Statistical analyses

We utilized the spls() function from the mixOmics package to carry out sparse partial least squares (sPLS) regression. sPLS regression involves integrating two matrices and modeling their structures. In this case, we employed variance-stabilized ASV abundances as the explanatory variables and continuous clinical parameters (TG, CHOL, HDL, LDL, VLDL, NEUper, WBC, NEU, LYM, MONO, HGB, PLT, UREA, UA, CR, ALT, AST, TP, ALB, TBIL, DBIL, and IBIL) as the response variables. This method allows for multiple response variables. To avoid multicollinearity in the sPLS analysis, we utilized the prcomp() function to conduct principal component analysis (PCA). No limitations were imposed on the number of response variables to be retained for each component (keepY) before model calculation. On the other hand, we set the number of explanatory variables (ASVs) to be retained for each component (keepX) at 25. To determine the relevant components, we utilized the perf() function. Subsequently, hierarchical clustering was carried out using the cim() function, employing the "complete linkage" clustering method and "Pearson's correlation" distance method. This process was performed based on the sPLS regression models for each body site. Consequently, we created coefficient matrices that reflected the correlations between ASV abundances and continuous clinical parameters. We then proceeded with canonical correspondence analysis (CCpnA), a multivariate constrained ordination method. This allowed us to evaluate the associations of both categorical and continuous clinical parameters with ASV abundances. In the CCpnA analysis, we included ASVs and variables that exhibited a correlation value greater than 0.2 or less than -0.2 (in the

oral and rectal data set) as identified in the sPLS analysis. The implemented method utilized the cca() function from the vegan package. This function performed a chi-square transformation of the log+1-transformed ASV count matrix. Subsequently, it conducted weighted linear regression and singular value decomposition. The results of CCpnA were visualized as a triplot, where the length of each axis reflected the percentage of variance explained by that axis.

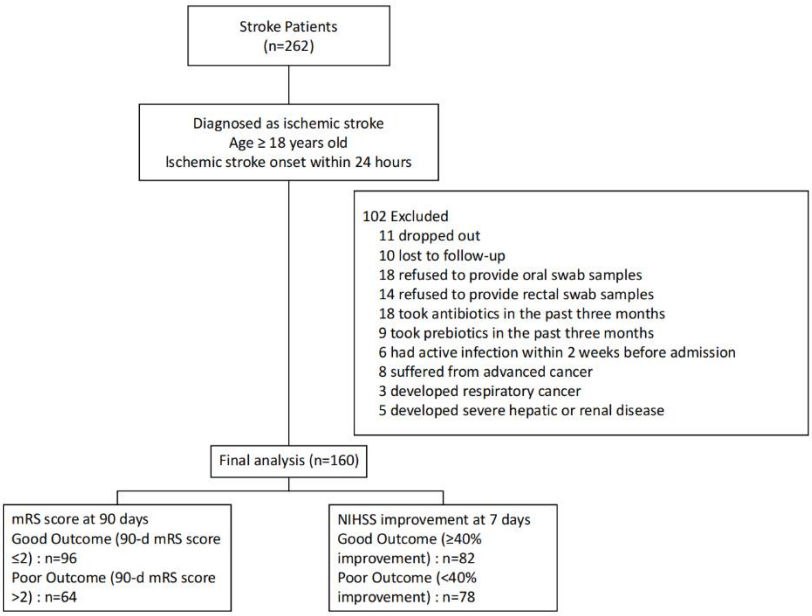

Figure S1. Flow diagram of the process for selecting patients.

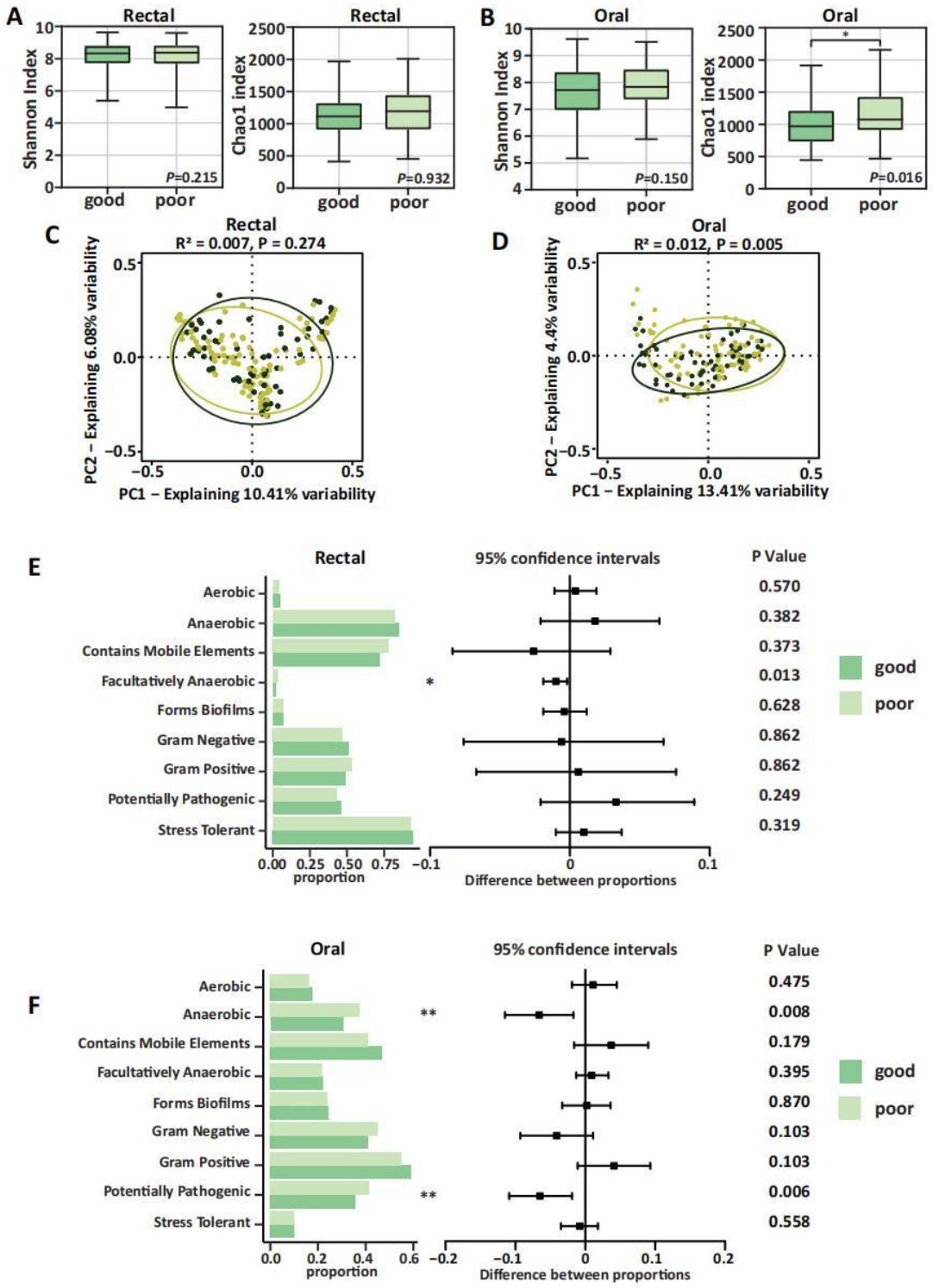

Figure S2. Gut and oral microbial community structures were analyzed based on 16S rRNA gene sequencing according to long-term prognosis. The α diversity (Shannon index) and richness (Chao1 index) in rectal swabs (A) and oral swabs(B). PCoA plots based on Bray–Curtis distances among the two groups in rectal swabs (C) and oral swabs(D). *PCoA* principal coordinate analysis. Phenotypic prediction by BugBase according to long-term prognosis. Bacterial compositions in the rectal samples (E) and

oral samples (F) predicted by BugBase.

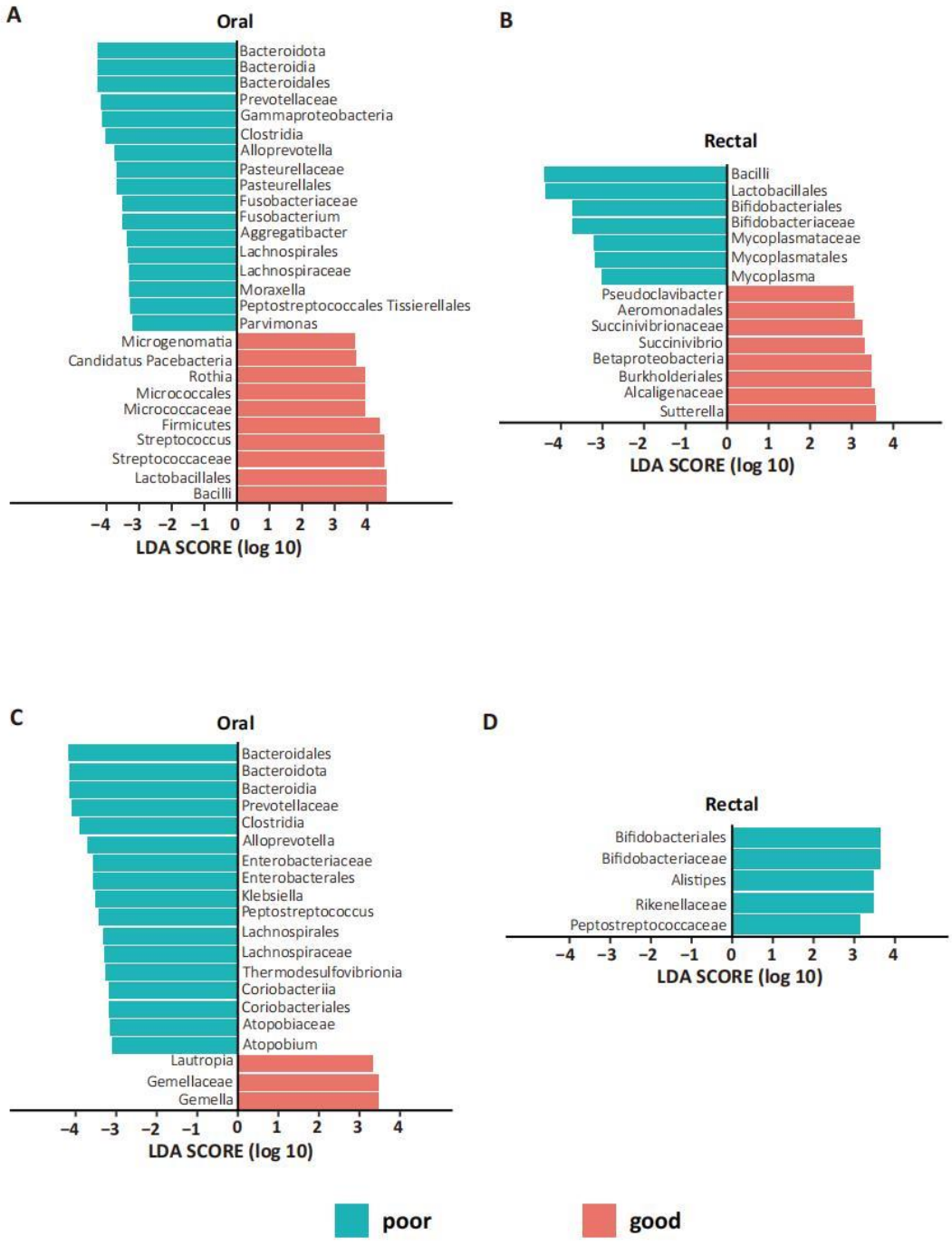

Figure S3. The LEfSe analysis indicates that the representation of various oral bacterial taxa (A) and gut bacterial taxa (B) differed between groups with good (red) and poor (green) outcomes based on short-term prognosis. Additionally, the analysis reveals that the representation of oral bacterial taxa (C) and gut bacterial taxa (D) varied between

groups with good (red) and poor (green) outcomes according to long-term prognosis.

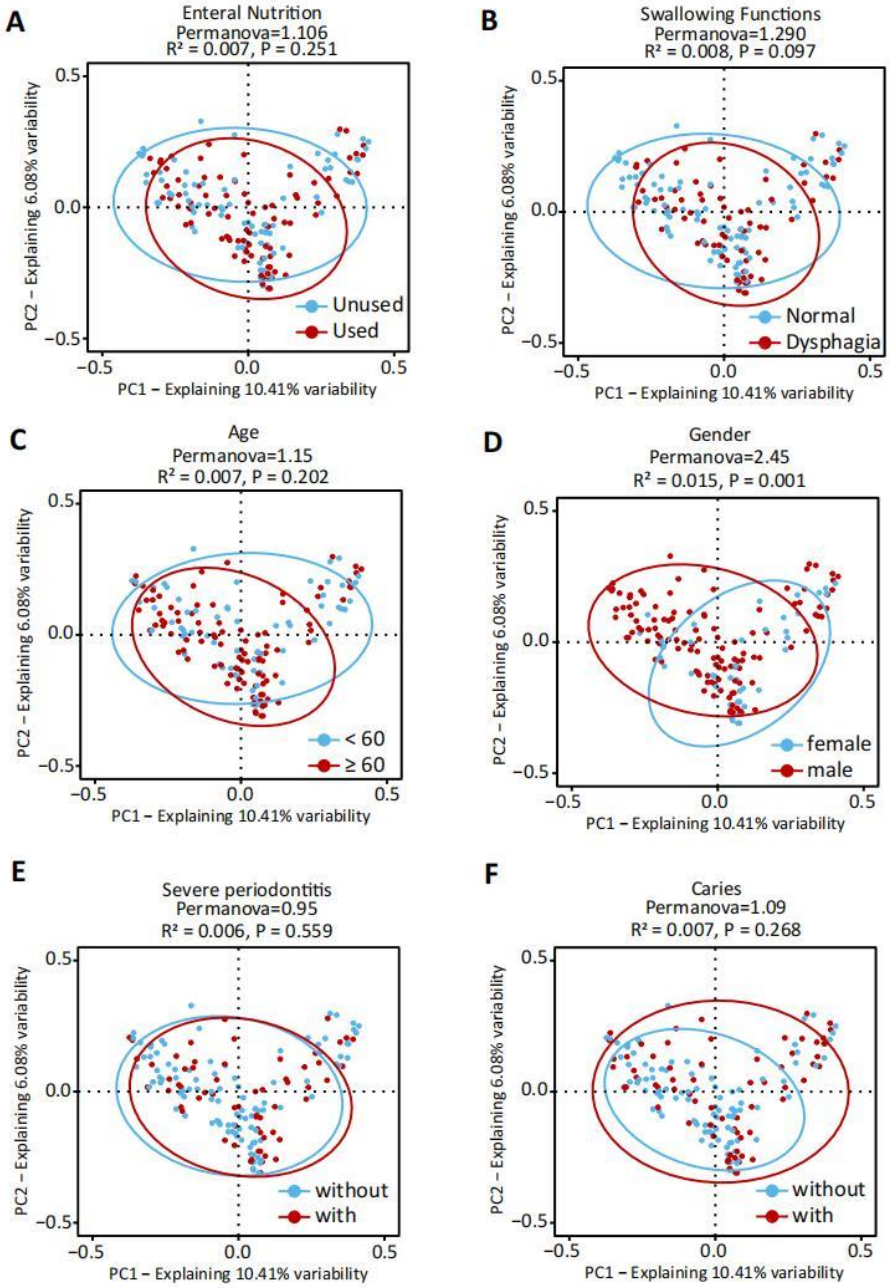

Figure S4 Influence of clinical variables on rectal microbiota after stroke. Explained variance (R2) for each variable was estimated using PERMANOVA. PERMANOVA results illustrate changes in microbial communities based on major variables: (A) Influence of enteral nutrition on rectal microbiota in AIS patients. (B) Influence of dysphagia on rectal microbiota in AIS patients. (C) Influence of age on rectal microbiota in AIS patients. (D) Influence of gender on rectal microbiota in AIS patients. (E) Influence of severe periodontitis on rectal microbiota in AIS patients. (F) Influence of caries on rectal microbiota in AIS patients.

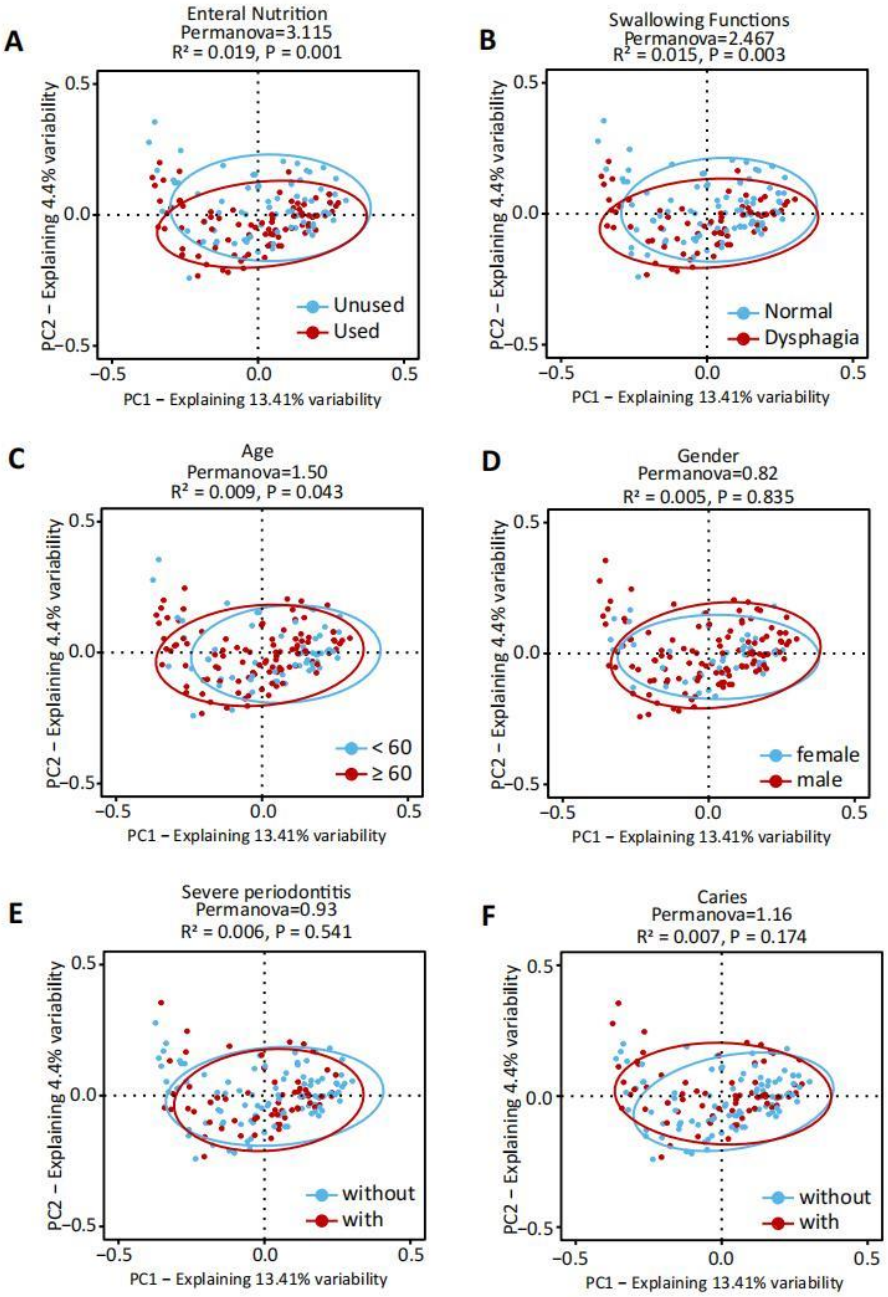

Figure S5. Influence of clinical variables on oral microbiota after stroke. Explained variance (R2) for each variable was estimated using PERMANOVA. PERMANOVA results illustrate changes in microbial communities based on major variables: (A) Influence of enteral nutrition on oral microbiota in AIS patients. (B) Influence of dysphagia on oral microbiota in AIS patients. (C) Influence of age on oral microbiota in AIS patients. (D) Influence of gender on oral microbiota in AIS patients. (E) Influence of severe periodontitis on oral microbiota in AIS patients. (F) Influence of caries on oral microbiota in AIS patients.

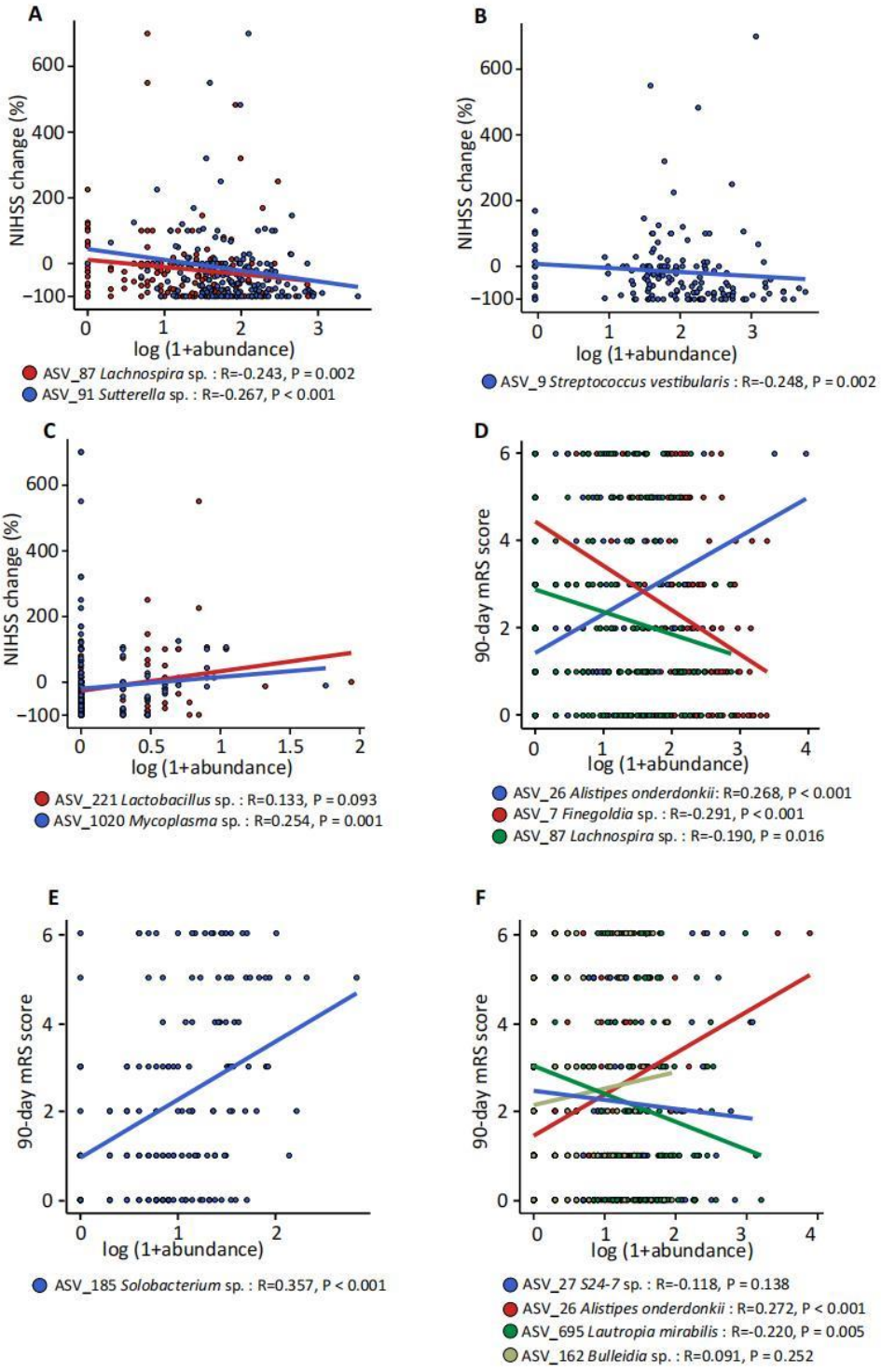

Figure S6. Correlation between predictive ASVs and clinical scores. (A-C) Spearman correlation analysis was used to examine the linear relationship between ASVs based on short-term prognosis with the percentage of NIHSS change. (D-F) Spearman correlation analysis was used to examine the linear relationship between ASVs based on long-term prognosis with mRS Score.

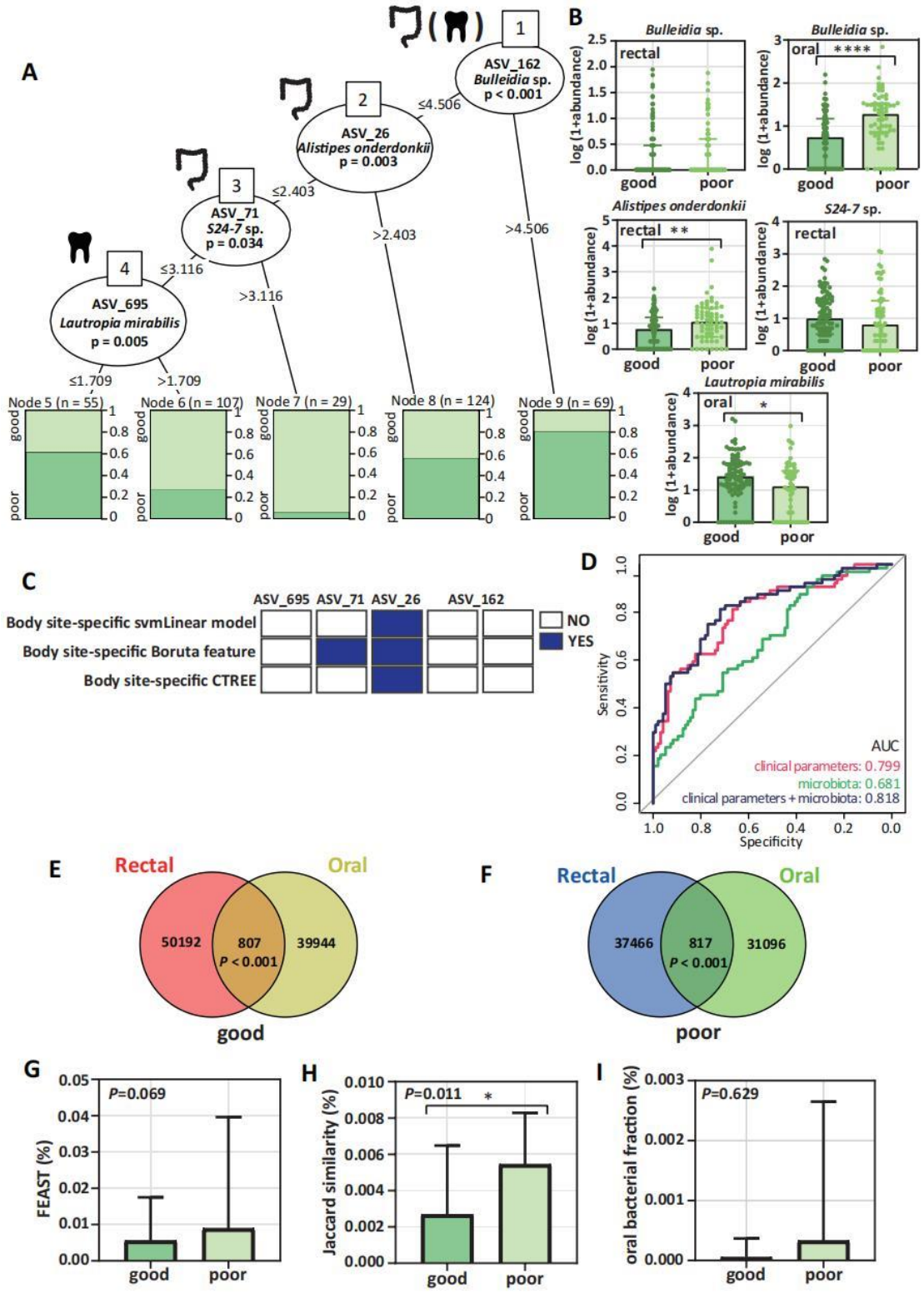

Figure S7. Combined machine learning-based prediction of long-term prognosis from microbiota composition of two body sites (rectal, mouth). (A) A Conditional Inference Tree (CTREE) is displayed to show ASVs identified as significant split nodes by nonparametric regression for the prediction of clinical outcomes. The numbers along the branches indicate split values of variance stabilized bacterial abundances. The terminal nodes present the proportion of samples originating from patients (n = number

of represented samples) with good versus poor outcomes. The icons indicate the body sites where the ASVs were detected. (B) The bar chart depicts the log-transformed relative abundances of the predictive ASVs in patients with good and poor functional outcomes. (C) Comparison of the combined machine learning model with the individual body site-specific models reveals the results obtained from the three body site-specific analyses: svmLinear model, Boruta feature selection, and regression framework with CTREE. The taxa identified in the combined machine learning model are highlighted. If a taxon is indicated in blue, it suggests agreement among the models, whereas white indicates that the ASV was not detected as significant in a specific analysis. (D) Predictive models for early differentiation of long-term poor outcomes: ROC curves of clinical parameters, including age, sex, initial NIHSS score, enteral nutrition, atrial fibrillation, dysphagia, and NLR. Also, ROC curves for predictive ASVs from oral and rectal swabs, and ROC curves for clinical parameters combined with microbiota. The model incorporating clinical parameters and microbiota includes age, sex, initial NIHSS score, enteral nutrition, atrial fibrillation, dysphagia, NLR, and predictive ASVs. Comparison of oral vs. rectal microbiota in patients among two groups according to long -term prognosis. (E-F) The Venn diagram shows the shared bacteria of oral and rectal samples. (G) The FEAST analysis shows oral-derived microbiota in the rectal samples. The numerical value of the vertical axis indicates the percentage of the two groups that had oral-derived bacteria in their gut microbiota. (H) Jaccard similarity analysis reveals the degree of similarity between concurrent rectal and oral samples. The greater the value, the greater the similarity between the two samples. (I) Distribution of oral bacterial fraction in rectal samples. FDR-corrected P values were calculated using a two-sided Mann-Whitney U test.

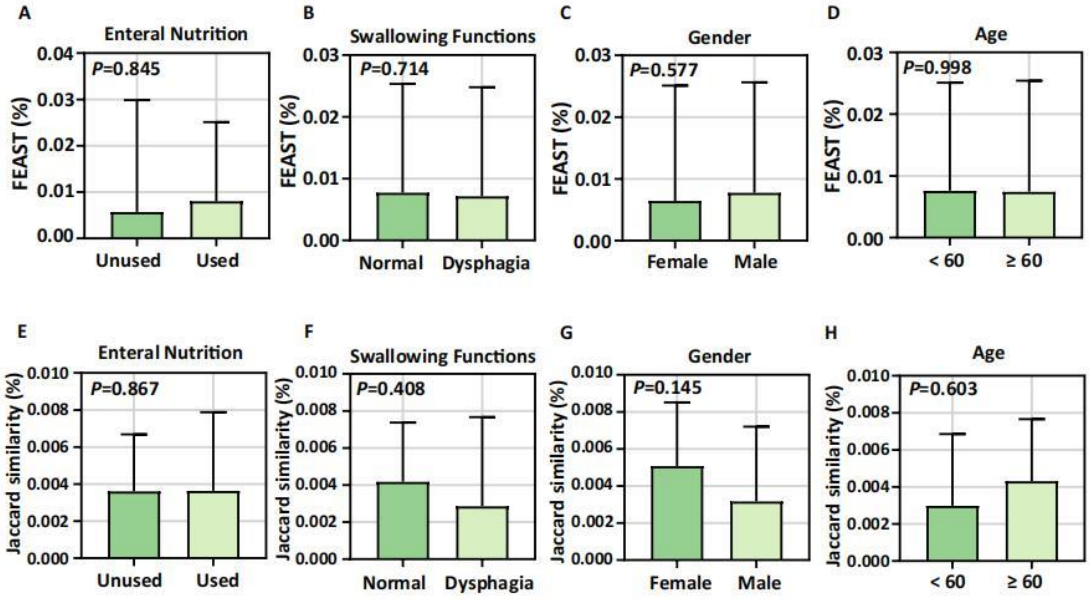

Figure S8. Comparison of the effects of enteral nutrition, dysphagia, age, gender, severe periodontitis, and caries on oral bacterial colonization in the gut.

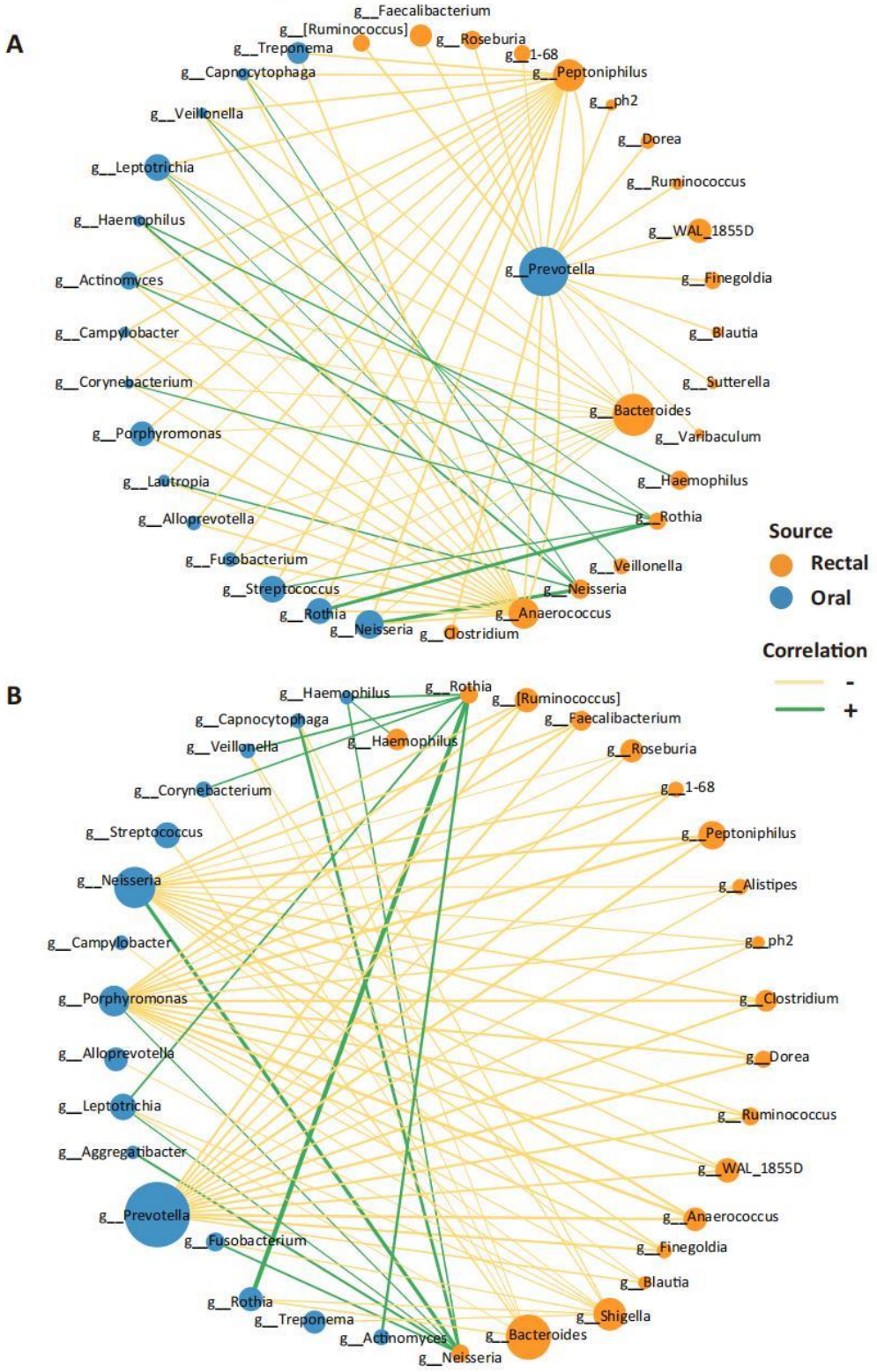

Figure S9. Co-occurrence network for oral and gut microbiota among two groups according to long-term prognosis. We have selected bacteria with a Spearman's correlation of ≥0.8 and a p-value of ≤0.05 in order to construct network diagrams of the oral and intestinal microbiota in groups with good (A) and poor (B) outcomes. In these diagrams, the orange circles represent the rectal microbiota, while the blue circles

represent the oral microbiota. The size of each circle corresponds to the abundance of the flora, with larger circles indicating higher abundance. The green lines represent positive correlations, and the yellow lines represent negative correlations. The thickness of the lines reflects the strength of the correlation coefficient.

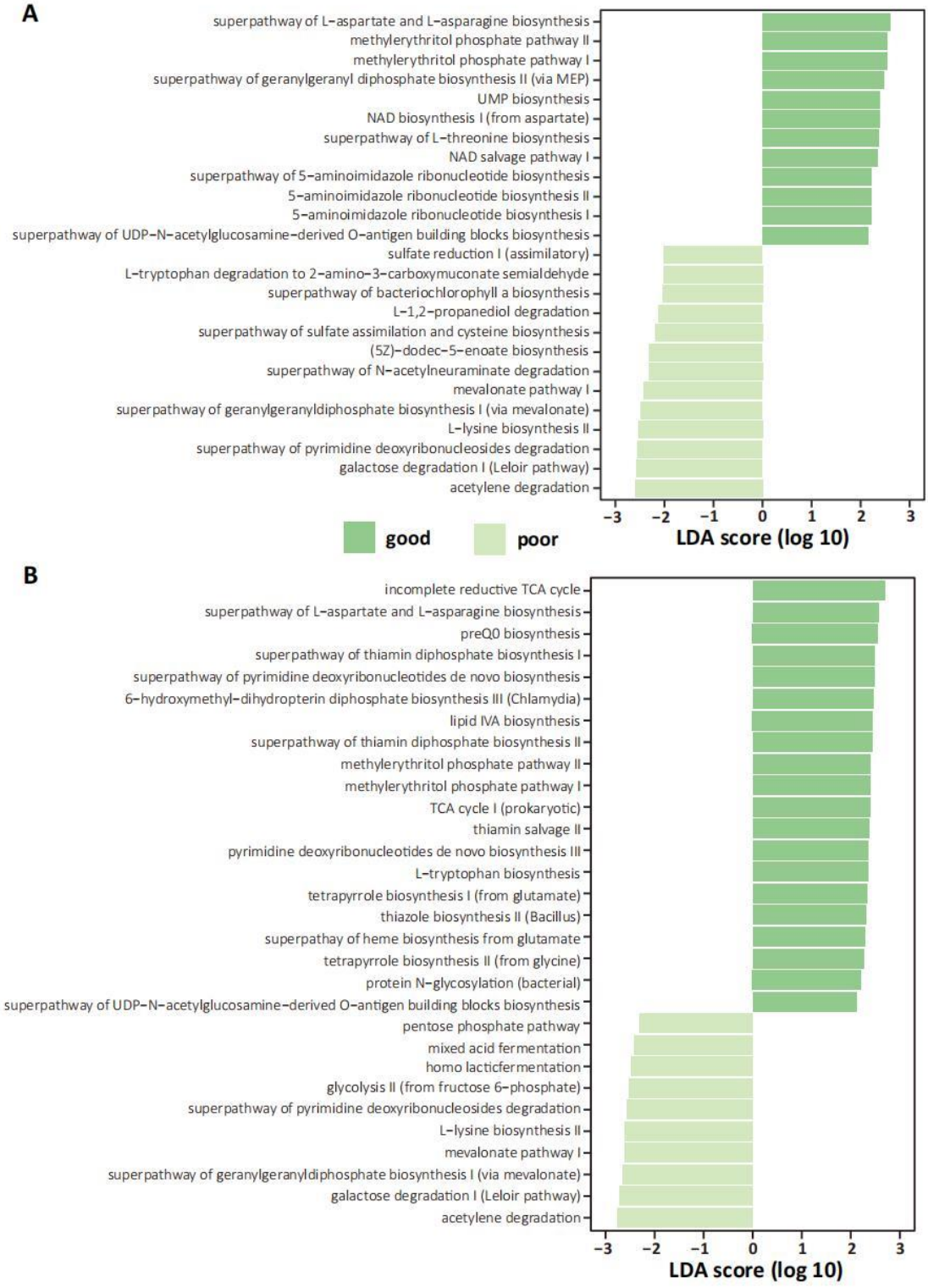

Figure S10. (A) KEGG analysis results of rectal microbiota comparing short-term good outcomes group with poor outcomes group. (B) KEGG analysis results of rectal

microbiota comparing long-term good outcomes group with poor outcomes group. LDA score threshold = 2.

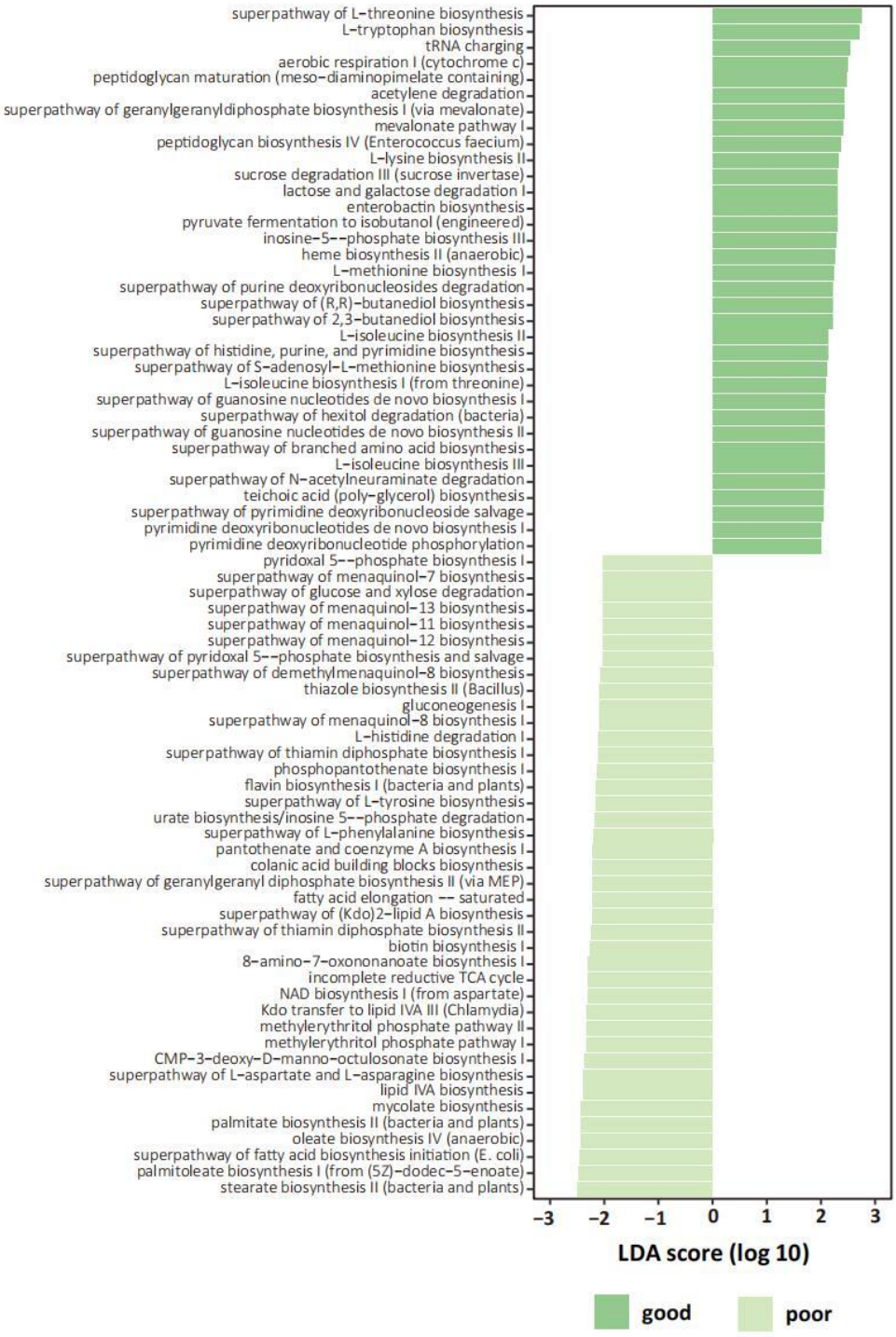

Figure S11. KEGG analysis results of oral microbiota comparing short-term good outcomes group with poor outcomes group. LDA score threshold = 2.

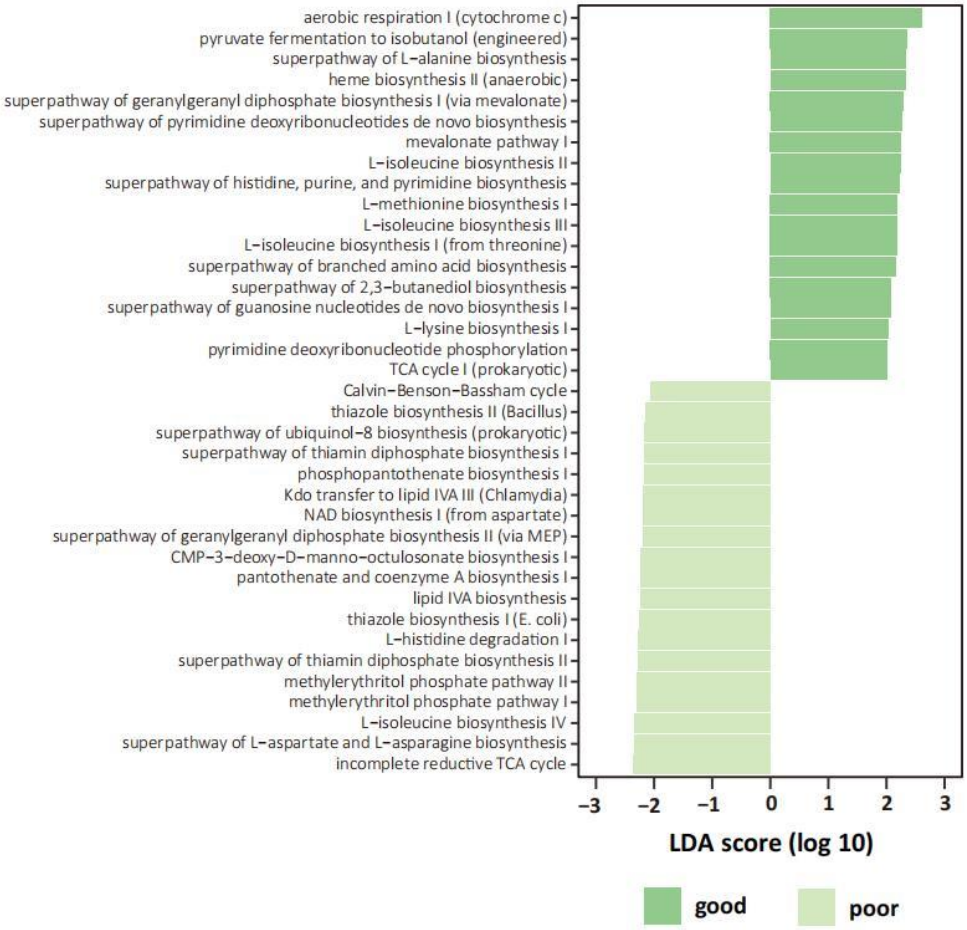

Figure S12. KEGG analysis results of oral microbiota comparing long-term good outcomes group with poor outcomes group. LDA score threshold = 2.