Additional file 1 — additional_file_1.xlsx

Biomarkers discovery : differentially abundant taxa between the clinical groups of interest, as identified by a two-sided Wilcoxon test with Benjamini-Hochberg correction for multiple hypothesis testing.

Additional file 2 — additional_file_2.png

Number of biomarkers identified in Loomba-2017 and Qin-2014, relatively to their abundance and the sequencing depth.

Additional file 3 — additional_file_3.png

Effect of filtering strategies on the recovery of differences between patients groups in different studies: significance of inter-group difference regarding Shannon diversity index for α-diversity (A), PERMANOVA analysis for β-diversity (B). AUC corresponding to random forest classification (C) was performed in Loomba-2017 and Qin-2014, with a split between discovery and validation cohorts in Qin-2014 as performed on the original paper of this study. Error bars represent standard deviation for 10 repetitions of training process in the classification model.