



Figure 2. Predictive microbiome signatures of PDAC.

(A) Normalized abundance of 27 selected species in the fecal microbiome across samples shown as a heatmap. The right panel represents the contribution of each selected feature to the overall model-1 and robustness (the percentage of models in which the feature is included as predictor) of each feature is presented as percentage. Classification scores from cross validation of each individual and condition for tested meta-variables are displayed at the bottom of the panel, yellow representing the missing information. (B-D) Internal cross validation results of unconstrained model-1 (without feature selection), enrichment-constrained model-2 (constrained to positive features) and combination of CA19-9 (using a threshold of 37U/ml) with microbial features (see Methods) are shown as receiver operating characteristic (ROC) curve with a 95% confidence interval shaded in corresponding color. True positive rates (TPR) are given as a percentage at 90% specificity cutoff. External validation of all models on German (DE) PDAC population (N=76) is represented as well. Published CA19-9 is displayed from a meta-study as orange. The yellow dots represent observed CA19-9 levels which were only available for 33/50 CTRs and 44/57 PDAC patients in the Spanish (ES) and for 8/32 CTRs and 44/44 PDAC patients in the German (DE) population (D) TPR of all models at different PDAC progression stages and in addition, the FPR for CP cases and controls with 90% specificity cut-off are shown as barplots. Stages I and II and stages III and IV are combined due to the overall low sample size. The number of predicted cases compared to total is also shown on the top of each bar. ES: Spanish, DE: German, DE-Val: German Validation