

Supplementary Figures

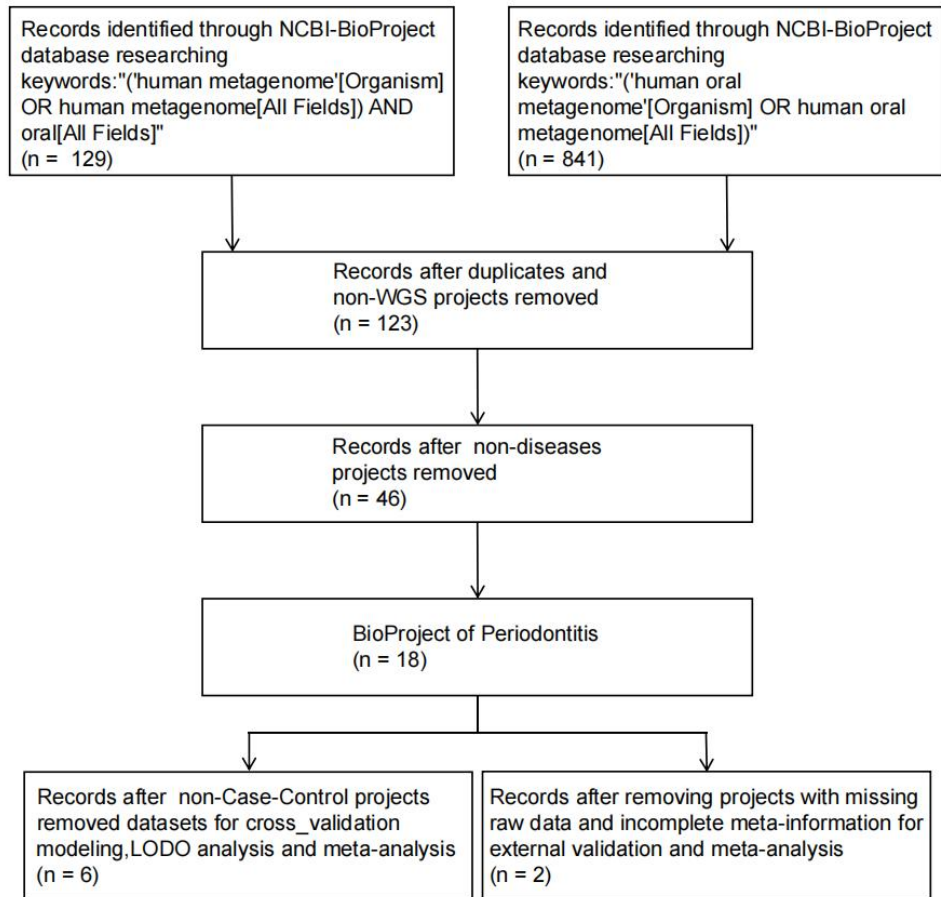
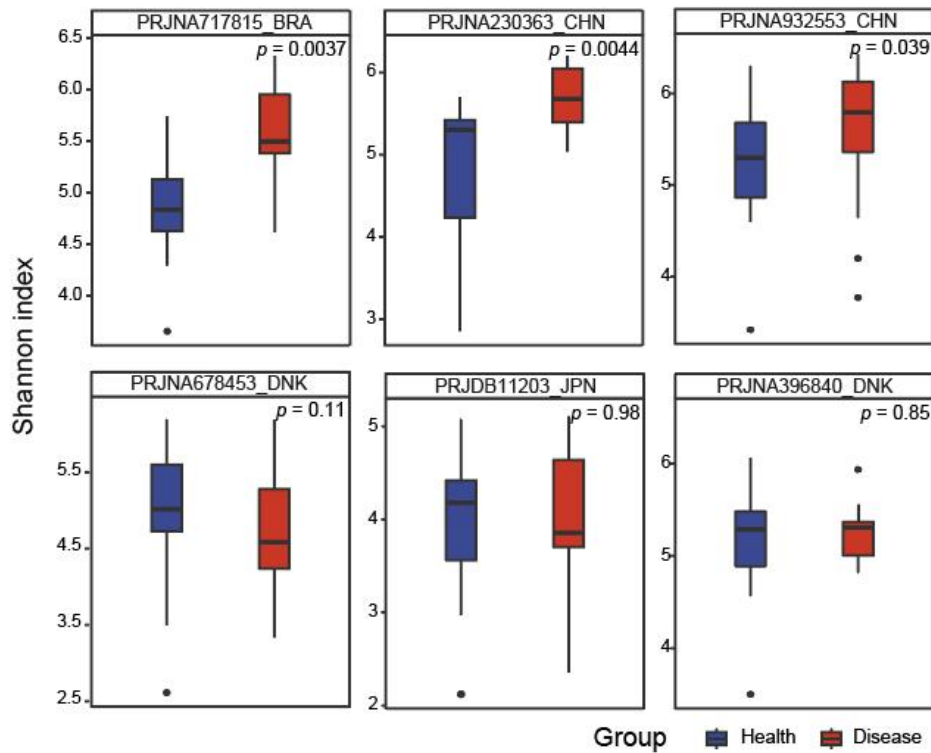


Figure S1. The pipeline for metagenomics data collection. We collected records about periodontitis in public databases until September 2023. The requirements for samples were that (i) at least 10 valid samples in each case group and control group, and (ii) there were no overlapping samples across datasets.

A



B

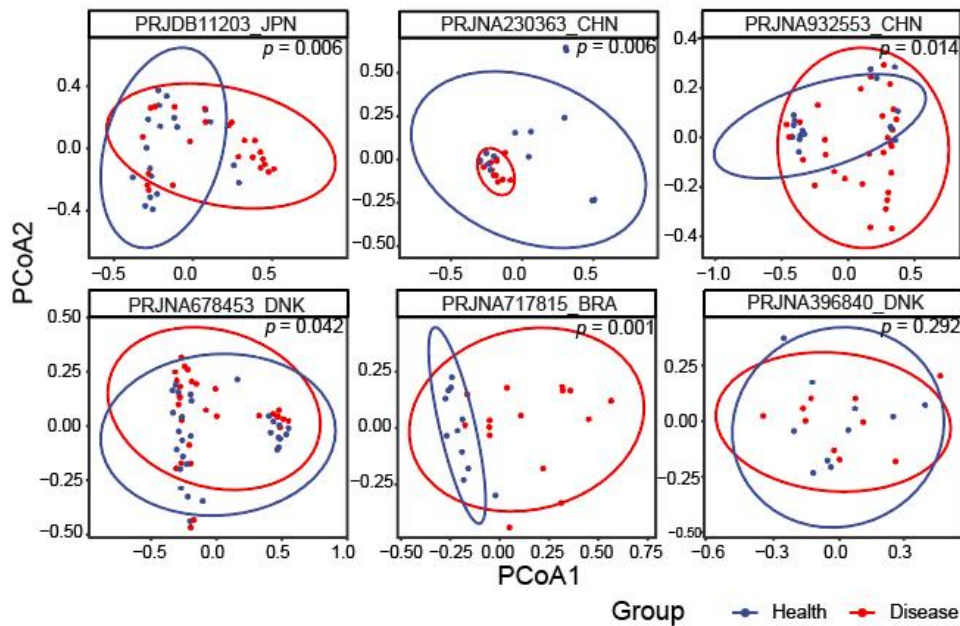


Figure S2. Species diversity was analyzed based on taxonomic relative abundance data. A, Within-cohort alpha-diversity (Shannon index) comparison between disease and healthy groups in the six case-control cohorts. Two-sided Wilcoxon rank-sum test was used for group-wise comparisons. **B,** Within-cohort beta-diversity analysis between disease and healthy groups using principal coordinates analysis (PCoA) in the six cohorts. P value was calculated with permutational multivariate analysis of variance (PERMANOVA) by 999 permutations (two-sided test). Each point is a sample. Each point is a sample.

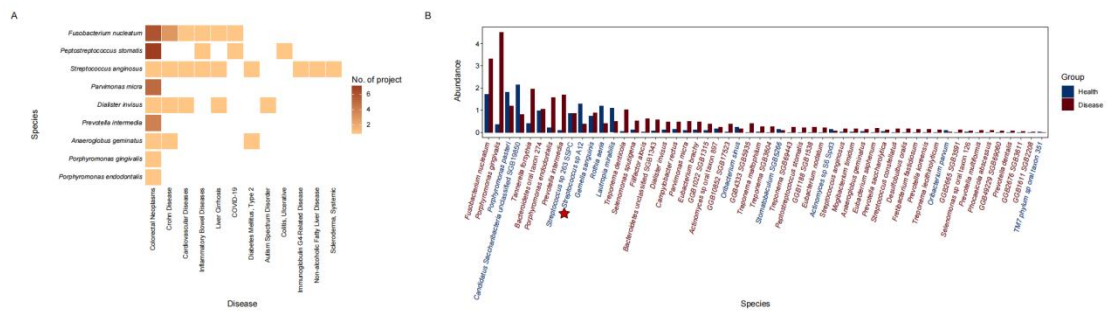


Figure S3. Disease type with biomarkers enriched as well as the abundance. A, Heatmap plot shows that the periodontitis oral biomarkers are also enriched in the gut microbiome of many diseases according to the GMrepo database. The color gradients indicate the number of projects of a particular disease in which the biomarkers are disease-enriched. **B,** The bar graph compares the abundance of marker species in healthy and diseased samples, ranked from highest to lowest total abundance. The red bars represent diseased samples, and the blue bars represent healthy samples. The red font indicates disease-enriched marker species, and the blue font indicates health-enriched species.

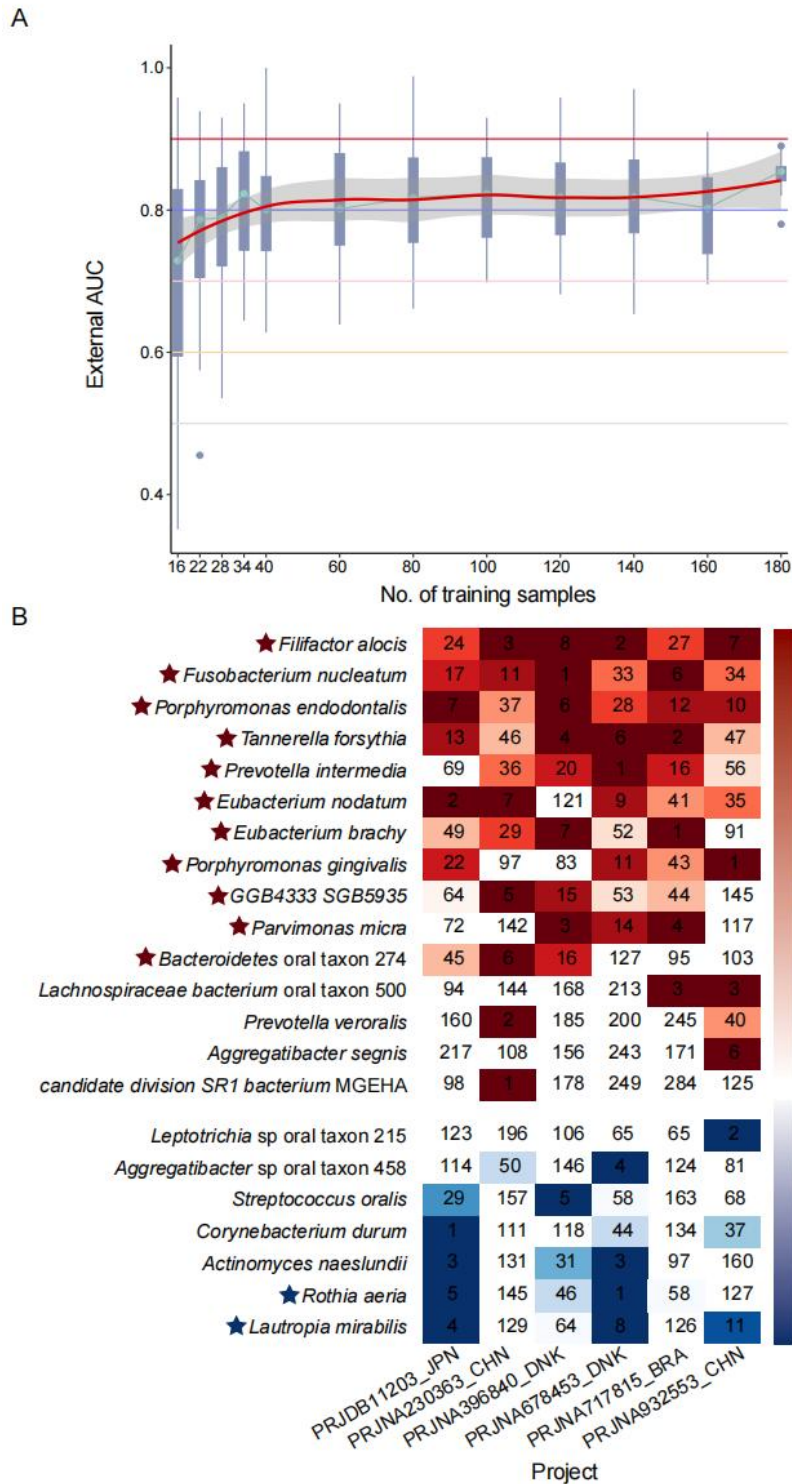


Figure S4. Trends in external AUC with number of training samples and heatmap of species ranked by RF classifiers. **A**, Box plot shows the trend of external AUCs as a function of the number of training samples, with the middle line representing the connection of the medians. **B**, Heatmap showing species ranked by the RF classifiers. Top five features ranked by each of the six intra-cohort classifiers were selected. Case- and control-enriched features were ranked independently. Color gradient represents the mean rank of each feature. Red and blue stars represent marker species that are

enriched in case and control groups according to the LefSe analysis, respectively.

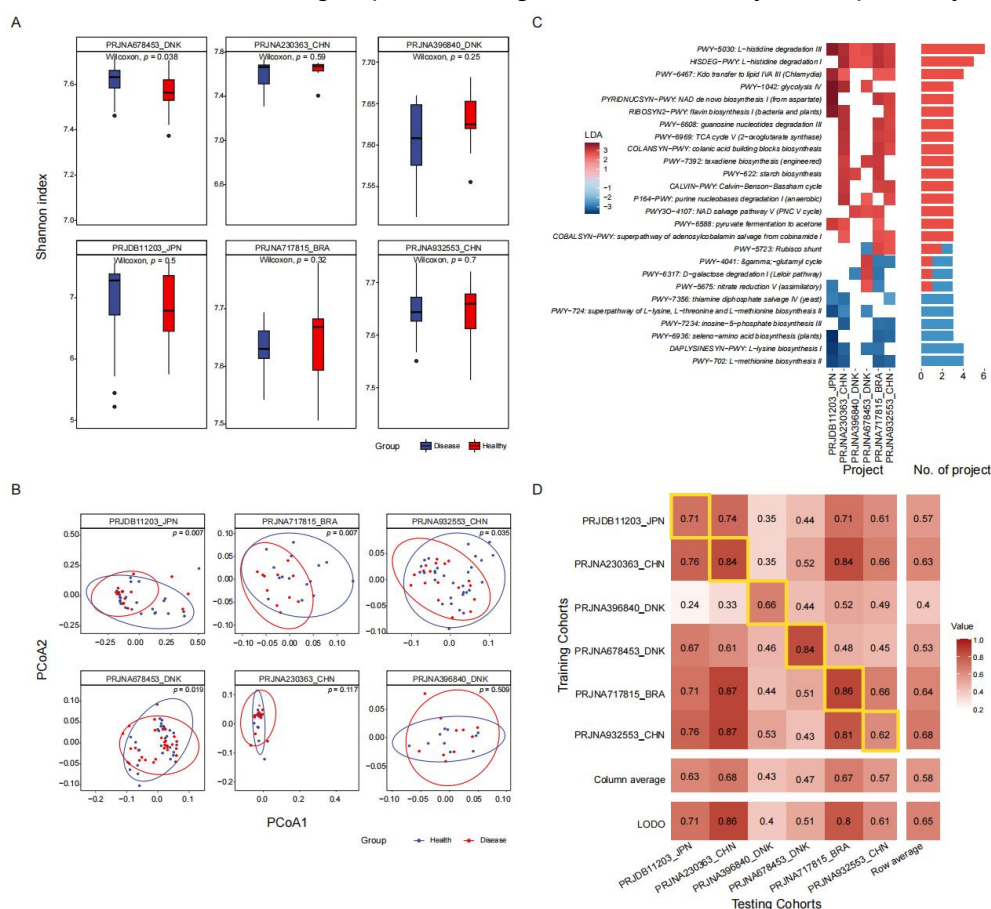


Figure S5. Function-based diversity analysis, identification of cross-cohort pathway biomarkers and modeling of machine learning classifiers for periodontitis.

A, Boxplots show alpha- diversity (Shannon index) of the six case-control cohorts using the functional profile. Significance was determined using the Wilcoxon rank-sum test (two-sided). **B**, Within-cohort principal-component analysis (PCA) between the case and control groups in the six cohorts. P value was calculated with PERMANOVA by 999 permutations (two-sided test). **C**, Heatmap-bar plot showing the metabolic pathway biomarkers that showed significantly differential abundances between case and control groups in ≥ 3 cohorts. The functional biomarkers were identified using LefSe on HUMAnN3 pathway profiles; a linear discriminant analysis (LDA) score cutoff of 2 was used to select the biomarkers. The red and blue blocks indicate case- and control enriched biomarkers, respectively. **D**, Cross prediction matrix reporting prediction performances as AUC values obtained using Random Forest (RF) classifiers on pathways-level relative abundances. The values boxed in yellow squares on the diagonal are the AUC values obtained by training and validating within individual cohorts. The non-diagonal values refer to training a classifier on the dataset corresponding to the row and applying it to the dataset corresponding to the column to obtain the AUC value. The "Row average" and "Column average" are the average AUCs of the corresponding rows and columns, excluding the diagonal values. The Leave-One-Dataset-Out (LODO) row refer to the performances obtained by training the model on the pathways-level abundances using all but the dataset of the corresponding column and applying it on the

dataset of the corresponding column. Color gradient represents the value of AUC.

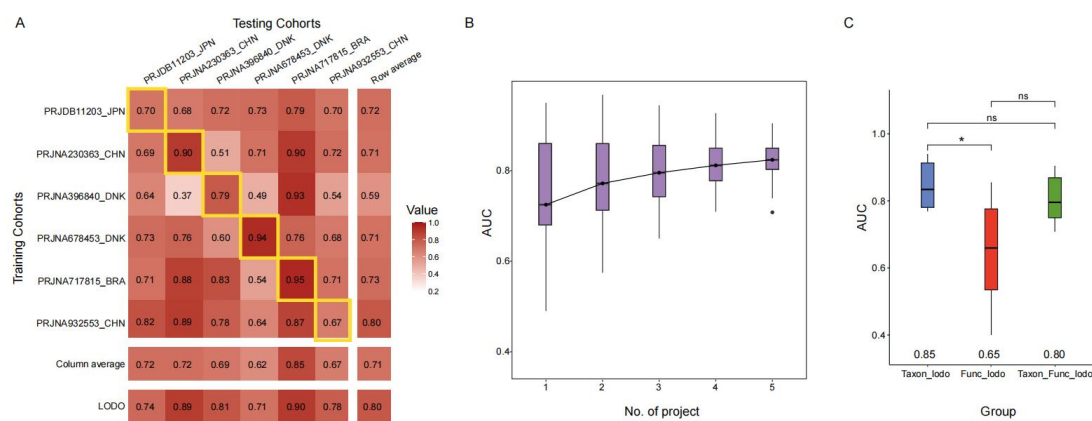


Figure S6. Prediction performances of the models based on the combined taxonomic-functional profiles in within-cohort validation and cross-cohort testing.

A, Cross prediction matrix reporting prediction performances as AUC values obtained using a Random Forest (RF) model on species-pathways combined relative abundances. The values boxed in yellow squares on the diagonal are the AUC values obtained by training and validating within the queue. The non-diagonal values refer to training a classifier on the dataset corresponding to the row and applying it to the dataset corresponding to the column to obtain the AUC value. The row average and column average are the average values of excluding the diagonal values. The Leave-One-Dataset-Out (LODO) row refer to the performances obtained by training the model on the species-pathways combined abundances using all but the dataset of the corresponding column and applying it on the dataset of the corresponding column. Color gradient represents the value of AUC. **B**, Box plot showing the external AUC as the function of increasing number of cohorts combined for model training. **C**, Box plot shows the comparison of AUC values obtained from LODO analysis of taxonomic profiles, functional profiles, and combined taxonomic-functional profiles. Two sides Wilcoxon rank sum test was used for comparisons. The numbers at the bottom indicate the average AUC values corresponding to each model. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

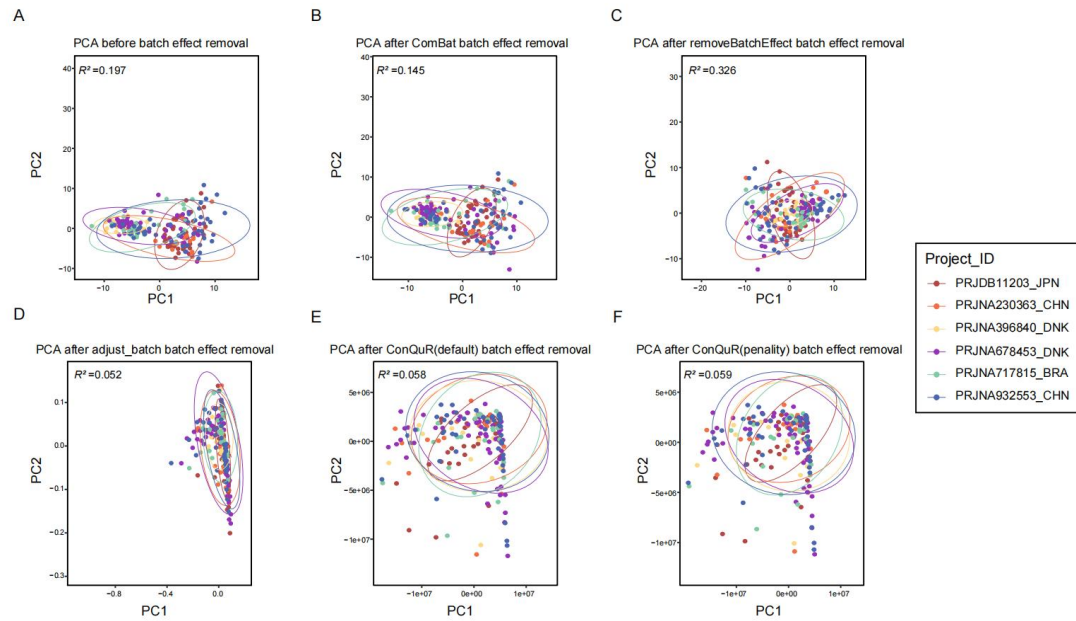


Figure S7. Evaluation of batch-effect removal tools on the six case-control projects. Principal coordinates analysis (PCoA) plots showing the sample distribution before and after batch-effect removal using state-of-art tools, with the tool names indicated in the panel titles. A, sample distribution before any batch-effect removal tools were applied. **B-F,** sample distributions after batch-effect removal tools were applied, including (B) “ComBat” function in the “sva” package, (C) “removeBatchEffect” in the “limma” package, (D) “adjust_batch” in the “MMUPHin” package, and I “ConQuR” in the “ConQuR” package with “default” parameter, and (F) “ConQuR” in the “ConQuR” package with “penalty” parameter.

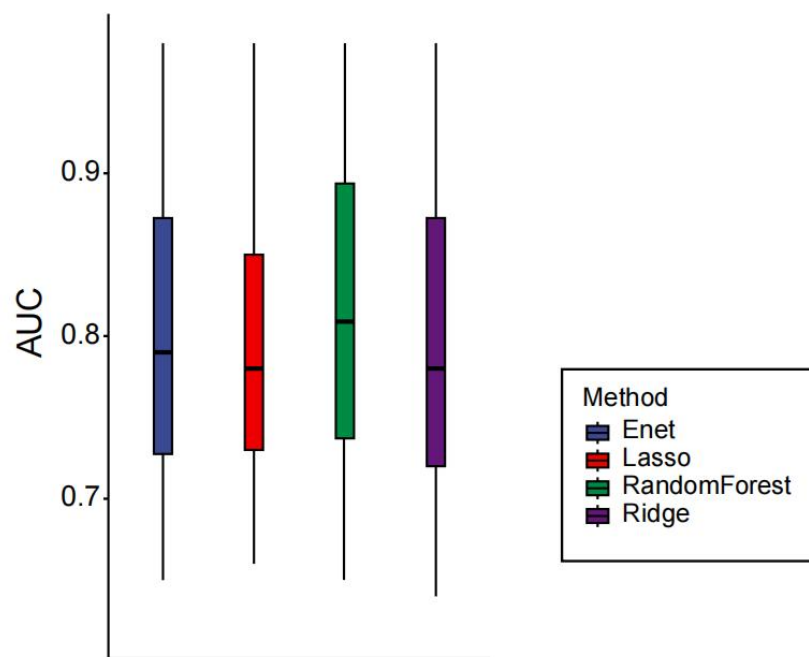


Figure S8. Boxplot showing the performance of four machine learning algorithms in cross-validation.