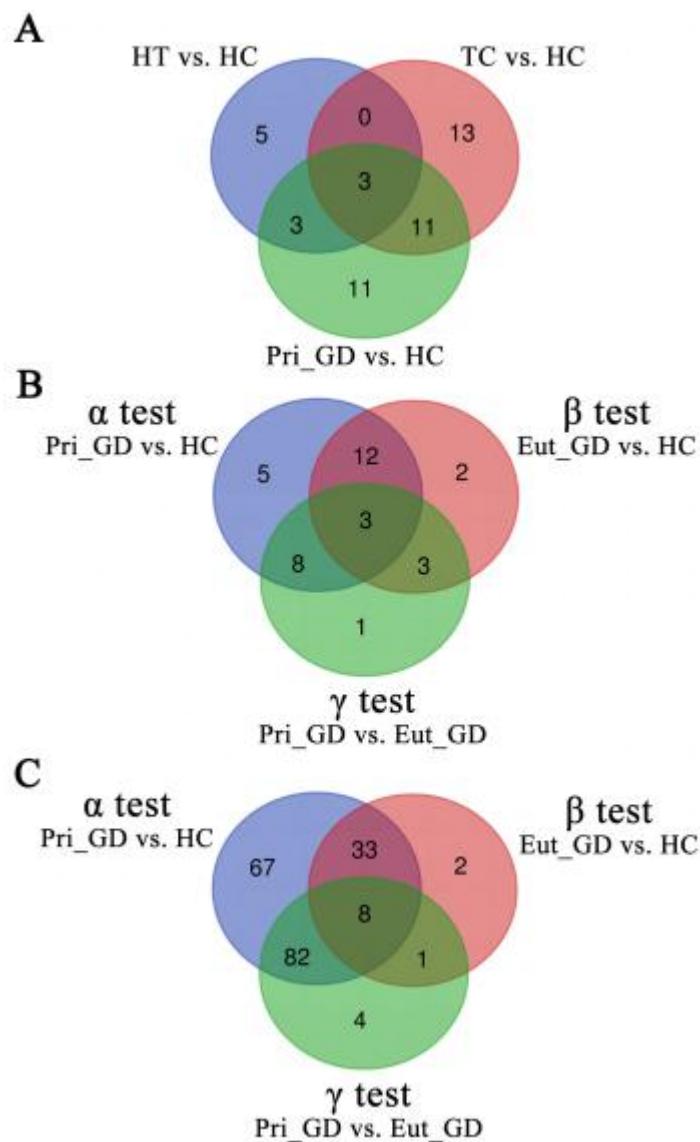


Figure S3: Linear correlation analysis between PC1 and clinical parameters



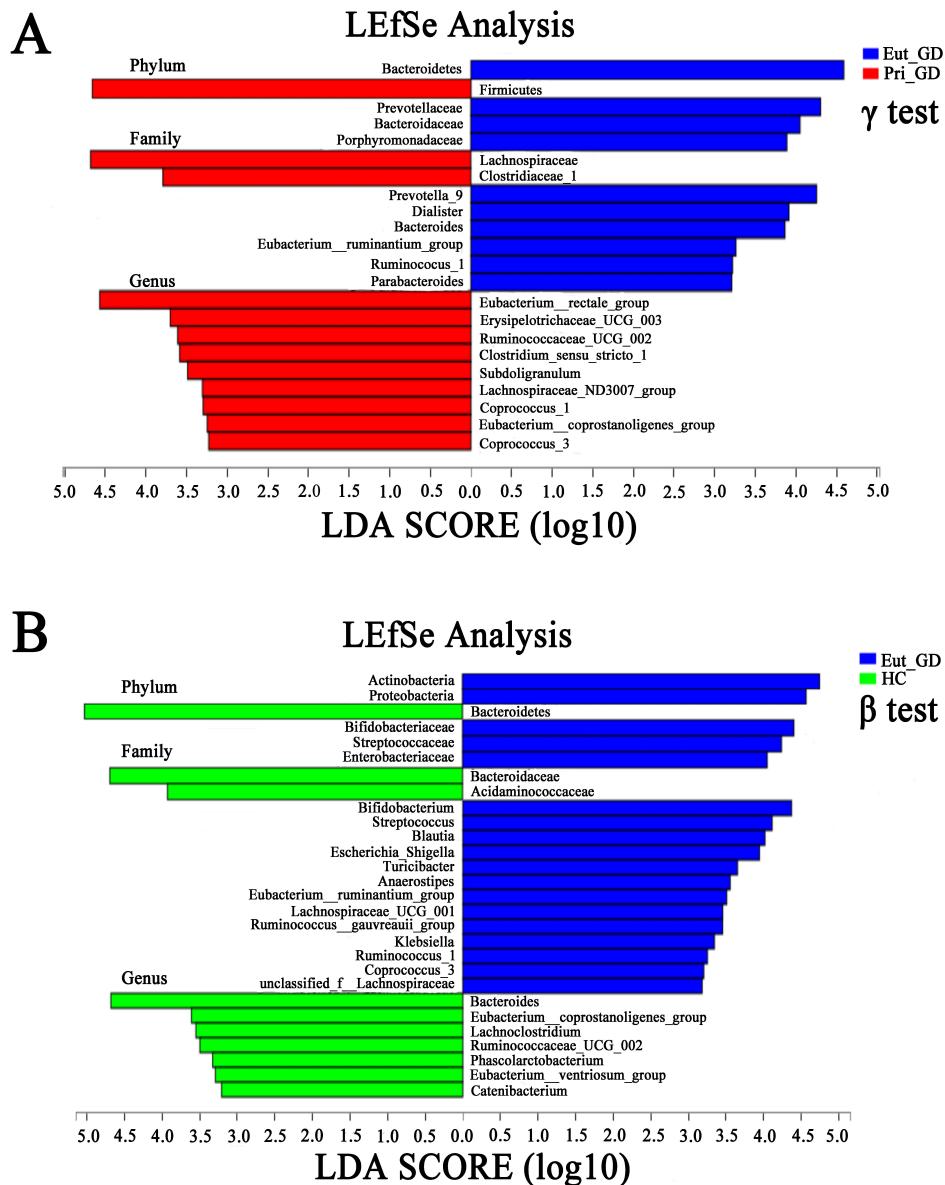
The black line represents the correlation curve between variables, and the blue shadow represents the 95% confidence interval. Points of different colors and shapes represent different groups.

Figure S4: The Venn analysis of different species or metabolites



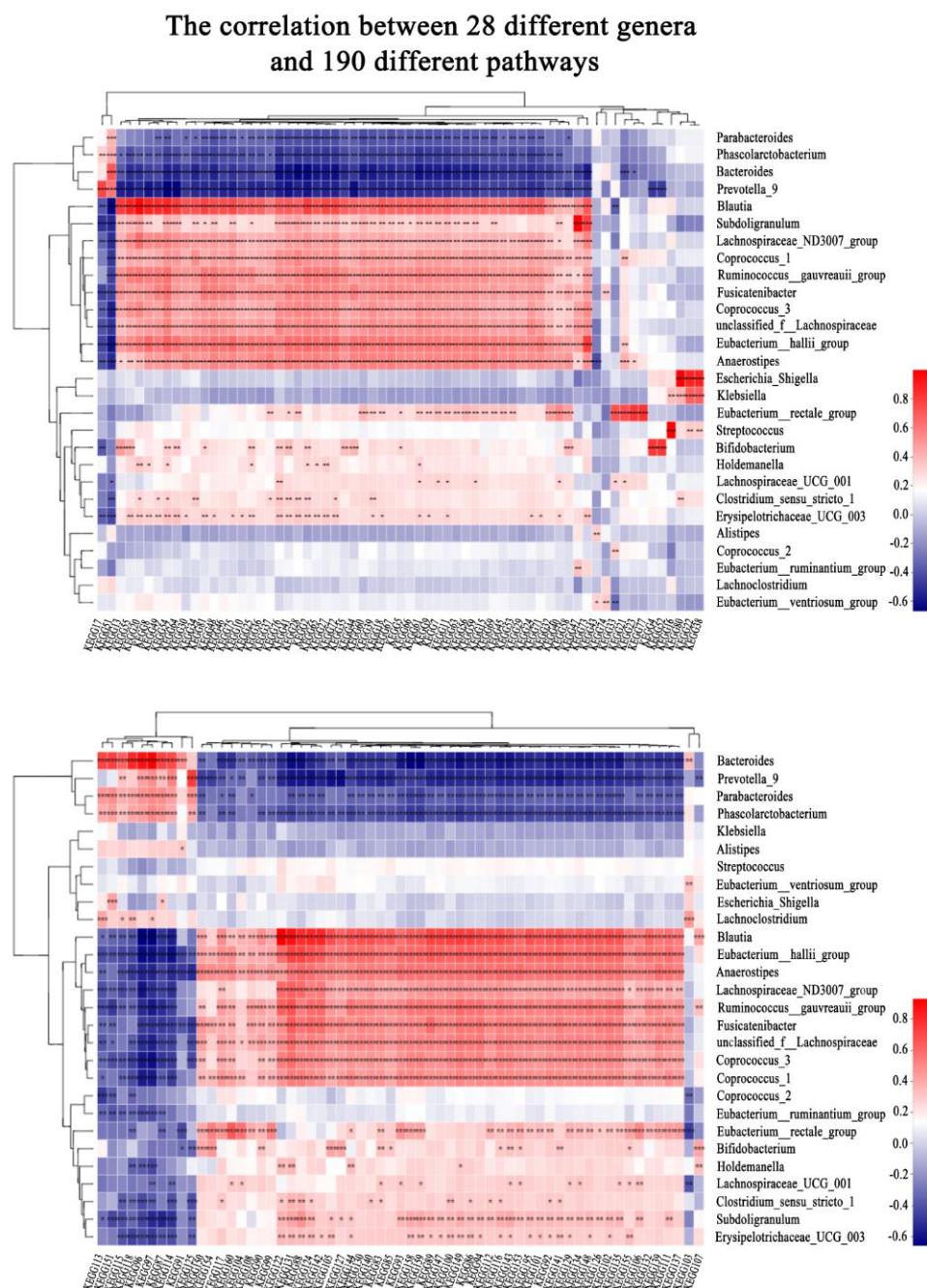
A, B. The Venn analysis of different species between studies or tests. **C.** The Venn analysis of different metabolites between tests.

Figure S5: Differential species at phylum, family and genus level



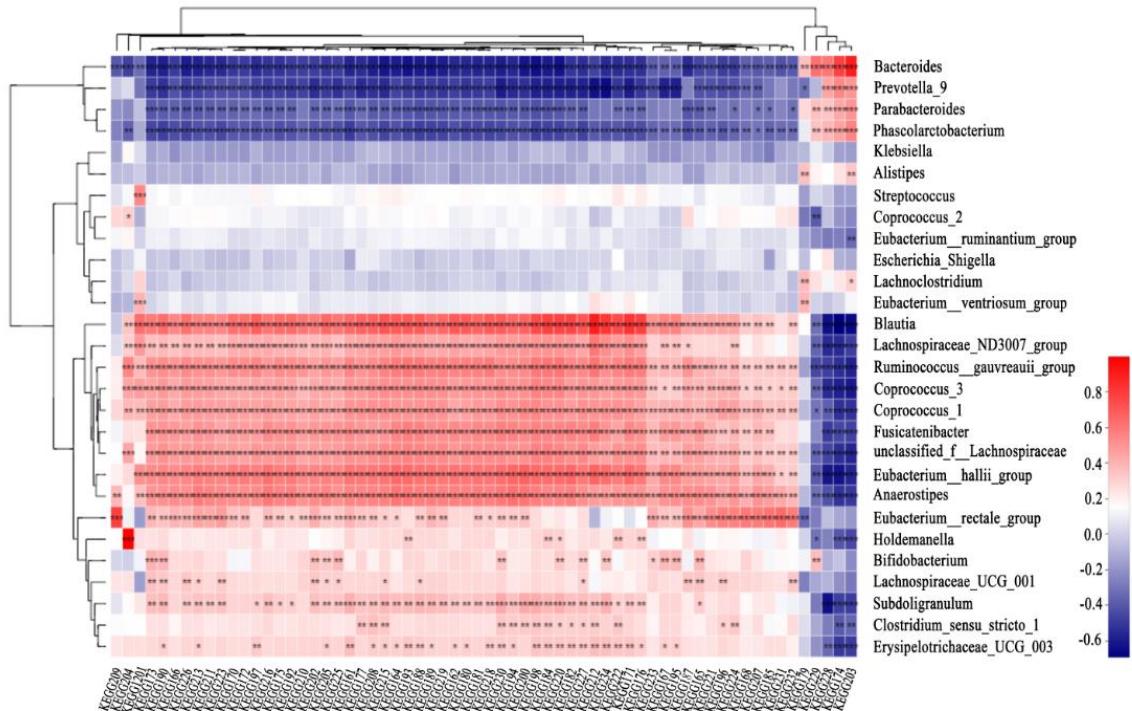
A. Species difference tests using LEfSe analysis, **B.** using Mann-Whitney U test.

Figure S6: The correlation of differential genera and differential KEGG pathways



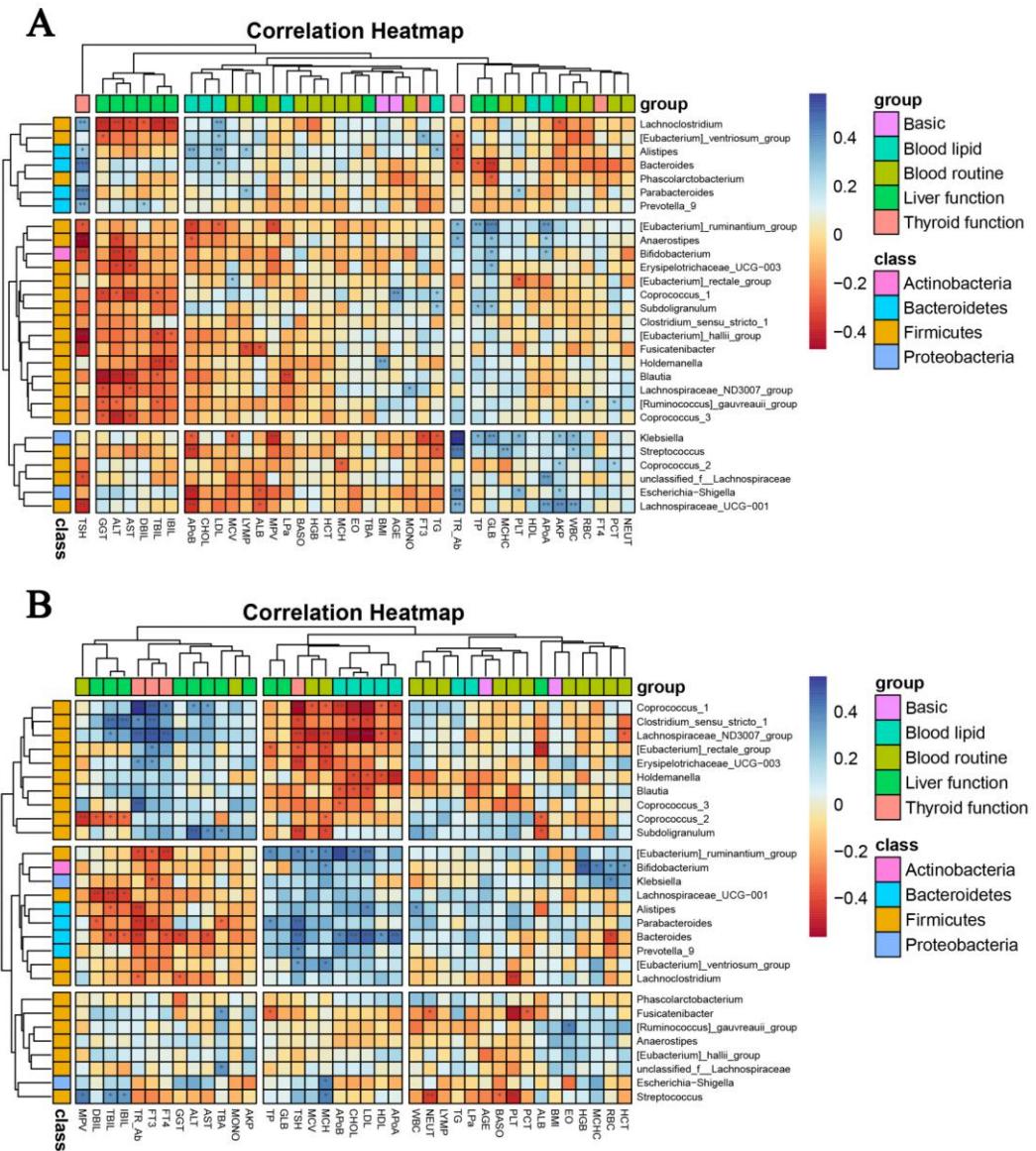
A, B. In the heatmap, color intensity represents the magnitude of correlation, and marker represents the significantly correlation. Red, positive correlations ($r > 0.3$); blue, negative correlations ($r < -0.3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Figure S7: The correlation of differential genera and differential KEGG pathways



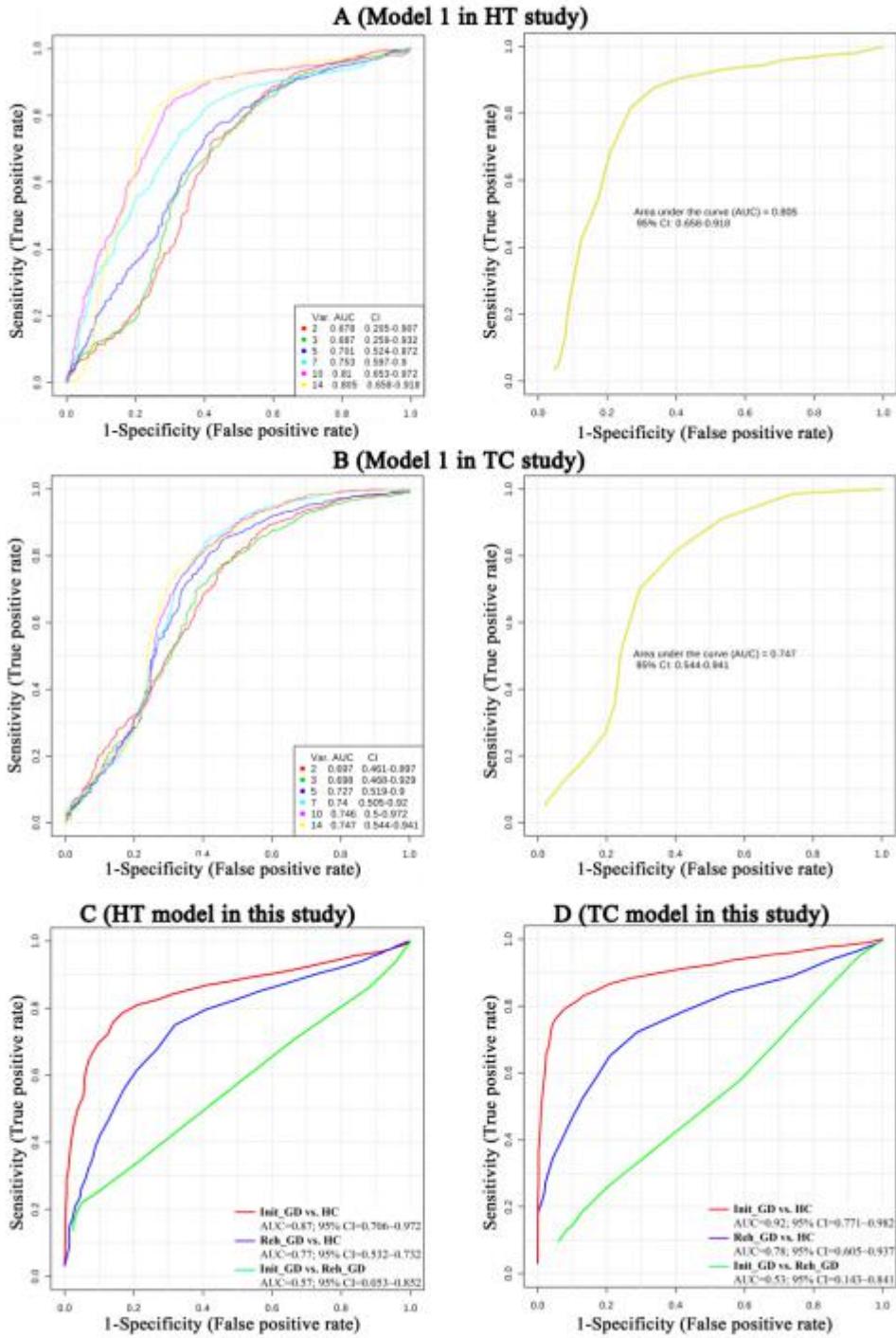
In the heatmap, color intensity represents the magnitude of correlation, and marker represents the significantly correlation. Red, positive correlations ($r > 0.3$); blue, negative correlations ($r < -0.3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Figure S8: The correlation of differential genera and clinical parameters



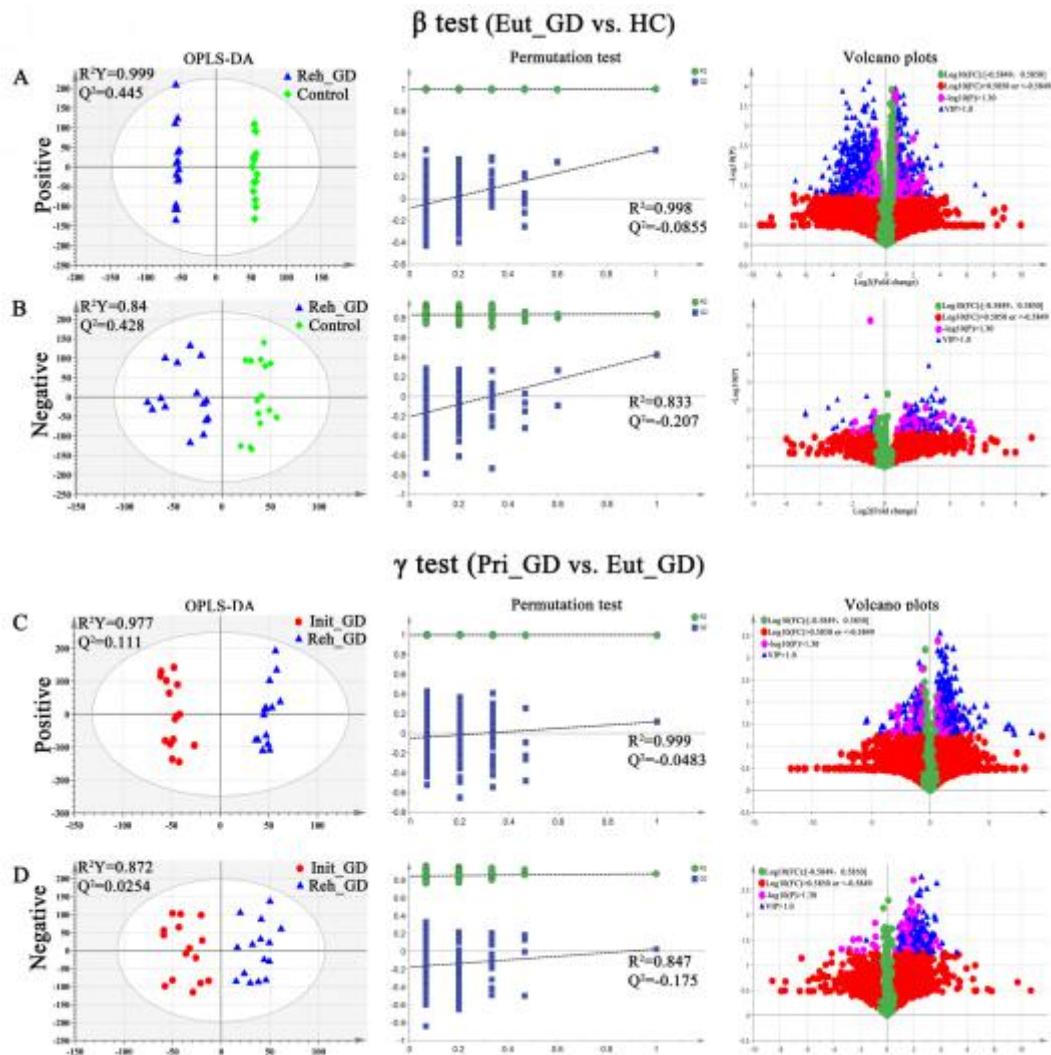
A. The correlation of 28 different genera from α test and 39 clinical parameters between Eut_GD and HC groups, and **B.** between Eut_GD and Pri_GD groups. In the heatmap, color intensity represents the magnitude of correlation, and marker represents the significantly correlation. Red, negative correlations ($r < -0.3$); blue, positive correlations ($r > 0.3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Figure S9: Cross-analysis between different disease prediction models



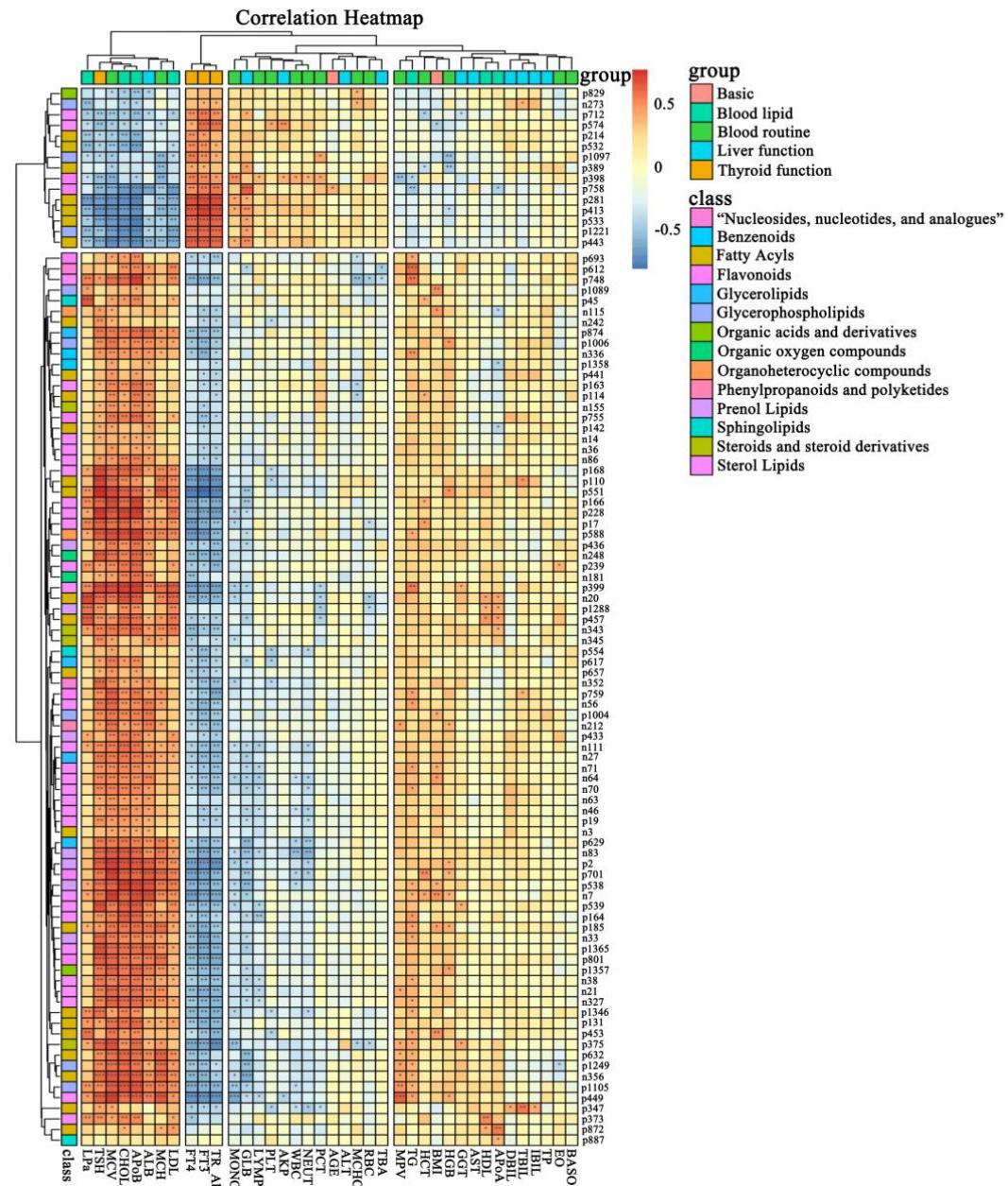
A. The accuracy of Model 1 was evaluated using ROC analysis based on the feature elimination step in distinguish HT and HC groups, and **B.** in distinguish TC and HC groups. **C, D.** The accuracy of HT model or TC model was evaluated using ROC analysis based on the feature elimination step in distinguish Pri_GD and HC groups, respectively.

Figure S10: Multivariate statistical analysis and differential metabolite identification



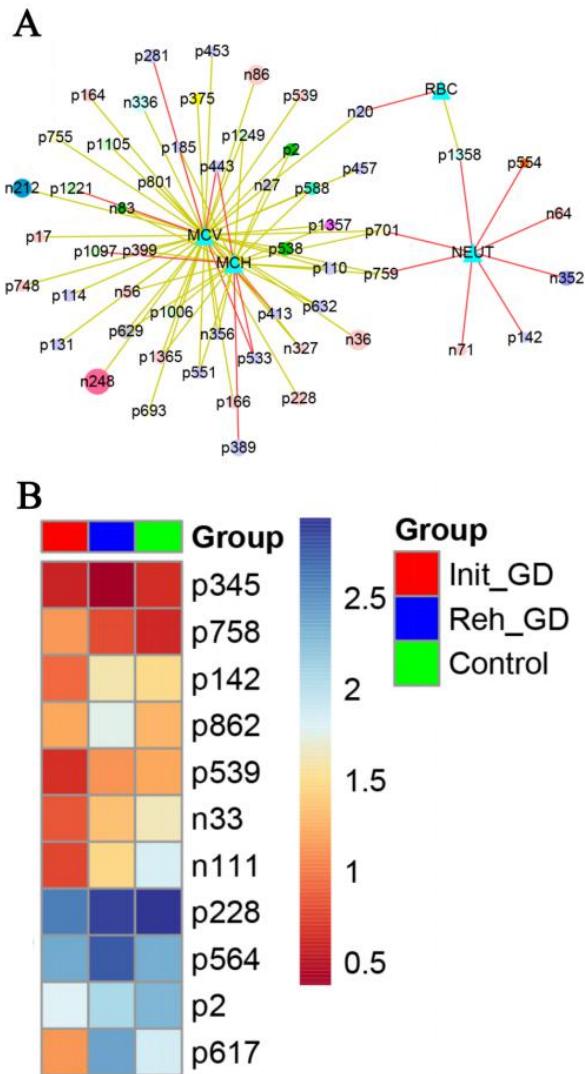
A, B. OPLS-DA analysis was used to construct a classification model for identifying differentiated metabolites, permutation test for OPLS-DA showed that the model is stable and available, and volcano plots for showing differential metabolites between Eut_GD and HC groups, and **C, D.** between Pri_GD and Eut_GD groups.

Figure S11: The correlation of differential gut metabolites and clinical parameters between Pri_GD and HC groups



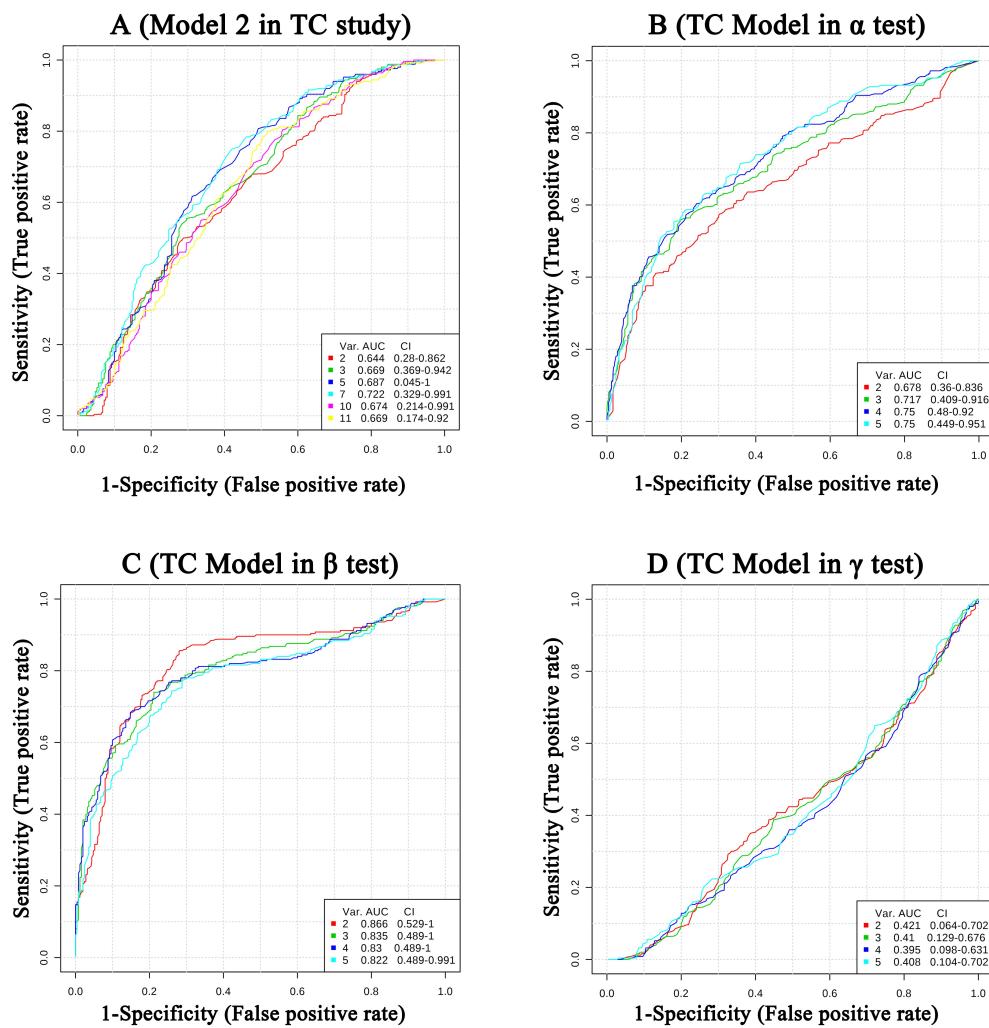
The correlation of 99 different metabolites and 39 clinical parameters between Eut_GD and HC groups. In the heatmap, color intensity represents the magnitude of correlation, and marker represents the significantly correlation. Red, positive correlations ($r>0$); blue, negative correlations ($r<0$). * $P<0.05$; ** $P<0.01$; *** $P<0.001$.

Figure S12: Network correlation analysis between blood routine parameters and gut metabolites, and model 2 classification analysis using heatmap



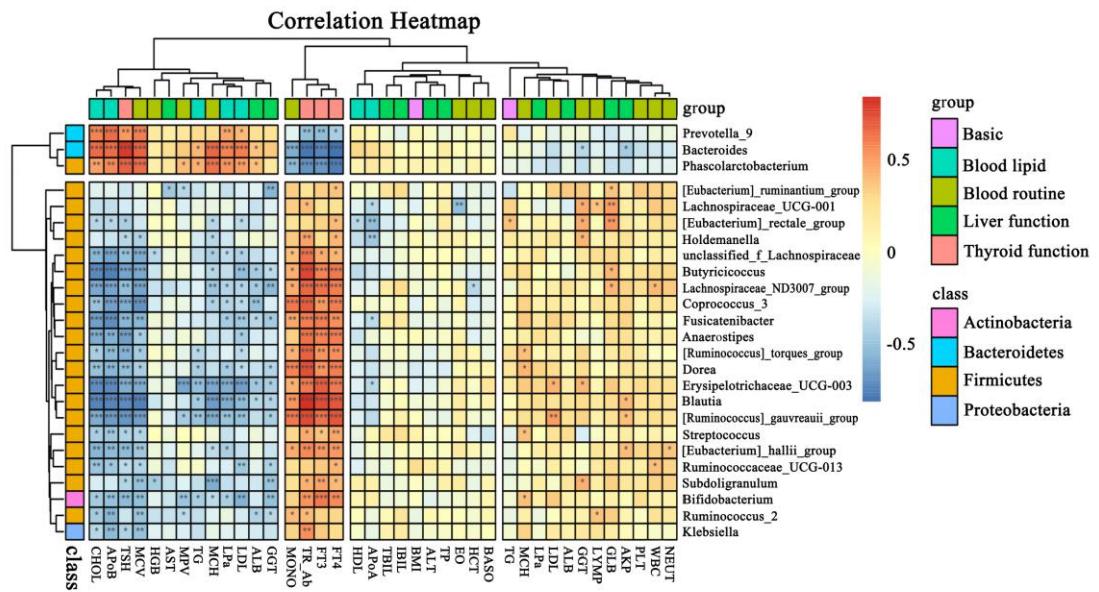
A. The significantly correlation of blood routine parameters and different metabolites was showing in the network diagrams ($P<0.05$, and $r> 0.5/-0.5$). In the network, triangle represents clinical parameters; circle represents metabolites and is classified according to color; the size of circle represents metabolite abundance; the red solid line represents negative correlation and the grass green solid line represents positive correlation; and the line length is meaningless. **B.** Heatmap diagram showed that three groups could be distinguished by model 2, and color intensity represents the magnitude of abundance.

Figure S13: The cross-analysis of prediction model between studies



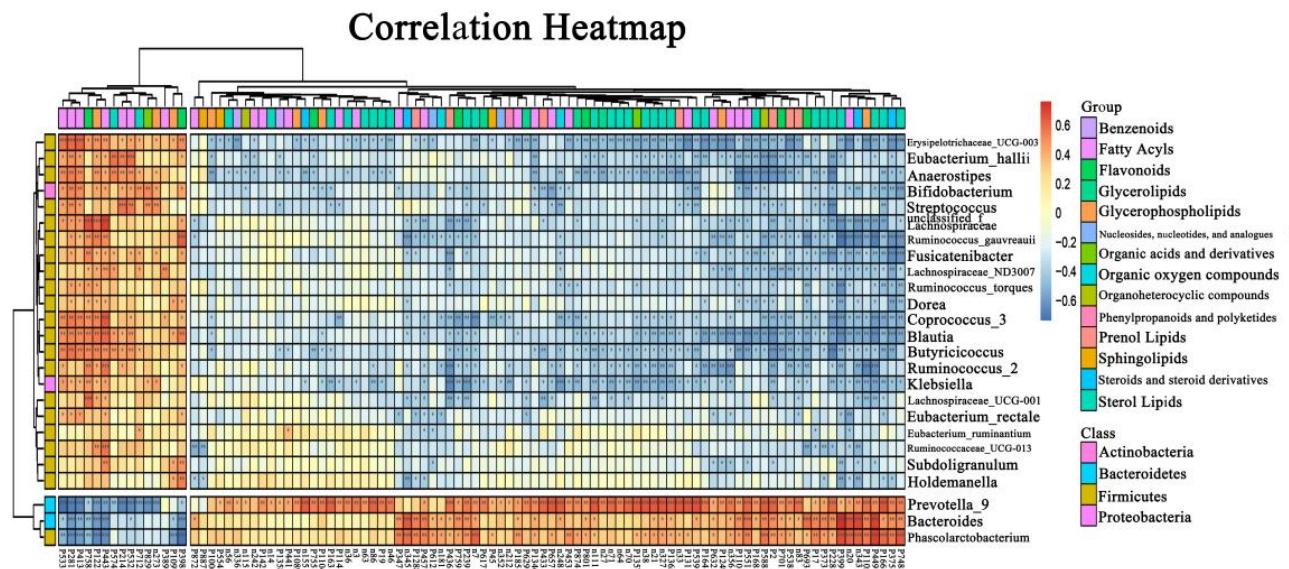
A. Model 2 was cross validated in TC patients; **B-D.** TC prediction model was cross validated in this study.

**Figure S14: Correlation of species and clinical parameters in paired samples
(Pri_GD vs. HC)**



In paired samples, the correlation of 25 different genera and 39 clinical parameters between Pri_GD and HC groups. In the heatmap, color intensity represents the magnitude of correlation, and marker represents the significantly correlation. Red, positive correlations ($r>0$); blue, negative correlations ($r<0$). * $P<0.05$; ** $P<0.01$; *** $P<0.001$.

Figure S15: The correlation analysis between 25 genera and 99 metabolites



In paired samples, the correlation of 25 different genera and 39 clinical parameters between Pri_GD and HC groups. In the heatmap, color intensity represents the magnitude of correlation, and marker represents the significantly correlation. Red, positive correlations ($r>0$); blue, negative correlations ($r<0$). * $P<0.05$; ** $P<0.01$; *** $P<0.001$.