Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Oral Multi-niche Microbiota and Age Stratification: A Novel Strategy for Precision Screening of Periodontitis

Yinan Chen
Fudan University
Enhua Mei
Fudan University
Wei Zhang
Fudan University
Yan Wang
Fudan University
Yuehua Liu
liuyuehua@fudan.edu.cn

Fudan University

Research Article

Keywords: Periodontitis, Oral microbiota, Age stratification, Non-invasive screening

Posted Date: May 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9227962/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Oral Health on August 24th, 2026. See the published version at https://doi.org/10.1186/s12903-026-09652-1.

Abstract

Background: periodontitis is a prevalent chronic inflammatory disease driven by oral microbiota dysbiosis, with profound oral and systemic impacts that impair healthy aging. Conventional screening methods are invasive and delay early detection, while microbiota-based strategies show promise for non-invasive screening—yet the roles of oral niche heterogeneity and age-related microbial shifts remain unclear.

Methods: a total of 120 participants (20–90 years) were classified into low-risk (non-periodontal, NP, n=67) and high-risk (periodontal, P, n=53) groups based on clinical and radiological indices. Non-invasive samples (saliva, dental plaque, tongue plaque) were collected for 16S rRNA gene sequencing. Bioinformatic analyses and machine learning models were used to evaluate diagnostic efficacy, with age stratification (≤50 years vs. >50 years).

Results: age differed significantly between NP and P groups (p=0.003). Age stratification improved model accuracy, with salivary microbiota showing optimal diagnostic efficacy (AUC=0.807), outperforming dental plaque (max AUC=0.780) and tongue plaque (AUC=0.753). Distinct site-specific biomarkers were identified, and beta-diversity analysis revealed significant microbial community differences across sites (PERMANOVA, p=0.001).

Conclusions: oral niche heterogeneity and age stratification are critical for microbiota-based periodontitis screening. Saliva is the optimal non-invasive sample, providing robust evidence for developing precise, accessible diagnostic tools for community and primary healthcare settings.

Background

Periodontitis is a highly prevalent chronic inflammatory disorder targeting periodontal tissues, primarily driven by oral microbiota dysbiosis[1]. The oral cavity harbors a complex, dynamically evolving microbial community; upon dysregulation, pathogenic bacteria (e.g., *Porphyromonas gingivalis*, *Tannerella forsythia*) outcompete commensal bacteria to establish dominance[2]. This shift elicits persistent, dysregulated host immune-inflammatory responses that progressively impair gingival tissues, the periodontal ligament, and alveolar bone, ultimately leading to tooth loss if left untreated. Epidemiological evidence indicates that the global prevalence of moderate-to-severe periodontitis among adults ranges from approximately 20% to 50%, with a disproportionately heavier disease burden in the elderly[3]. Beyond compromising oral health, growing evidence links periodontitis to an increased risk of various systemic diseases, including cardiovascular diseases, type 2 diabetes mellitus, and adverse pregnancy outcomes —highlighting its substantial clinical and public health impact[4]. Given the irreversibility of periodontal tissue damage and its systemic sequelae, early screening and prompt intervention are pivotal to halting disease progression and improving patient prognosis.

Traditionally, periodontitis screening relies on clinical parameters, including periodontal probing depth (PD), clinical attachment loss (CAL), bleeding on probing (BOP), and visual assessment of gingival

inflammation—collectively regarded as the "gold standard" for evaluating periodontal status[5]. Nevertheless, these conventional methods have inherent limitations: invasiveness, reliance on well-trained clinicians (rendering them susceptible to inter-examiner variability), and limited accessibility in primary care or community settings[6]. More critically, these metrics primarily reflect established tissue damage rather than early-stage microbial dysbiosis, leading to delayed disease detection and intervention[7]. To address these drawbacks, microbiota-based diagnostic strategies have gained increasing traction in recent years[8]. Leveraging the tight association between oral microbial community perturbations and periodontitis pathogenesis, these approaches enable early identification of disease susceptibility prior to the onset of overt clinical manifestations.

The oral microbiota is characterized by pronounced spatial heterogeneity and age-associated dynamic shifts. Different oral niches (e.g., saliva, dental plaque, tongue plaque) harbor unique microenvironmental conditions (e.g., oxygen tension, pH, nutrient availability), which shape niche-specific microbial communities with distinct compositional and functional profiles[9]. Additionally, aging drives profound alterations in the oral microbiota, typified by reduced diversity, enrichment of pathogenic taxa, and depletion of beneficial commensals—modulating an individual's susceptibility to periodontitis[10]. Currently, most microbiota-based periodontitis diagnostic studies focus on subgingival plaque and gingival crevicular fluid (GCF), the primary colonization site of periodontal pathogens[11]. However, subgingival plaque sampling is cumbersome, invasive, and requires professional clinical expertise, severely limiting its widespread application in community and primary healthcare settings. In contrast, saliva, dental plaque, and tongue plaque are obtainable non-invasively and conveniently, making them ideal candidates for routine diagnostics. Yet, their diagnostic efficacy for periodontitis remains incompletely elucidated.

To address this knowledge gap, the present study recruited 120 participants across different age groups, collecting microbial samples from the three aforementioned easily accessible niches (saliva, dental plaque, tongue plaque). Microbial community composition was analyzed via 16S rRNA gene sequencing, coupled with comprehensive periodontal clinical assessments. Based on sequencing data and clinical findings, multiple machine learning models were constructed to evaluate the diagnostic efficacy of microbiota from each niche. Results demonstrated that stratifying participants into two age groups (≤50 years and >50 years) substantially improved the diagnostic accuracy of the models, with salivary microbiota exhibiting the optimal performance (AUC=0.807). This study confirms the necessity of accounting for niche heterogeneity and age stratification in microbiota-based periodontitis screening, thereby providing robust data support and novel insights for the development of non-invasive, precise diagnostic tools for periodontitis.

Methods

Participants

This study is a cohort study aiming to observe oral phenotypic multimodal characteristics and the risk of aging-related oral-maxillofacial diseases (Project No.: 2023YFC3605601 and 2023SHZDZX02D05). This study was approved by the Ethics Committee of Shanghai Stomatological Hospital (Ethics Approval Nos.: [2023]019 and [2024]021). Eligible participants were recruited from communities and tertiary oral specialist hospitals in Shanghai, with enrollment conducted from December 2024 to March 2025. Eligible participants were adults aged 20–90 years, who were conscious, willing to cooperate with the study, and free of major systemic diseases. The following individuals were excluded: (1) those with complete edentulism; (2) those with major underlying diseases; (3) those who were receiving, had received periodontal treatment within the past 3 months, or planned to receive periodontal treatment; (4) those who had taken antibiotics for a long-term period within the past 6 months or for a short-term period within the past 1 month; (5) those who were considered by the researchers to have factors that might affect the study results.

A total of 120 participants were enrolled in the study. Approximately 40 volunteers were evenly recruited in each of the three age groups: 20–39 years, 40–59 years, and ≥60 years (median, 49.5; interquartile range [IQR], 33.5–68). Participants were required to have a light diet 3 days before the examination. On the day of the examination, fasting samples of saliva, dental plaque, and tongue plaque were collected, rapidly frozen at -80 °C, and sent for testing. Oral examinations were performed by professional dentists, and panoramic radiographs were taken to analyze the alveolar bone condition. Other relevant information was collected through self-reported questionnaires completed by participants.

Microbiota 16S rRNA Sequencing

DNA Extraction and Amplification

Total genomic DNA was extracted using the MagPure Soil DNA LQ Kit (Magan) following the manufacturer's instructions. DNA concentration and integrity were measured with a NanoDrop 2000 (Thermo Fisher Scientific, USA) and agarose gel electrophoresis. Extracted DNA was stored at -20 °C until further processing. The extracted DNA was used as a template for PCR amplification of bacterial 16S rRNA genes with barcoded primers and Takara Ex Taq (Takara). For bacterial diversity analysis, the V3-V4 variable regions of 16S rRNA genes were amplified with universal primers 343F (5'-TACGGRAGGCAGCAG-3') and 798R (5'-AGGGTATCTAATCCT-3').

Library Construction and Sequencing

Amplicon quality was visualized using agarose gel electrophoresis. The PCR products were purified with AMPure XP beads (Agencourt) and subjected to another round of PCR amplification. After re-purification with AMPure XP beads, the final amplicons were quantified using the Qubit dsDNA Assay Kit (Thermo Fisher Scientific, USA). The concentrations were then adjusted for sequencing. Sequencing was performed on an Illumina NovaSeq 6000 with 250 bp paired-end reads (Illumina Inc., San Diego, CA; OE Biotech Company, Shanghai, China).

Bioinformatic Analysis

Library sequencing and data processing were performed by OE Biotech Co., Ltd. (Shanghai, China). Raw data in FASTQ format were preprocessed with Cutadapt to trim adapters. Based on the QIIME2 platform[12], the DADA2 algorithm[13] was used for sequence filtering, denoising, merging, and chimera removal to obtain Amplicon Sequence Variant (ASV) representative sequences and an abundance table. Representative sequences were selected via QIIME2, and taxonomic annotation was completed by aligning with the Silva 138 database using the q2-feature-classifier plugin.

Alpha and beta diversity analyses were performed using QIIME2 software. The construction of genus-level microbial co-occurrence networks and machine learning models was implemented with R language packages. The remaining bioinformatic analyses were conducted via the OECloud tools (https://cloud.oebiotech.com).

Results

Table 1 presents the demographic and clinical characteristics of the 120 participants, who were classified into the low-risk group (n=67) and high-risk group (n=53) based on periodontal indices including PD, CAL, and radiological bone resorption. For subsequent microbiota analyses, the low-risk group was defined as the non-periodontal (NP) group and the high-risk group as the periodontal (P) group (NP: low risk; P: high risk). Age was identified as a key distinguishing factor with a significant inter-group difference (p=0.003, Wilcoxon rank sum test), highlighting its critical role in risk stratification of age-related oral diseases. No statistically significant differences were observed between the two groups in other demographic (sex) and lifestyle variables (overweight status, smoking, alcohol consumption, tea drinking, sugar-free diet adherence, insufficient toothbrushing frequency), with all p-values >0.05 (Pearson's Chi-squared test for categorical variables).

Table 1 Demographic and clinical characteristics of the study participants

Variables	Low risk (NP, n=67)	High risk (P, n=53)	P-value[1]
Age, Mean ± SD	46.40 ± 17.82	55.98 ± 17.45	0.004
Male, n(%)	33 (49.25)	24 (45.28)	0.665
Overweight, n(%)	12 (17.91)	16 (30.19)	0.114
Smoker, n(%)	12 (17.91)	10 (18.87)	0.893
Alcohol drinker, n(%)	7 (10.45)	8 (15.09)	0.445
Tea drinker, n(%)	30 (44.78)	25 (47.17)	0.794
Sugar-free diet, n(%)	13 (19.40)	17 (32.08)	0.111
Insufficient toothbrushing, n(%)	22 (32.84)	18 (33.96)	0.897

[1]T test for continuous variables; chi-square test for categorical variables.

Fig 1 illustrates the study workflow and oral microbial community characteristics. The study was divided into three components (Fig 1.A): first, 120 participants were stratified into NP and P groups based on Community Periodontal Index (CPI) scores; second, demographic/lifestyle data, oral clinical examination results, and samples (dental plaque, tongue plaque, saliva) were collected; third, microbial community structure analysis, meta-analysis, and functional prediction were performed. Genus-level stacked bar charts show the taxonomic composition of microbial communities across the three sampling sites, with *Streptococcus* exhibiting the most prominent variation in tongue plaque, *Actinomyces* in dental plaque, and *Prevotella* in saliva between the NP and P groups (Fig 1.B). Alpha-diversity indices (Shannon, Chao1) showed no significant differences between the NP and P groups within each sampling site (Fig 1.C). The principal coordinate analysis (PCoA) plot of beta-diversity (Bray-Curtis dissimilarity) revealed distinct clustering by sampling site (permutational multivariate analysis of variance [PERMANOVA], p=0.001), indicating significant microbial community differences across sites between the two groups (Fig 1.D). Genus-level co-occurrence networks showed that the microbial interaction intensity was significantly enhanced in the dental plaque of the P group, while no obvious changes were observed in tongue plaque and saliva—suggesting that altered microbial co-occurrence patterns are niche-specific to dental plaque in periodontal conditions (Fig 1.E).

Fig 1 Community characteristics of multi-site oral microbiota in populations with different periodontal risks

A Schematic diagram of the overall study workflow. The main procedures include participant enrollment, sample collection, and data analysis. 16S rRNA sequencing data analysis involves community structure analysis, feature selection and comparison, machine learning (ML) model construction, model performance evaluation, and functional prediction. **B** Stacked bar charts showing the relative abundances of the top 20 genera across different sampling sites and periodontal risk statuses (NP: low risk; P: high risk). **C** Alpha-diversity indices of the microbial communities in different study groups (Shannon and Chao1). **D** Principal coordinates analysis (PCoA) based on Bray-Curtis distance (PERMANOVA p-value: 0.001). **E** Co-occurrence networks of genera based on correlation analysis. A connection represents a strong and significant correlation ($r > 0.5$, $p < 0.05$). The size of each node is proportional to the number of connections. Phylum legend: *Firmicutes*, *Bacteroidota*, *Proteobacteria*, *Fusobacteriota*, *Actinobacteriota*, *Campilobacterota*, *Spirochaetota*, *Patescibacteria*, *Deferribacterota*, *Desulfobacterota*.

Fig 2 identifies periodontitis-specific bacterial biomarkers across different oral sites via linear discriminant analysis effect size (LEfSe) and random forest (RF) models. In dental plaque, LEfSe analysis revealed significant enrichment of *Lawsonella* and *Dialister* in the P group (LDA score > 2, $p < 0.05$), with *Aerococcus* validated as the core discriminatory biomarker by RF (Fig 2.A/B). For tongue plaque, *Streptococcus* was the most prominently enriched taxon in the P group by LEfSe, while *Porphyromonas* emerged as the key biomarker in RF analysis (Fig 2.C/D). In saliva, LEfSe showed marked enrichment of

Treponema in the P group, and RF confirmed *Leptotrichia* as the core discriminatory taxon (Fig 2.E/F). Collectively, periodontitis-associated bacterial biomarkers exhibit distinct site-specificity, yet all core markers effectively distinguish the P group from the NP group, providing targeted candidates for site-specific screening.

Fig 2 Bacterial biomarkers identified from three sampling sites via feature selection analyses

Panels **A and B** represent comparisons between the NP and P groups in dental plaque samples; Panels **C and D** represent comparisons in tongue plaque samples; Panels E and F represent comparisons in saliva samples. Panels **A, C, and E** show LEfSe analysis results, with the top 10 genera ranked by absolute values of linear discriminant analysis (LDA) scores (LDA score > 2) and significant differences validated by the Kruskal-Wallis test ($p < 0.05$). Panels **B, D, and F** show mean decrease Gini values calculated by the RF model, where the order from top to bottom indicates the importance ranking of bacterial genera. The sum of the relative abundances of these important genera across all samples is also presented on the right.

Fig 3 presents the receiver operating characteristic (ROC) curves of multiple machine learning models constructed using site-specific oral microbiota from distinct niches. Notably, among the three sampling sites (saliva, tongue plaque, dental plaque), salivary microbiota exhibited the highest diagnostic efficacy, with a maximum area under the curve (AUC) of 0.734, outperforming both tongue plaque and dental plaque microbiota. Additionally, model selection analyses revealed that Boosting-based algorithms yielded superior predictive performance compared to other machine learning approaches.

Fig 3 Evaluation of diagnostic efficacy of genus-level core taxa biomarkers using different feature selection methods

Panels **A, C, and E** show training ROC curves of models constructed based on LEfSe features, while Panels **B, D, and F** represent training ROC curves of models built using RF features. The curve corresponding to the model with the optimal training performance is highlighted in red. All models were validated using leave-one-out cross-validation (LOOCV).

Previous studies have well established distinct age-related alterations in the oral microbiota of elderly individuals[14-16]. We therefore hypothesized that constructing predictive models with stratified analysis of older (>50 years) and prime-age (≤50 years) populations would better align with clinical practice (demographic and clinical characteristics of the age-stratified participants are provided in Supplementary Table 2). Fig 4 depicts the key signature taxa identified by the RF model and their relative abundances across the two age groups. Notably, a greater number of key signature taxa (with high mean decrease Gini values and elevated relative abundances) were detected in the elderly population for both dental plaque and tongue plaque niches, whereas no notable difference was observed in the number of signature taxa between elderly and prime-age individuals in saliva. This observation suggests that periodontitis-associated pathogenic taxa are more diverse in the elderly relative to younger individuals. For dental plaque, *Aerococcus* represented a prominent signature taxon in individuals aged ≤50 years,

while *Dialister*, Porphyromonas, and other taxa emerged as key discriminatory taxa in those aged >50 years (Fig. 4A/B). In the tongue plaque niche, *Fusobacterium* and *Neisseria* were the characteristic taxa for individuals ≤50 years, whereas *Actinomyces*, *Porphyromonas*, and other taxa were identified as periodontitis-specific taxa in the >50 years group (Fig. 4C/D). For saliva, *Corynebacterium* and *Tannerella* were the predominant signature taxa in individuals ≤50 years, while *Catonella*, *Oribacterium*, and related taxa represented the key discriminatory taxa in those >50 years. Collectively, these findings demonstrate that periodontitis-associated signature taxa in the elderly differ significantly from those in prime-age individuals, both in terms of taxon number and phylogenetic composition.

Fig 4 Age-specific bacterial biomarkers identified from three sampling sites via RF model

Panels **A, C, and E** show mean decrease Gini values of microbiota in dental plaque, tongue plaque, and saliva of the prime-age group (≤50 years), calculated by the RF model. Panels **B, D, and F** show mean decrease Gini values of microbiota in the above three sampling sites of the older group (>50 years), calculated by the RF model. The order from top to bottom indicates the importance ranking of bacterial genera, and the sum of the relative abundances of these important genera across all samples is shown on the right.

Fig. 5 displays the ROC curves of multiple machine learning models constructed with site-specific microbiota upon age stratification (≤50 years and >50 years). We observed a universal improvement in overall predictive accuracy across all models, with the salivary niche still exhibiting the highest diagnostic efficacy (AUC=0.807).

Fig 5 Evaluation of age-stratified diagnostic efficacy of genus-level core taxa biomarkers

Panels **A, C, and E** show training ROC curves of models constructed based on RF features of the prime-age group (≤50 years), while Panels **B, D, and F** represent training ROC curves of models built using RF features of the older group (>50 years). The curve corresponding to the model with the optimal training performance is highlighted in red. All models were validated using LOOCV.

Finally, functional pathway enrichment analysis was conducted on the microbiota from distinct oral niches (Fig. 6). In dental plaque, the NADSYN-PWY (NAD biosynthesis II pathway) and PWY-5651 (pyrimidine nucleotide biosynthesis pathway) were significantly enriched in the P group, indicative of robust proliferative activity of the dental plaque microbiota. In tongue plaque, PWY1G-0 (vitamin B12-associated biosynthesis pathway) and POLYAMSYN-PWY (polyamine biosynthesis super pathway) were markedly enriched in the P group. A prior study established a correlation between microbial vitamin B12 metabolism and taste acuity in the elderly, and vitamin B12 is an essential nutrient for maintaining neurological function[17]; this implies that microbial dysbiosis in periodontitis patients may induce taste impairment. Aberrant glycogen accumulation in lingual muscles is a key contributor to macroglossia, and periodontitis-associated microbiota may compromise lingual motor function by modulating glycogen biosynthesis in lingual muscles[18]. In saliva, the LEU-DEG2-PWY (L-leucine degradation II pathway) was significantly enriched, suggesting that periodontitis-associated bacteria may obtain

energy and carbon skeletons through leucine catabolism. Collectively, our findings demonstrate niche-specific functional enrichment patterns of periodontitis-associated microbiota, providing empirical data to support the development of precision metabolic interventions for periodontitis targeting oral microbial metabolism.

Fig 6 Functional predictions of the oral microbiome among NP and P groups from distinct oral niches

Mean proportion of Metacyc pathway abundances between the NP and P groups in A dental plaque, B tongue plaque, and C saliva. Statistical significance between groups was analyzed using t-test (*p<0.05).

Discussion

One of the key findings of the present study is that age stratification significantly enhanced the diagnostic efficacy of periodontal risk assessment based on oral microbiota—a result closely associated with the age distribution characteristics of the study population and age-related remodeling patterns of the oral microbiota. Demographic data revealed that the median age of subjects in the high-risk (P) group was significantly higher than that in the low-risk (NP) group, with a statistically significant intergroup difference (p=0.003, Table 1), which corroborates the epidemiological characteristic of age as a major risk factor for periodontitis. Further analyses showed that in the absence of age stratification, the maximum AUC of the optimal diagnostic model across the three non-invasive sampling sites was only 0.734 (salivary samples, Fig. 3F). Following stratification into ≤50 years and >50 years age groups, however, the diagnostic AUC of salivary samples increased to 0.807 and 0.743, respectively—representing the optimal diagnostic approach. Additionally, the oral microbiota distribution in the elderly exhibited distinct niche specificity. Several studies have demonstrated that the abundances of *Lactobacillus*, *Candida albicans*, and *Staphylococcus* in saliva are significantly higher in the elderly than in prime-age individuals, with increased levels of *Actinomyces* and aciduric bacteria in dental plaque and tongue plaque niches, respectively[19-21]. Our data further confirmed that periodontitis signature taxa in the elderly also displayed remarkable niche-specific differences. Screening via the RF model and relative abundance ranking identified *Oribacterium* and *Bergeyella* in saliva, *Porphyromonas* in dental plaque, and *Actinomyces* in tongue plaque as the prominent signature taxa in elderly patients with periodontitis. Collectively, given the unique oral microbiota characteristics of middle-aged and elderly individuals, age stratification in microbiota-based diagnostic models for periodontitis enables the models to focus more specifically on periodontitis-associated microbial dysbiosis per se, thereby providing a critical methodological reference for the standardized construction of microbiota-based diagnostic models for periodontitis.

The present study also systematically compared the diagnostic efficacy across three non-invasive sampling sites (saliva, dental plaque, tongue plaque), with definitive results demonstrating that salivary samples exhibited superior diagnostic performance. After age stratification, the AUC of salivary samples reached 0.807, which outperformed dental plaque (maximum AUC=0.780) and tongue plaque (AUC=0.753). This finding is likely attributable to the combined effects of the biological properties and

microbial community characteristics of saliva. In terms of microbial community structure, saliva acts as a systemic mixed reservoir for microbiota from distinct oral microenvironments (e.g., supragingiva, subgingiva, tongue dorsum), and thus integrates microbial signals across the entire oral cavity rather than being restricted to community alterations in a single local microenvironment[22]. In terms of biomarker specificity, screening results from LEfSe and the RF model confirmed that core differentially enriched biomarkers in saliva (e.g., *Treponema*, *Tannerella*) are internationally recognized periodontal pathogens, and their altered abundances are directly associated with periodontal risk status—ensuring the specificity and reliability of screening (Fig. 2E/F). Furthermore, salivary sampling offers notable clinical advantages including non-invasiveness, ease of collection, and high patient compliance. It does not require the intervention of professional dental clinicians, rendering it highly suitable for early screening in large-scale population settings such as community screening and primary care, and thus laying a solid foundation for the clinical translation of non-invasive diagnostic tools for periodontitis.

Despite the meaningful findings of the present study, several limitations exist that warrant further refinement in subsequent research. First, there is considerable room for improvement in diagnostic efficacy: the maximum AUC of the optimal model in this study was 0.807, which has not yet reached the high accuracy threshold of >0.9. This limitation is potentially attributed to interindividual variability in the oral microbiota and insufficient functional validation of core biomarkers, among other factors. Future studies should further optimize feature selection strategies and integrate multi-omics biological information to improve model accuracy. Second, the sample size and population representativeness are limited: this was a single-center cohort study enrolling only 120 subjects, all recruited from communities and a tertiary dental hospital in Shanghai. The relatively concentrated geographical distribution and lifestyle characteristics of the study population may have introduced selection bias, thus limiting the generalizability of the developed models. Subsequent investigations should conduct large-sample, multicenter prospective cohort studies (e.g., ≥500 cases) enrolling populations with diverse geographical origins, ethnicities, and systemic health statuses to validate the stability and applicability of the models. Third, limitations exist in the sequencing and analytical techniques employed: 16S rRNA gene sequencing, as used in this study, only reveals the taxonomic composition and relative abundances of the microbial community, and cannot capture functional metabolic characteristics (e.g., virulence factor expression, metabolite secretion) or strain-level variations—hindering a comprehensive elucidation of the molecular mechanisms by which microbial dysbiosis drives periodontitis. Future studies should integrate metagenomic sequencing and metabolomic technologies to conduct an in-depth analysis of the microbiota-function-disease axis. Finally, an independent external validation cohort was not established to verify the clinical applicability of the diagnostic cut-offs, which has also compromised the efficiency of clinical translation of the study findings to some extent. These limitations collectively point to future research directions; strategies including expanding the sample size and optimizing technical approaches are expected to further enhance the accuracy and clinical utility of microbiota-based screening for periodontitis.

Conclusion

This study confirmed that oral niche heterogeneity and age stratification are essential for microbiota-based periodontitis screening. Age stratification significantly improved model diagnostic accuracy, with salivary microbiota showing the optimal efficacy (AUC=0.807) among non-invasive samples, and distinct site-specific microbial biomarkers and functional enrichment patterns were identified in periodontitis patients. Saliva is the ideal non-invasive sampling matrix for periodontitis screening, providing key evidence for developing precise diagnostic tools for community and primary healthcare settings.

List of abbreviations

AUC：Area Under the Curve

ASV：Amplicon Sequence Variant

BOP：Bleeding on Probing

CAL：Clinical Attachment Loss

CPI：Community Periodontal Index

GCF：Gingival Crevicular Fluid

IQR：Interquartile Range

LEfSe：Linear Discriminant Analysis Effect Size

LOOCV：Leave-One-Out Cross-Validation

ML：Machine Learning

NP：Non-Periodontal

P：Periodontal

PD：Periodontal Probing Depth

PCoA：Principal Coordinate Analysis

PERMANOVA：Permutational Multivariate Analysis of Variance

PLS：Partial Least Squares

RF：Random Forest

ROC：Receiver Operating Characteristic

rRNA：Ribosomal Ribonucleic Acid

Declarations

Ethics Approval and Consent to Participate

This cohort study was conducted at communities and tertiary oral specialist hospitals in Shanghai following approval from the Ethics Committee of Shanghai Stomatological Hospital (Ethics Approval Nos.: [2023]019 and [2024]021). The study was conducted in accordance with the tenets of the Declaration of Helsinki and relevant clinical research guidelines. All eligible participants were fully informed of the study content, and provided written informed consent to participate prior to enrollment.

Consent for publication

Not applicable.

Availability of data and materials

The raw 16S rRNA gene sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1444982. A temporary reviewer access link is provided as https://dataview.ncbi.nlm.nih.gov/object/PRJNA1444982?reviewer=omef69sdq1ucj0unqcfr6eh3e6. The full dataset will be made publicly available upon acceptance of the manuscript for publication. The detailed study protocol and additional bioinformatic analysis results are available in the Additional files.

Competing interests

All authors have no conflicts of interest to declare.

Funding

This work was supported by National Key Research and Development Program of China (2023YFC3605600), National Key Clinical Program on Orthodontics and Shanghai Municipal Science and Technology Major Project (2023SHZDZX02).

Authors' Contributions

Yinan Chen and Enhua Mei contributed equally to this work. Yan Wang and Yuehua Liu conceived and designed the study. Wei Zhang provided technical support for sample collection, microbial sequencing and experimental operation. Yinan Chen and Enhua Mei collected and assembled the clinical, demographic and sequencing data, and performed the data analysis and interpretation with the guidance of Yan Wang and Yuehua Liu. All authors contributed to manuscript writing, had full access to the data, and participated in reviewing and editing the text. All authors read and approved the final manuscript.

Acknowledgements

We thank all participating subjects, clinical dentists, research staff who involved in sample collection, clinical examination and experimental operation, as well as the technical team of OE Biotech Co., Ltd. for their professional support in 16S rRNA gene sequencing and bioinformatic analysis. The authors did not receive payment from any pharmaceutical company or other agency for writing this article. The corresponding authors affirm that no author was precluded from accessing the data in the study, that all authors had access to or could obtain the underlying data upon reasonable request, and that the corresponding authors accept responsibility for the decision to submit the manuscript for publication.

References

1. Curtis MA, Diaz PI, Van Dyke TE: **The role of the microbiota in periodontal disease**. *Periodontol 2000* 2020, **83**(1):14-25.
2. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL, Jr.: **Microbial complexes in subgingival plaque**. *J Clin Periodontol* 1998, **25**(2):134-144.
3. Chen MX, Zhong YJ, Dong QQ, Wong HM, Wen YF: **Global, regional, and national burden of severe periodontitis, 1990-2019: An analysis of the Global Burden of Disease Study 2019**. *J Clin Periodontol* 2021, **48**(9):1165-1188.
4. Beck JD, Papapanou PN, Philips KH, Offenbacher S: **Periodontal Medicine: 100 Years of Progress**. *J Dent Res* 2019, **98**(10):1053-1062.
5. Tonetti MS, Greenwell H, Kornman KS: **Staging and grading of periodontitis: Framework and proposal of a new classification and case definition**. *J Periodontol* 2018, **89 Suppl 1**:S159-s172.
6. Gul SS, Abdulkareem AA, Sha AM, Rawlinson A: **Diagnostic Accuracy of Oral Fluids Biomarker Profile to Determine the Current and Future Status of Periodontal and Peri-Implant Diseases**. *Diagnostics (Basel)* 2020, **10**(10).
7. Tanner AC, Paster BJ, Lu SC, Kanasi E, Kent R, Jr., Van Dyke T, Sonis ST: **Subgingival and tongue microbiota during early periodontitis**. *J Dent Res* 2006, **85**(4):318-323.
8. Di Stefano M, Polizzi A, Santonocito S, Romano A, Lombardi T, Isola G: **Impact of Oral Microbiome in Periodontal Health and Periodontitis: A Critical Review on Prevention and Treatment**. *Int J Mol Sci* 2022, **23**(9).
9. **Structure, function and diversity of the healthy human microbiome**. *Nature* 2012, **486**(7402):207-214.
10. Zaura E, Pappalardo VY, Buijs MJ, Volgenant CMC, Brandt BW: **Optimizing the quality of clinical studies on oral microbiome: A practical guide for planning, performing, and reporting**. *Periodontol 2000* 2021, **85**(1):210-236.
11. Blanco-Pintos T, Regueira-Iglesias A, Relvas M, Alonso-Sampedro M, Chantada-Vázquez MP, Balsa-Castro C, Tomás I: **Using SWATH-MS to identify new molecular biomarkers in gingival crevicular fluid for detecting periodontitis and its response to treatment**. *J Clin Periodontol* 2024, **51**(10):1342-1358.

12. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F *et al*: **Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2**. *Nat Biotechnol* 2019, **37**(8):852-857.
13. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP: **DADA2: High-resolution sample inference from Illumina amplicon data**. *Nat Methods* 2016, **13**(7):581-583.
14. Feres M, Teles F, Teles R, Figueiredo LC, Faveri M: **The subgingival periodontal microbiota of the aging mouth**. *Periodontol 2000* 2016, **72**(1):30-53.
15. Zawadzki PJ, Perkowski K, Padzik M, Mierzwińska-Nastalska E, Szaflik JP, Conn DB, Chomicz L: **Examination of Oral Microbiota Diversity in Adults and Older Adults as an Approach to Prevent Spread of Risk Factors for Human Infections**. *Biomed Res Int* 2017, **2017**:8106491.
16. Schwartz JL, Peña N, Kawar N, Zhang A, Callahan N, Robles SJ, Griebel A, Adami GR: **Old age and other factors associated with salivary microbiome variation**. *BMC Oral Health* 2021, **21**(1):490.
17. Jeon S, Kim Y, Min S, Song M, Son S, Lee S: **Taste Sensitivity of Elderly People Is Associated with Quality of Life and Inadequate Dietary Intake**. *Nutrients* 2021, **13**(5).
18. Keeler AM, Zieger M, Todeasa SH, McCall AL, Gifford JC, Birsak S, Choudhury SR, Byrne BJ, Sena-Esteves M, ElMallah MK: **Systemic Delivery of AAVB1-GAA Clears Glycogen and Prolongs Survival in a Mouse Model of Pompe Disease**. *Hum Gene Ther* 2019, **30**(1):57-68.
19. Jiang Q, Liu J, Chen L, Gan N, Yang D: **The Oral Microbiome in the Elderly With Dental Caries and Health**. *Front Cell Infect Microbiol* 2018, **8**:442.
20. Asakawa M, Takeshita T, Furuta M, Kageyama S, Takeuchi K, Hata J, Ninomiya T, Yamashita Y: **Tongue Microbiota and Oral Health Status in Community-Dwelling Elderly Adults**. *mSphere* 2018, **3**(4).
21. Mittal R, Tan KS, Wong ML, Allen PF: **Correlation between microbial host factors and caries among older adults**. *BMC Oral Health* 2021, **21**(1):47.
22. Jia G, Zhi A, Lai PFH, Wang G, Xia Y, Xiong Z, Zhang H, Che N, Ai L: **The oral microbiota - a mechanistic role for systemic diseases**. *Br Dent J* 2018, **224**(6):447-455.

Figures

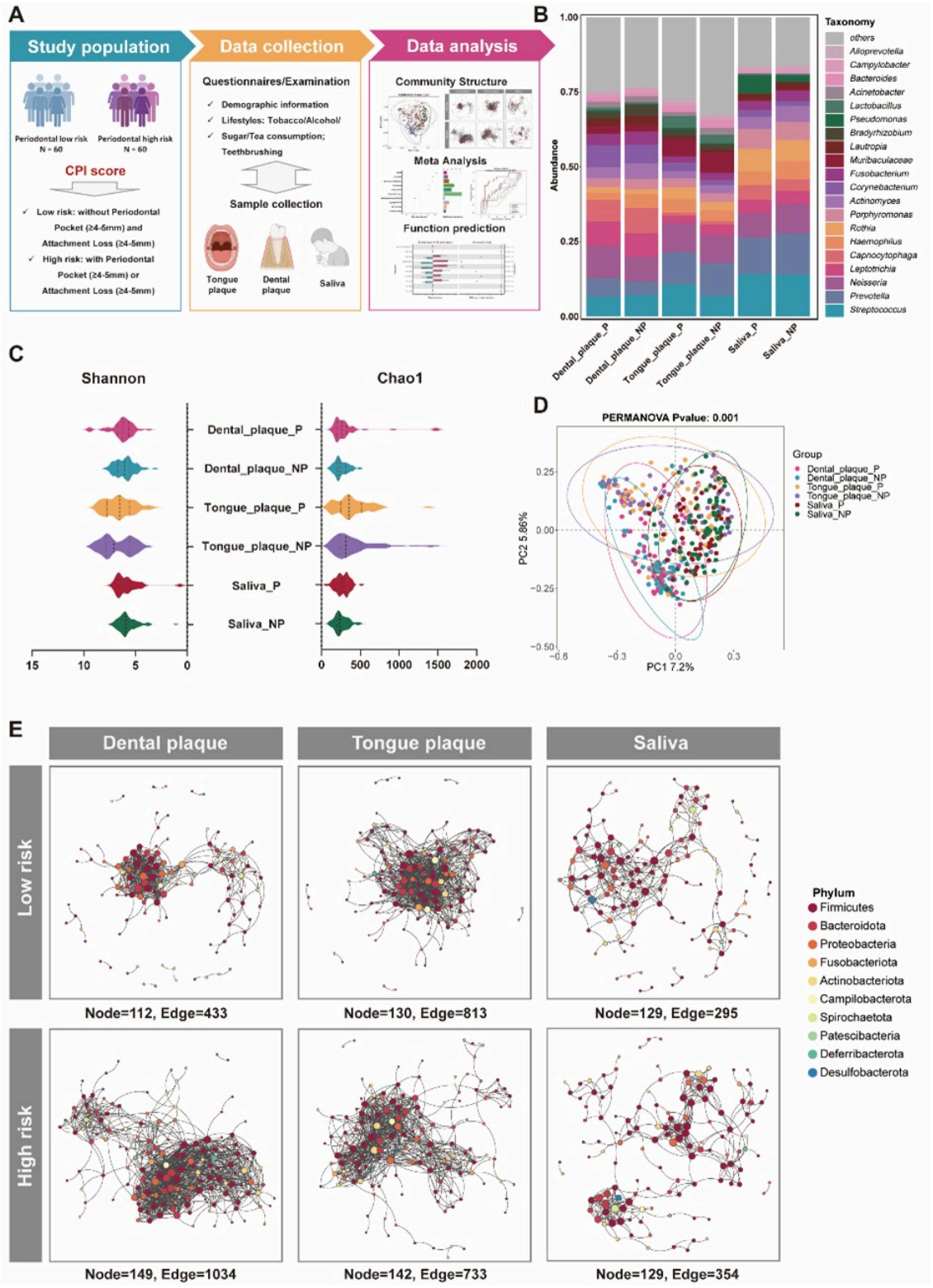


Figure 1

Community characteristics of multi-site oral microbiota in populations with different periodontal risks

A Schematic diagram of the overall study workflow. The main procedures include participant enrollment, sample collection, and data analysis. 16S rRNA sequencing data analysis involves community structure analysis, feature selection and comparison, machine learning (ML) model construction, model

performance evaluation, and functional prediction. **B** Stacked bar charts showing the relative abundances of the top 20 genera across different sampling sites and periodontal risk statuses (NP: low risk; P: high risk). **C**Alpha-diversity indices of the microbial communities in different study groups (Shannon and Chao1). **D** Principal coordinates analysis (PCoA) based on Bray-Curtis distance (PERMANOVA p-value: 0.001). **E** Co-occurrence networks of genera based on correlation analysis. A connection represents a strong and significant correlation ($r > 0.5$, $p < 0.05$). The size of each node is proportional to the number of connections. Phylum legend: *Firmicutes*, *Bacteroidota*, *Proteobacteria*, *Fusobacteriota*, *Actinobacteriota*, *Campilobacterota*, *Spirochaetota*, *Patescibacteria*, *Deferribacterota*, *Desulfobacterota*.

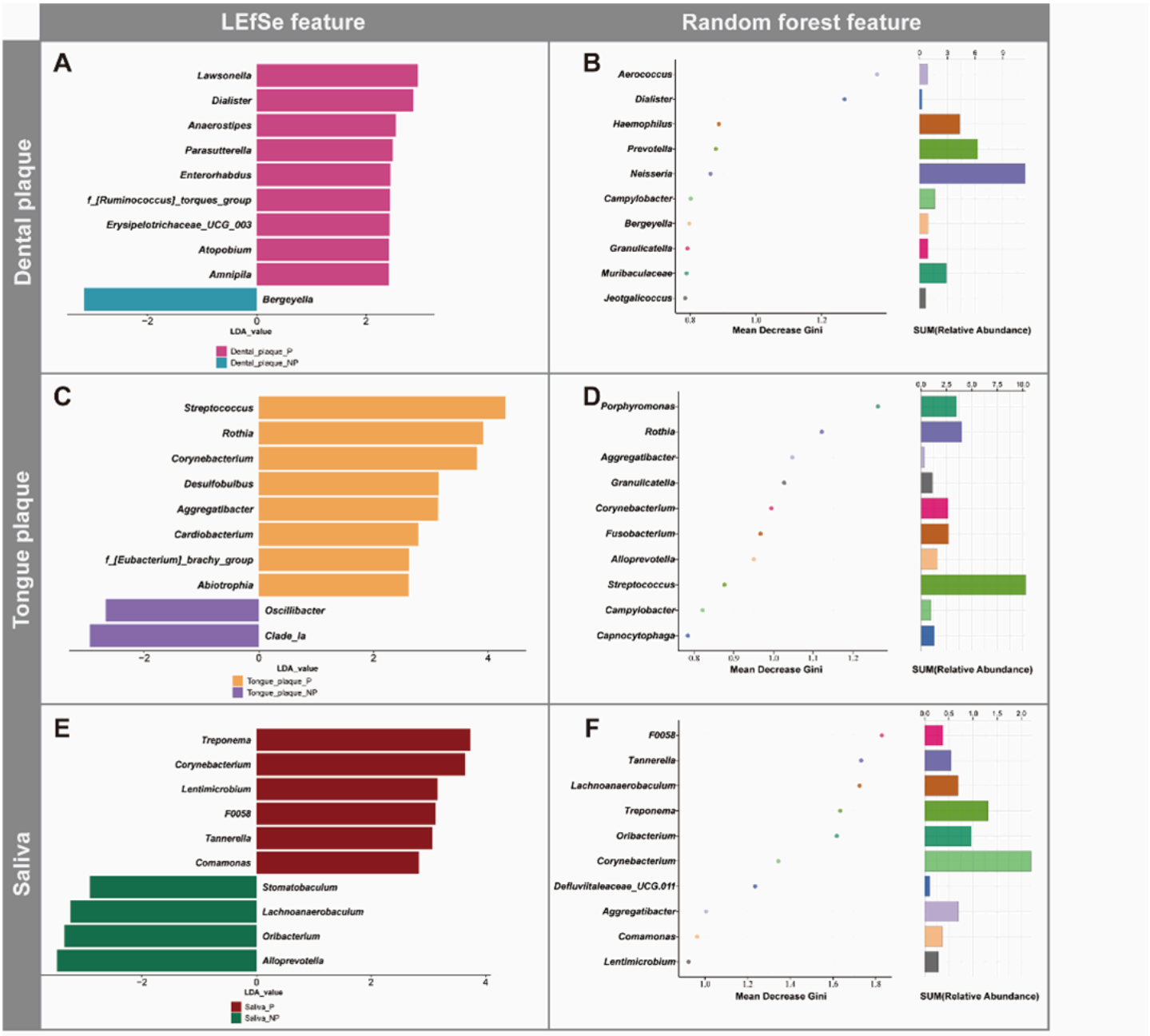


Figure 2

Bacterial biomarkers identified from three sampling sites via feature selection analyses

Panels **A and B** represent comparisons between the NP and P groups in dental plaque samples; Panels **C and D** represent comparisons in tongue plaque samples; Panels E and F represent comparisons in saliva samples. Panels **A, C, and E** show LEfSe analysis results, with the top 10 genera ranked by absolute values of linear discriminant analysis (LDA) scores (LDA score > 2) and significant differences validated by the Kruskal-Wallis test ($p < 0.05$). Panels **B, D, and F** show mean decrease Gini values calculated by the RF model, where the order from top to bottom indicates the importance ranking of bacterial genera. The sum of the relative abundances of these important genera across all samples is also presented on the right.

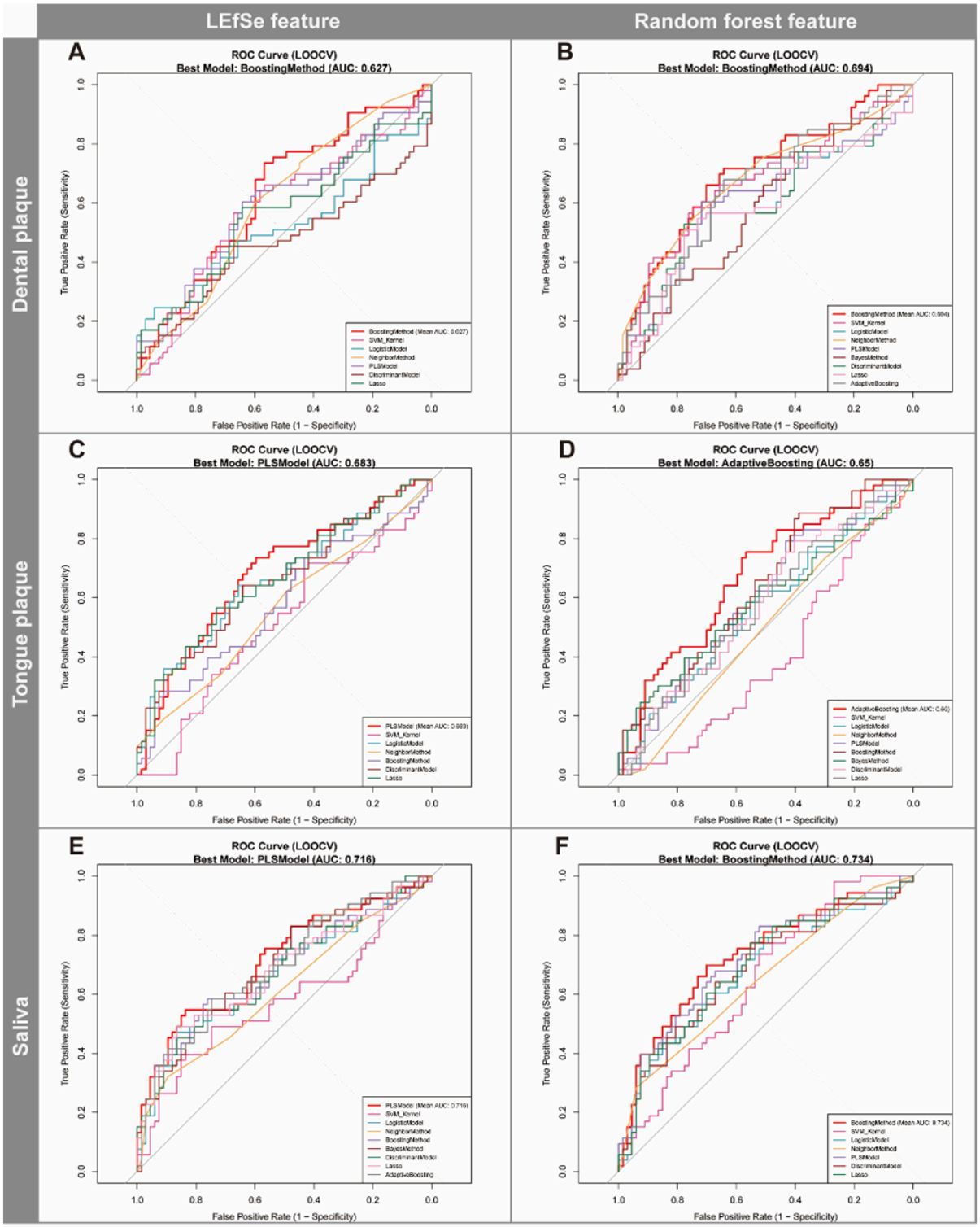


Figure 3

Evaluation of diagnostic efficacy of genus-level core taxa biomarkers using different feature selection methods

Panels **A, C, and E** show training ROC curves of models constructed based on LEfSe features, while Panels **B, D, and F** represent training ROC curves of models built using RF features. The curve

corresponding to the model with the optimal training performance is highlighted in red. All models were validated using leave-one-out cross-validation (LOOCV).

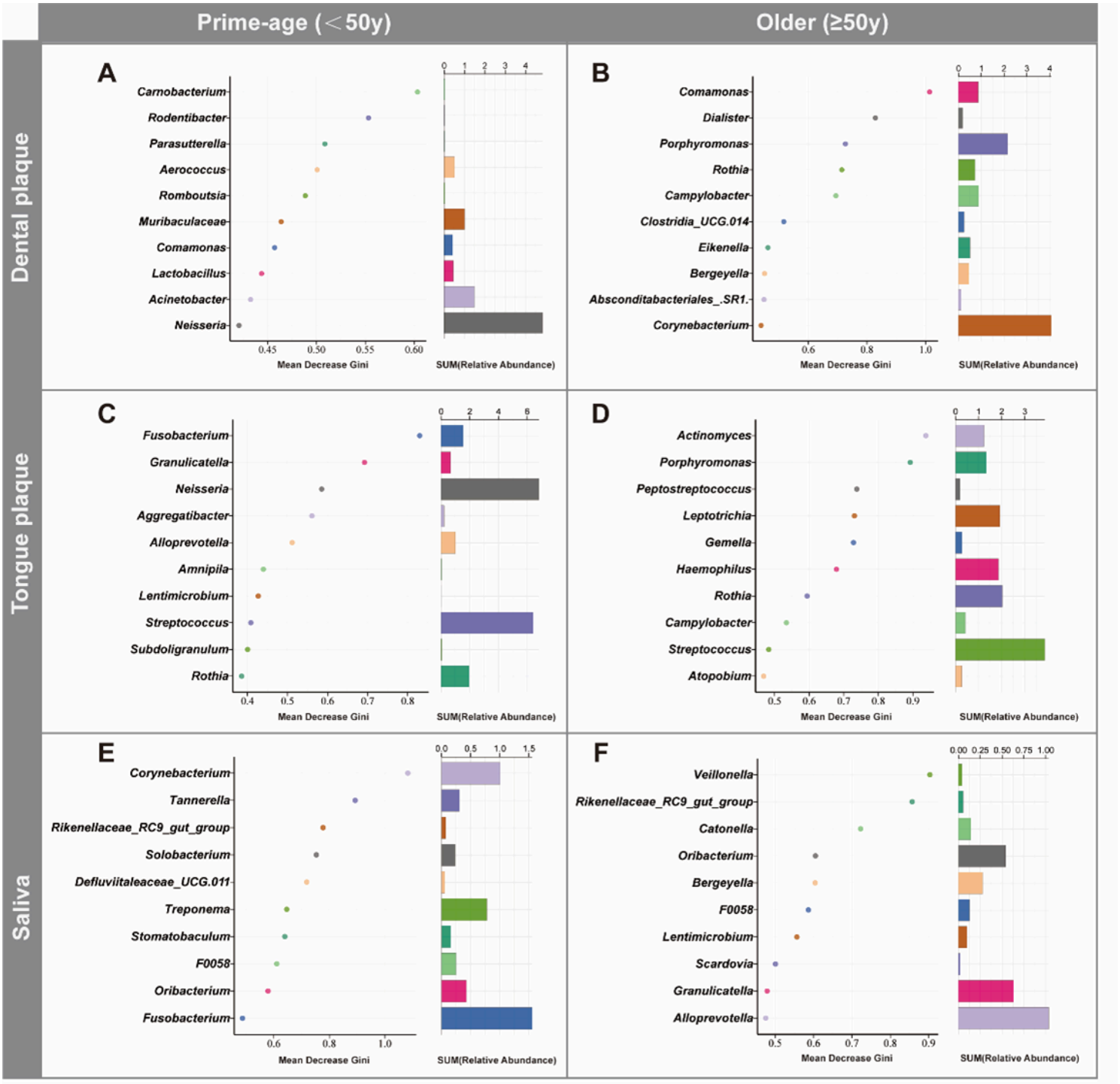


Figure 4

Age-specific bacterial biomarkers identified from three sampling sites via RF model

Panels **A, C, and E**show mean decrease Gini values of microbiota in dental plaque, tongue plaque, and saliva of the prime-age group (≤50 years), calculated by the RF model. Panels **B, D, and F** show mean decrease Gini values of microbiota in the above three sampling sites of the older group (>50 years),

calculated by the RF model. The order from top to bottom indicates the importance ranking of bacterial genera, and the sum of the relative abundances of these important genera across all samples is shown on the right.

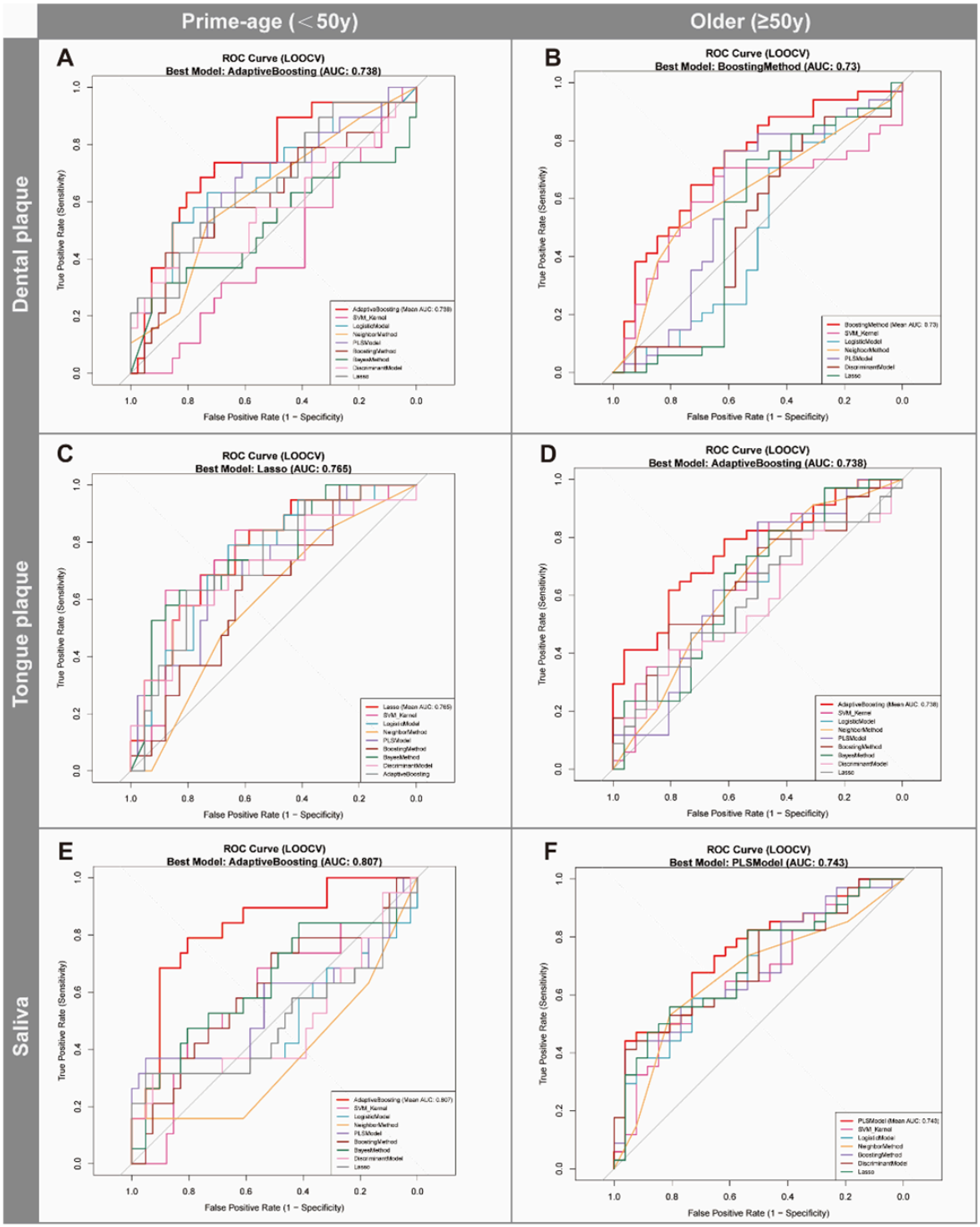


Figure 5

Evaluation of age-stratified diagnostic efficacy of genus-level core taxa biomarkers

Panels **A, C, and E** show training ROC curves of models constructed based on RF features of the prime-age group (≤50 years), while Panels **B, D, and F** represent training ROC curves of models built using RF features of the older group (>50 years). The curve corresponding to the model with the optimal training performance is highlighted in red. All models were validated using LOOCV.

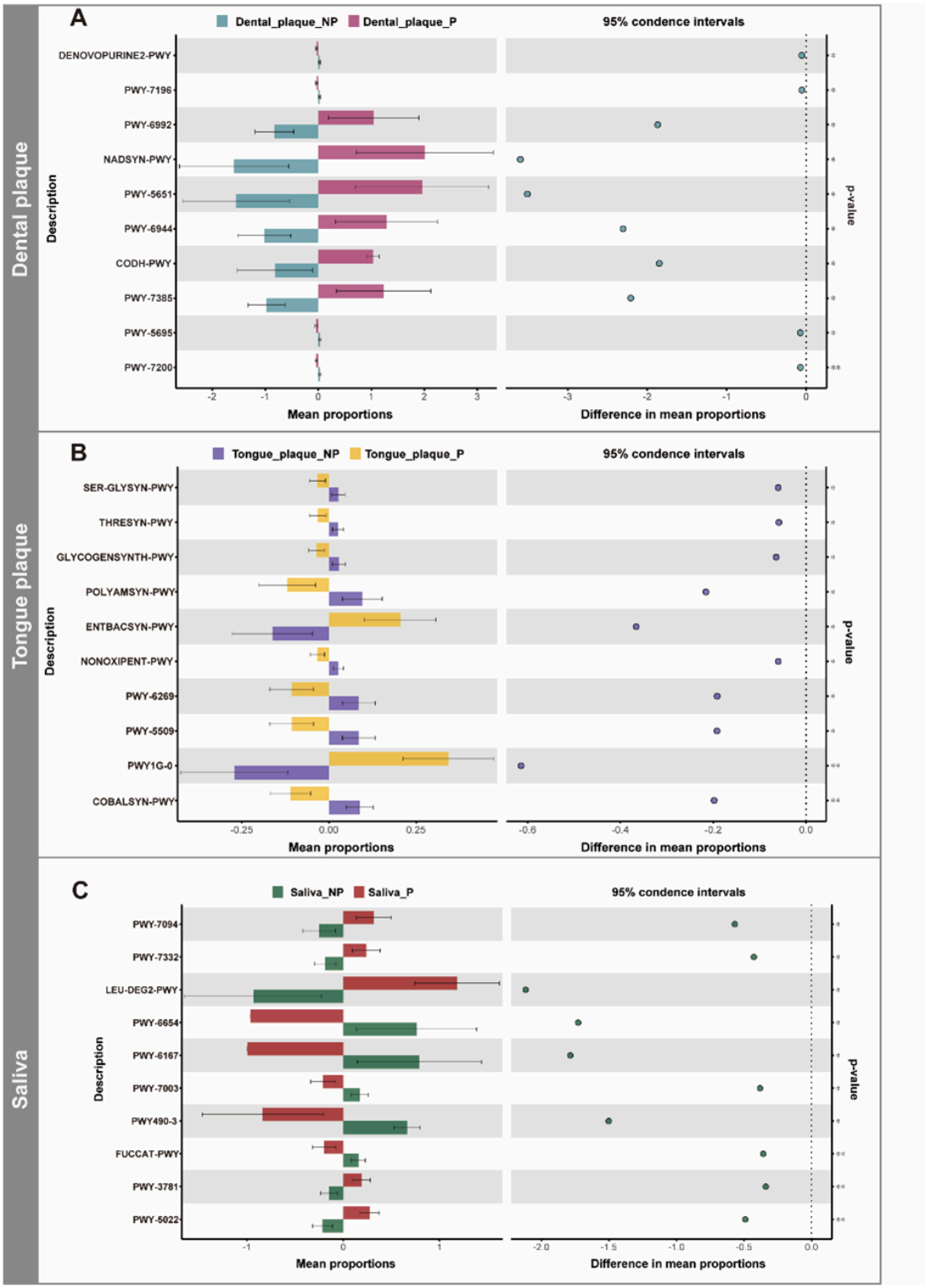


Figure 6

Functional predictions of the oral microbiome among NP and P groups from distinct oral niches

Mean proportion of Metacyc pathway abundances between the NP and P groups in A dental plaque, B tongue plaque, and C saliva. Statistical significance between groups was analyzed using t-test (*p<0.05).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.xlsx