8-2019

Periodontal pathogens are a risk factor of oral cavity squamous cell carcinoma, independent of tobacco and alcohol and human papillomavirus

Ian Ganly

Liying Yang

Rachel Giese

Yuhan Hao

Carlos W. Nossa

See next page for additional authors

Authors

Ian Ganly, Liying Yang, Rachel Giese, Yuhan Hao, Carlos W. Nossa, Luc G. T. Morris, Matthew Rosenthal, Jocelyn Migliacci, Dervla Kelly, and Wenzhi Tseng

Periodontal pathogens are a risk factor of oral cavity squamous cell carcinoma, independent of tobacco and alcohol and human papillomavirus

Ian Ganly[1], Liying Yang[2,3], Rachel A. Giese[1], Yuhan Hao[2,4,5], Carlos W. Nossa[6], Luc G. T. Morris[1], Matthew Rosenthal[1], Jocelyn Migliacci[1], Dervla Kelly[2], Wenzhi Tseng[2], Jiyuan Hu[7], Huilin Li[7], Stuart Brown[4], and Zhiheng Pei[2,8]

[1]Head and Neck Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, NY, USA

[2]Department of Pathology, New York University School of Medicine, NY, USA, 10016

[3]Department of Medicine, New York University School of Medicine, NY, USA, 10016

[4]Applied Bioinformatics Laboratories, New York University School of Medicine, NY, USA, 10016

[5]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY, 10003

[6]Ascend Genetics, Bellaire, TX 77402

[7]Department of Population Health and the Department of Environmental Medicine, NY, USA, 10016

[8]Department of Veterans Affairs New York Harbor Healthcare System, NY, USA, 10010

Abstract

Over the past decade, there has been a change in the epidemiology of oral cavity squamous cell cancer (OC-SCC). Many new cases of OC-SCC lack the recognized risk factors of smoking, alcohol and human papilloma virus. The aim of this study was to determine if the oral microbiome may be associated with OC-SCC in nonsmoking HPV negative patients. We compared the oral microbiome of HPV-negative nonsmoker OC-SCC(n=18), premalignant lesions(PML) (n=8) and normal control patients (n=12). Their oral microbiome was sampled by oral wash and defined by 16S rRNA gene sequencing. We report that the periodontal pathogens *Fusobacterium, Prevotella, Alloprevotella* were enriched while commensal *Streptococcus* depleted in OC-SCC. Based on the four genera plus a marker genus *Veillonella* for PML, we classified the oral microbiome into two types. Gene/pathway analysis revealed a progressive increase of genes encoding HSP90 and ligands for TLRs 1, 2 and 4 along the controls→PML→OC-SCC progression sequence. Our

findings suggest an association between periodontal pathogens and OC-SCC in non smoking HPV negative patients.

Keywords

Microbiome; oral cavity squamous cell carcinoma; periodontal pathogens; leukoplakia; risk factor; nonsmoking; *Fusobacterium*; *Streptococcus*

Introduction

Smoking and alcohol are the two main risk factors for oral cancer. Other factors are also implicated in the aetiology of squamous cell head and neck cancer such as poor oral hygiene[1–6], diet, viruses, occupational agents, pollutants, genetic influences, but few case-controlled epidemiological studies have been carried out. Since 1990, there has been a steady increase in oral cancer in patients in the USA who do not smoke[7]. Despite a decline in the prevalence of cigarette smoking in the USA since 1975 (from ~40% to 20%) the incidence of oral cancer has remained virtually unchanged [8]. The prevalence of heavy alcohol consumption in the USA has only slightly increased from 7% to 8.2% between 2005 and 2012[9]. Oral SCC can be divided into oropharyngeal SCC and oral cavity SCC (OC-SCC). The prevalence of oropharyngeal SCC related to high-risk human papillomaviruses (HPV) has increased from 40.5% in 2000 to 72.2% in 2010[10]. The recognition of HPV etiology in oropharyngeal SCC has improved the clinical outcomes and led to specific prevention for HPV infection by vaccination. In contrast, the prevalence of HPV in OC-SCC is unclear and greatly varied across multiple studies with an average of 23.3% [11]. Thus, a significant proportion of newly diagnosed OC-SCC in the USA have no known risk factor. It is possible that this may be directly related to poor oral hygiene. There are now several studies showing an association between poor oral hygiene and oral cancer. These studies show an association with infrequent tooth brushing, gum bleeding and periodontitis[1–6]. Poor oral hygiene will result in a change in the oral microbiome of such patients. Very recently, several studies reported alteration of microbiome in oral cancer [12–14]. However, the significance of many of the studies associating the oral microbiome with OC-SCC remains unclear due to the lack of statistical correction for false discoveries associated with multiple comparisons, heterogeneous patient population with mixed oropharyngeal and OC-SCC, and no control for important confounding risk factors for OC-SCC. In our study, we have compared the oral microbiome in negative controls, in nonsmoking patients with oral premalignant lesions and nonsmoking patients with OC-SCC who lack HPV infection.

Methods

Supplementary methods are shown in supplementary material.

Human subjects.

A case-control study was approved by the Institutional Review Board of Memorial Sloan Kettering Cancer Center (IRB 15–256). Written informed consent was obtained from each participant. All participants had no recent history of using tobacco products, including 27

who never smoked and 11 ex-smokers who have not smoked for at least 11 years (average 21 years, range 11–38 years). Alcohol consumption was defined according to the previously described criteria[15], with four participants across both cases and controls classified as heavy drinkers (Supplementary Table S1). The cases included two groups: oral cavity squamous cell carcinoma (OC-SCC) and premalignant lesions (PML) based on histopathological examination. All premalignant lesions had leukoplakia with dysplasia confirmed on pathological analysis. The negative controls comprised patients with thyroid nodules (benign or malignant). These patients were deemed to be a representative "normal" population because they attended our head and neck clinics and had received a complete head and neck examination including flexible laryngoscopy or mirror laryngoscopy of the laryngopharynx to show no evidence for oral cavity pathology. All these patients have been followed and no patient has developed any oral cancer. All patients with OC-SCC were negative for high risk HPV infection. This was determined by negative staining by p16 immunohistochemistry. p16 immunohistochemistry is now used as the surrogate marker for high risk HPV infection. Pathological stage of OC-SCC was determined by histopathological examination at the time of surgical resection (Supplementary Table S1). All oral rinse specimens were taken prior to surgical resection of the oral cavity cancer or premalignanat lesion. Patients were followed for up to 22.2 months or until death for recurrence.

Oral sample collection and processing.

To sample the oral microbiome, the participants rinsed the mouth vigorously with 10 ml sterile saline for 30 seconds and spit, and bacteria were recovered from the rinse liquid by centrifugation at 3,220g for 20min. The pellets were transferred into 2ml tube and stored at −80° freezer for further study.

Detection of HPV in patients with OC-SCC.

All patients with OC-SCC were negative for high risk HPV infection as evaluated using immunohistochemical stain for p16INK4a expression in tumor cells and PCR using HPV specific primers in all the saliva samples. [16].

DNA extraction and 16S rRNA gene library preparation.

Total genomic DNA was extracted from the specimens using a modified QIAGEN DNA extraction method (QIAGEN, Germantown, MD), as previously described[17]. From the extracted DNA, the V3 and V4 regions of bacterial16S rRNA genes were amplified using the primer set 347F 5′-GGAGGCAGCAGTRRGGAAT and 803R 5′-CTACCRGGGTATCTAATCC, which we previously designed[18]. PCR reactions were carried out as previously described[17]. PCR products were purified using Agencourt AMPure XP (Beckman Coulter Life Sciences, IN) and quantified using Qubit 2.0 Fluorometer (Life Technologies, CA). Amplicon libraries were pooled at equimolar concentrations and sequenced on 454 FLX platform. Two types of negative controls were included to detect exogenous contamination: (i) Template negative controls (PCR reaction without DNA template) along with the samples while constructing 16S PCR amplicon libraries; (ii) Specimen-negative controls (use PBS in lieu of a human specimen) along with the samples while extracting DNA from the human samples.

Taxonomic classification.

Using the QIIME pipeline[19], reads were demultiplexed and filtered using default parameters with quality score 25. The sequences were grouped into operational taxonomic units (OTUs) and were classified to taxonomic levels from the phylum to genus levels. OTUs that were unclassifiable to specific taxa at these levels were excluded from further analysis (Table S4).

α-and β-diversities.

α-diversities (within subject diversity) were calculated by Monte Carlo permutations using compare_alpha_diversity.py, a built-in function in the QIIME pipeline, which included taxa observed, taxa predicted (Chao1), and Shannon with richness and evenness indices. Variation in alpha-diversity measures among three phenotypes were tested using ANOVA and between two phenotypes by t-test. β-diversity (between-subject diversity) was analyzed using weighted and unweighted UniFrac distances matrices, visualized in principal coordinate plots[20], statistically tested Adonis test from a built-in function in the QIIME pipeline.

Taxonomic analysis of differences between diseases and control.

Comparisons among and between the case control groups were mainly performed at the genus level. To reduce false discoveries due to noise, genera with very low abundance (<0.01%) were excluded from the comparisons (Table S4). The algorithm random forest (RF) was used as a feature selection model, which classifies the three case control groups by constructing 2,000 of decision tress by randomly sampling genera in the samples as predicators[21]. Out-of-bag (OOB) error rate from RF was used to rank the importance of genera[22]. To identify the most predictive genera alone or in combination and avoid model bias, Support Vector Machine (SVM) was applied as a classification model[23]. Leave-one-out error of SVM was calculated and classification accuracy (one minus "leave-one-out error") was used as the criteria to choose optimal combination of genera leading to maximal accuracy[24]. We put all combinations of top 10 important genera estimated by RF as features into SVM and calculated their classification accuracy. Finally, we identified the combinations achieving the peak accuracy. All these analyses were performed in MATLAB_R2017a. All genera that led to peak classification accuracy were selected for statistical comparisons among and between the case control groups while those caused reduction in the accuracy after the peak were excluded. Comparison of median relative abundance of a particular taxon among the three case control groups was done with nonparametric Kruskal Wallis test and between two groups with Mann-Whitney U test. The change in abundance of a taxon along the sequence of negative controls→PML→OC-SCC was analyzed using Jonckheere's trend test[25, 26]. All statistical tests were two-sided, with a p value <0.05 considered of nominal statistical significance. A false discovery rate (FDR)-adjusted p value (q value) <0.10 was used as the threshold for significance, as previously described[27], after adjustment for multiple comparisons for Kruskal Wallis test and Mann-Whitney U test in the follow up analysis. All statistical tests were conducted using R version 3.2.1.

Classification of bacterial communities into microbiome types by cluster analysis.

The bacterial communities sampled were classified into microbiome types based on community structures determined by genera composition and their relative abundances. Using genera that were significantly different among the three case control groups, a Euclidean metrics between all possible pairs of samples were calculated and used to cluster the samples by Ward's linkage algorithm. Major clusters generated were termed microbiome types. The significance of the association between the microbiome types and the disease states was evaluated by two-tailed Fisher's exact test.

Microbial network analysis and visualization.

Networks among genera were determined by correlation analysis[28]. A correlation matrix was generated by calculating the pairwise Spearman's rank correlations between all genera whose average relative abundance was higher than 0.01% across all samples, plus the three disease states. To allow quantitative correlation with bacterial genera, negative controls was weighted as 0, PML as 1 and OC-SCC as 2. Two genera were considered statistically correlated if the absolute value of correlation coefficient was >0.3 and the p value was <0.05. The network included primary genera that directly correlated to the disease states and secondary genera correlated to the primary genera. Correlation analyses and matrix were generated in R environment by corrplot[29], igraph[30] and Hmisc[31] packages. Visualization of the network was performed in the platform of Gephi[32].

Prediction of functional difference of oral microbiome in diseases and controls.

The functional potentials of the oral microbiome were inferred from the taxonomic compositions determined by the 16S rRNA gene survey using PICRUSt (v1.1.1)[33]. Operational taxonomic units (OTUs) were picked through closed-reference against the Greengenes (v13.5) at the 97% identity to create an OUT. The OUT 16S abundance was first normalized to the genome abundance using the "normalize_by_copy_number.py" command in the PICRUSt and metagenomics functions were then predicted according to the Kyoto Encyclopedia of Genes and Genome (KEGG) Orthology (KO)[34, 35] by the "predict_metagenomes.py" command in PICRUSt. One sample was excluded as its large Nearest Sequenced Taxon Index value was >0.15, representing low accuracy of the predicted KEGG groups. q value was adjusted by Benjamini–Hochberg FDR multiple test[36].

Results

To search for new risk factors of OC-SCC in nonsmokers, we conducted a case control study with 18 cases of OC-SCC, 8 cases of PML and 12 negative controls. The cases and the controls differed significantly by age but not by sex and race (Supplementary Table S1). All participants never smoked or were free of tobacco use for at least 11 years. Groups varied but did not statistically differ in alcohol consumption. All cancer patients were negative for high risk HPV. Their oral microbiome was sampled by oral wash and defined by 16S rRNA gene sequencing. This assay of the oral microbiome identified 12 phyla, 21 classes, 35 orders, 66 families, 116 genera and 172 species.

Oral microbiome is globally altered in OC-SCC.

We first assessed α-diversity of the oral microbiome and found that no significant difference of the disease groups from the control, as measured by observed OTUs; both Chao1 estimate for total species richness, and Shannon diversity index for both richness and evenness showed $p>0.05$ (Supplementary Table S2). In contrast, β-diversity by principal coordinates analysis revealed an overall separation among the three groups (Supplementary Table S3) (Figure 1A) using weighted UniFrac distance metrics [37]. Follow-up tests indicated that the differences were significant between OC-SCC and controls (Figure 1A) but not between PML and negative controls or between OC-SCC and PML. These findings suggest that the global alteration of the oral microbiome was due to changes in the abundance of certain taxa in OC-SCC.

Periodontal pathogens were enriched and commensal bacteria depleted along the sequence of controls→PML→OC-SCC regardless of status of heavy alcohol consumption.

To identify major bacterial taxa that were responsible for the global alteration of the oral microbiome , we filtered out very low abundant (<0.01%) genera from the dataset (Supplementary Table S4) and ranked the remaining 50 genera by the importance in discriminating the three groups using the algorithm Random Forest[21]. With the importance rank, we evaluated the maximum classification accuracy (MCA) achieved by selecting the optimal combination of the top ten most important genera using Support Vector Machine method. We achieved a maximal 73.7% classification accuracy among all three groups with six genera including *Prevotella, Alloprevotella, Veillonella, Actinomyces, Kingella,* and *TG5.* We further performed statistical analyses to compare their relative abundance among the three groups with Kruskal-Wallis test followed by correction for false discovery by Benjamini–Hochberg procedure[36]. *Alloprevotella* was the only genus significantly different among the three groups (Supplementary Table S5).

In pairwise analyses, we achieved 95% MCA between controls and PML with five genera, 90% MCA between negative control and OC-SCC with seven genera, and 88.5% MCA between PML and OC-SCC with three genera (Supplementary Table S5). Follow-up testing with Mann-Whitney U test revealed five genera statistically different between the controls and disease groups but none between the PML and OC-SCC groups after FDR adjustment (Figure 1B). Specifically, *Fusobacterium* (3.15% vs. 1.37%) and *Veillonella* (14.86% vs. 5.72%) were significantly more abundant in PML than in negative controls while *Alloprevotella* (2.31% vs. 0.13%), *Fusobacterium* (3.88% vs. 1.37%), and *Prevotella* (14.01% vs. 6.21%) were significantly more abundant and *Streptococcus* (18.09% vs. 34.63%) less abundant in OC-SCC than in controls (Supplementary Table S5). All the statistical conclusions remained unchanged after adjustment for alcohol use and age in linear regression models except *Streptococcus* whose enrichment in OC-SCC depended on age but not alcohol consumption (Supplementary Table S6).

In trend analyses, three of the five genera, all being periodontal pathogens, were progressively enriched along the sequence of controls→PML→OC-SCC, including *Prevotella* (6.21%→13.61%→14.01%), *Alloprevotella* (0.13%→1.50%→2.31%), and

Fusobacterium (median 1.37%→3.15%→3.88%) while commensal *Streptococcus* (34.63%→22.30%→18.09%) progressively decreased along the sequence (Figure 1C).

In PCA, reduction of all 50 genera to only the five genera resulted in a better separation among controls, PML, and OC-SCC (Figure 1D) and between the controls and OC-SCC (Figure 1D) but not between the controls and PML.

Samples of oral microbiome can be classified into two microbiome types.

Based on relative abundance of the five genera which differentiate the disease states, samples of the oral microbiome can be clustered into two distinct groups (Figure 2A), which we termed periodontal pathogen-low (PPL) and periodontal pathogen-high (PPH) types of microbiome, respectively. The two microbiome types were well separated on PCA (Figure 2B) using weighted UniFrac distance metrics. The PPL microbiome was characterized by higher abundant *Streptococcus* (median 38.67% vs. 18.72%), while the PPH microbiome by higher abundance of the three periodontal pathogens (23.31% vs. 6.26%) including *Prevotella* (16.81% vs. 4.50%), *Fusobacterium* (4.34% vs. 1.64%), and *Alloprevotella* (2.16% vs. 0.12%) (Figure 2C). Additionally, *Veillonella* (10.93% vs. 3.11%) was also significantly higher in the PPH microbiome. The microbiome in the large majority (66.67%, 8/12) of the control samples was classified as the PPL microbiome, while 88.46% (23/26) of the diseased samples (7/8 of PML samples and 16/18 OC-SCC samples) as the PPH microbiome (Odds ratio: 15.33, 95% Confidence Intervals: 2.80–83.89, p=0.0011, Fisher exact test, two tailed) (Figure 2D).

Periodontal pathogens and nonpathogens were collaborative among themselves but inhibitive to each other.

To explore the relationships among bacteria in the oral microbiome, we performed network analysis between the disease states and the 50 genera in the oral microbiome using pairwise Spearman's rank correlations with r>0.3 and p<0.05 considered significant (Figure 3A). To allow quantitative correlation of the disease states with the relative abundances of microbes, we weighted the control as 0, PML as 1, and OC-SCC as 2. The network formed under these conditions was composed of 11 known periodontal pathogens (*Fusobacterium, Prevotella, Alloprevotella, Camplylobacter, Parviomonas, Peptostreptococcus, Porphyromonas, Tannerella, Enterococcus, Selenomonas,* and *Dialister*) and 11 nonpathogens based on the recent classification of periodontal pathogens[38]. The disease states were directly correlated to four primary genera, including three positive (collaborative) connections to periodontal pathogens *Fusobacterium, Prevotella*, and *Alloprevotella*, and one negative (inhibitive) connection to commensal bacterium *Streptococcus*. Each of the four primary genera formed a cluster by networking with several secondary genera. Among the four clusters, the 11 connections between the *Streptococcus* cluster and the three pathogen clusters were all inhibitive while all 5 connections among the three pathogen clusters were collaborative. Overall, of a total of 33 connections, 19 were collaborative while 14 were inhibitive. 84.21% (16/19) of the collaborative connections were between two pathogens (n=13) or between two non-pathogens (n=3) while 92.86% (13/14) of the inhibitive connections were between a pathogen and a nonpathogen (p=0.000013, Fisher exact test, two tailed). Promoted by the significant association of periodontal pathogens with the disease states, we combined all 14

periodontal pathogens found in the samples and found a progressive increase in their relative abundance along the sequence of controls-PML-OC-SCC (median 9.96%→20.90%→27.84%). Pairwise comparison showed that periodontal pathogens were more abundant in OC-SCC than in controls (Figure 3B).

Proinflammatory pathways were progressively enriched along the controls→PML→OC-SCC sequence.

To explore functional differences among the three groups, we deduced microbial gene contents in the samples from 16S rRNA gene profiles using PICRUSt[33]. To minimize false discovery, we only concentrated on the genes whose abundance showed progressive changes along the controls→PML→OC-SCC sequence by Jonckheere trend test with r value >0.3 and q value <0.1 after adjustment for multiple comparisons. Overall, 132 KEGG proteins met these criteria, representing 102 different pathways (KO_C) and 34 processes (KO_B) and 7 biological categories (KO_A) (Supplementary Table S7). In particular, genes encoding heat shock protein 90 (HSP90) and ligands for Toll Like Receptor 1 (TLR1) (lipoproteins: pal, bamD) and enzymes for synthesis of ligands for TLR2 (peptidoglycans: amiA/B/C, vanX, mltD/dniR, mrdA) and TLR4 (lipopolysaccharides, LPS: lpxA, lpxB, lpxC, lpxD, lpxK, kdsA, kdsD/F, kdtA/waaA) were progressively enriched along the disease sequence (Figure 4A). *Prevotella* was the major contributor to the enrichment of these proinflammatory genes, with additional significant contributions from *Alloprevotella*, *Fusobacterium*, *Veillonella,* and *Porphyromonas*.

Capnocytophaga was associated with recurrence of OC-SCC.

In the 18 patients with OC-SCC, cancer recurred in 10 patients after resection. *Capnocytophaga*, a putative periodontal pathogen, was significantly more abundant in patients with recurrence (median 1.54% vs. 0.27%) (Figure 4B). In contrast, no taxa were associated with primary tumor stage, lymph node metastasis, or remote metastasis.

Discussion

Increased data indicating the association of the microbiome with cancer has triggered interest in the analysis of the oral microbiome in oral cancer [39]. A recent study by Lee et al reported a significant difference in bacterial genera between premalignant lesions and oral cancer with 5 genera *Bacillus, Enterococcus, Parvimonas, Peptostreptococus* and *Slackia* [40]. Another recent study by Hsiao et al also reported an association between 3 species of periodontopathogenic bacteria *Prevotella tannerae, Fusobacterium nucleatum, Prevotella intermedia* and oral cancer risk [41]. This study showed the association was correlated to increased use of cigarette smoking, betal quid use and also poor oral hygiene. However, these studies did not specifically examine changes in the microbiome in patients who did not smoke. As many host factors may confound microbiome changes, we specifically minimized the impact of tobacco use and HPV by enrolling only HPV-negative nonsmokers in this study. In addition, because the oropharyngeal subtype of oral cancer is mostly caused by high-risk HPV, we designed our study to focus only on the oral cavity subtype of oral cancer OC-SCC in which the large majority is unrelated to HPV infection. To eliminate false discoveries, we statistically corrected false discoveries in multiple statistical comparisons.

We found that the oral microbiome was globally altered in patients with OC-SCC due to enrichment of periodontal pathogens *Fusobacterium, Prevotella, Alloprevotella* and depletion of commensal *Streptococcus* in OC-SCC. In addition, *Fusobacterium*, and *Veillonella* were more abundant in PML than in controls. Further analysis demonstrated these changes were independent of alcohol consumption despite age as a confounding factor for the depletion of *Streptococcus*. Based on these marker genera, the oral microbiome can be classified into two types of microbiome, periodontal pathogen-low and periodontal pathogen-high. This classification had >80% accuracy in predicting PML and OSC-SCC. Beyond the three periodontal pathogens, the combined abundance for all 14 periodontal pathogens found in the samples progressively increased along the sequence of negative controls-PML-OC-SCC sequence and the pathogens were approximately three times as abundant in OC-SCC as in negative controls. Network analysis revealed collaborative relationships among the periodontal pathogens or among commensals per se but inhibitive relationships between pathogens and commensals. These findings consistently suggest periodontal pathogens are an independent risk factor in subjects lacking major risks of OC-SCC.

Periodontal pathogens are well known for their ability to cause chronic inflammation in periodontitis. To explore potential mechanisms, we deduced microbial gene and pathway contents based on 16S rRNA genes found in the samples. Overall, the microbiome showed a progressive enrichment of genes encoding HSP90A and ligands for TLR1 (lipoproteins) and enzymes for synthesis of ligands for TLR2 (peptidoglycans) and TLR4 (LPS) along the disease sequence. Periodontal pathogens were the major source for the enrichment of these genes. HSP90 is a new target for cancer therapy because it regulates a variety of cellular processes and many of its client proteins are oncogenic drivers that can regulate tumor intrinsic pathways, steroid hormone signaling, immunity and inflammation [42]. Because HSP90 functions in a highly conserved macromolecular complex in eukaryotes [43], whether bacterial HSP90 is able to interfere with the interaction between human HSP90 and its clients needs to be experimentally evaluated. TLRs are a type of pattern recognition receptor and recognize molecules that are broadly shared by pathogens including bacterial lipoproteins, peptidoglycans, and LPS. Activation of TLRs causes production of inflammatory cytokines via activation of NFκB. While triggering TLRs 5 and 9 promotes oral cancer growth, the roles of TLR1, 2, and 4 in OC-SCC have not been sufficiently studied [44]. A recent study by Hsiao et al, reported that a strong association between 3 species of periodontopathogenic bacteria: *Prevotella tannerae, Fusobacterium Nucleatum, Prevotella intermedia* and oral cancer risk [41]. However, this association was not present in patients who had genetic polymorphisms in TLR2 and TL4R suggesting these polymorphisms were protective to patients by inducing less inflammation. Further evidence for the association between inflammation induced by bacteria and the development of oral cancer was increased levels of salivary IL1β, an inflammatory cytokine found in the patients with these 3 bacterial species. These recent reports along with our own findings open new avenues to explore whether inflammation cause OC-SCC or play a secondary role in enhancing its malignancy.

The association of *Capnocytophaga* with OC-SCC recurrence is a novel finding. The median abundance of *Capnocytophaga* was 5.62 fold higher in patients with recurrence of OC-SCC after tumor resection than in controls. *Capnocytophaga* is an opportunistic pathogen

following pet bites[45] and a putative periodontal pathogen[46]. However, knowledge about its pathogenesis is very limited.

Our study is not without its limitations. Our sample size of the oral cavity and premalignant cohorts is small and therefore our results should be corroborated on a larger group of patients. Although our control cohort had no evidence of PML or OC-SCC it is possible that these patients may have a different oral hygiene habit to patients with PML or OC-SCC and therefore differences in the microbiome identified may be due to the differences in the oral hygiene of our patients. Unfortunately we did not have any data on the oral hygiene of our patients and this would need to be addressed in future studies. We have assumed that all cases of PML and OC-SCC do not harbor high risk HPV infection based upon p16 immunohistochemistry which is used as a surrogate marker for HPV related oropharyngeal cancer. However, it is possible that in oral cancer, p16 immunohistochemistry may not be quite as reliable as in oropharyngeal cancer as a marker for HPV positivity. This was recently suggested in a report by Lechner et al[47]. We also used 16S rRNA gene sequencing of the V3 and V4 regions of bacterial16S rRNA genes to identify the bacteria in the oral rinse specimens. New next generation sequencing techniques using metagenomics analysis of WGS sequences generated by lllumina HiSeq are now available which can give a more comprehensive identification of bacteria in oral rinse specimens.

Our study has important public health implications. Pre-diagnostic bacteria informative of risk of OCSCC, if identified, could potentially be used in clinical practice for more efficient screening and early detection of OCSCC. Furthermore, these patients could be treated with antibiotics or probiotic agents to prevent cancer. Specific bacteria types could also be incorporated into an existing vaccine to extend the spectrum of protection to include prevention of OCSCC.

In conclusion, our findings suggest periodontal pathogens are associated with OC-SCC in patients who lack risk factors of HPV and smoking. Microbiome-mediated inflammation may be responsible for OC-SCC in these patients. Our study does suggest that further studies are needed to determine whether bacterial HSP90 or TLR ligands contribute to OC-SCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

References

1. Wen BW, Tsai CS, Lin CL, et al. Cancer risk among gingivitis and periodontitis patients: a nationwide cohort study. Qjm-an International Journal of Medicine 2014;107: 283–290. [PubMed: 24336850]

2. Tezal M, Sullivan MA, Reid ME, et al. Chronic periodontitis and the risk of tongue cancer. Archives of Otolaryngology - Head and Neck Surgery 2007;133: 450–454. [PubMed: 17515503]

3. Zheng TZ, Boyle P, Hu HF, et al. Dentition, oral hygiene, and risk of oral cancer: a case-control study in Beijing, People's Republic of China. Cancer Causes and Control 1990;1: 235–241. [PubMed: 2102296]

4. Michaud DS, Fu Z, Shi J, Chung M. Periodontal Disease, Tooth Loss, and Cancer Risk. Epidemiologic Reviews 2017;39: 49–58. [PubMed: 28449041]

5. Hashim D, Sartori S, Brennan P, et al. The role of oral hygiene in head and neck cancer: results from International Head and Neck Cancer Epidemiology (INHANCE) consortium. Annals of Oncology 2016;27: 1619–1625. [PubMed: 27234641]

6. Chang JS, Lo HI, Wong TY, et al. Investigating the association between oral hygiene and head and neck cancer. Oral Oncology 2013;49: 1010–1017. [PubMed: 23948049]

7. Montero PH, Patel PD, Palmer FL, et al. Changing trends in smoking and alcohol consumption in patients with oral cancer treated at Memorial Sloan-Kettering Cancer Center from 1985 to 2009. Archives of Otolaryngology - Head and Neck Surgery 2012;138: 817–822. [PubMed: 22986714]

8. Weatherspoon DJ, Chattopadhyay A, Boroumand S, Garcia I. Oral cavity and oropharyngeal cancer incidence trends and disparities in the United States: 2000–2010. Cancer Epidemiol 2015;39: 497–504. [PubMed: 25976107]

9. Dwyer-Lindgren L, Flaxman AD, Ng M, Hansen GM, Murray CJ, Mokdad AH. Drinking Patterns in US Counties From 2002 to 2012. American Journal of Public Health 2015;105: 1120–1127. [PubMed: 25905846]

10. Mehanna H, Beech T, Nicholson T, et al. Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer--systematic review and meta-analysis of trends by time and region. Head and Neck 2013;35: 747–755. [PubMed: 22267298]

11. Isayeva T, Li Y, Maswahu D, Brandwein-Gensler M. Human papillomavirus in non-oropharyngeal head and neck cancers: a systematic literature review. Head Neck Pathol 2012;6 Suppl 1: S104–120. [PubMed: 22782230]

12. Zhao H, Chu M, Huang Z, et al. Variations in oral microbiota associated with oral cancer. Sci Rep 2017;7: 11773. [PubMed: 28924229]

13. Wolf A, Moissl-Eichinger C, Perras A, Koskinen K, Tomazic PV, Thurnher D. The salivary microbiome as an indicator of carcinogenesis in patients with oropharyngeal squamous cell carcinoma: A pilot study. Sci Rep 2017;7: 5867. [PubMed: 28725009]

14. Al-Hebshi NN, Nasher AT, Maryoud MY, et al. Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. Sci Rep 2017;7: 1834. [PubMed: 28500338]

15. Bouchery EE, Harwood HJ, Sacks JJ, Simon CJ, Brewer RD. Economic costs of excessive alcohol consumption in the U.S., 2006. American Journal of Preventive Medicine 2011;41: 516–524. [PubMed: 22011424]

16. Lindemann ML, Dominguez MJ, de Antonio JC, et al. Analytical comparison of the cobas HPV Test with Hybrid Capture 2 for the detection of high-risk HPV genotypes. J Mol Diagn 2012;14: 65–70. [PubMed: 22079749]

17. Yang L, Lu X, Nossa CW, Francois F, Peek RM, Pei Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. Gastroenterology 2009;137: 588–597. [PubMed: 19394334]

18. Nossa CW, Oberdorf WE, Yang L, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World Journal of Gastroenterology 2010;16: 4135–4144. [PubMed: 20806429]

19. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7: 335–336. [PubMed: 20383131]

20. Lozupone C, Hamady M, Knight R. UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 2006;7.

21. Breiman L Random forests. Machine Learning 2001;45: 5–32.

22. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006;7: 3. [PubMed: 16398926]

23. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learning 2002;46: 389–422.

24. Elisseeff A, Pontil M. Leave-one-out error and stability of learning algorithms with applications. NATO science series sub series iii computer and systems sciences, 2003:111–130.

25. Cochran W Some methods for strengthening the common chi-squared tests. Biometrics. International Biometric Society 1954;10 417–451.

26. Armitage P Tests for Linear Trends in Proportions and Frequencies. Biometrics. International Biometric Society 1955;11 375–386.

27. Zhernakova A, Kurilshikov A, Bonder MJ, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 2016;352: 565–569. [PubMed: 27126040]

28. Steinhauser D, Krall L, Mu¨ssig C, Bu¨ssis D, Usadel B. Correlation networks In: Junker BHSF, editor. Analysis of Biological Networks Hoboken, USA John Wiley & Sons, 2008:305–335.

29. Wei T, Simko V. Corrplot: Visualization of a Correlation Matrix

30. Csardi G, Nepusz T. The igraph software package for complex network research. International Journal of Complex Systems 2006;1695: 1–9.

31. Harrell FE Jr. Hmisc: Harrell Miscellaneous Available from URL: https://CRAN.R-project.org/package=Hmisc.

32. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. ICWSM 8 361–362. 2009.

33. Langille MG, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 2013;31: 814–821.

34. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research 2000;28: 27–30. [PubMed: 10592173]

35. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 2017;45: D353–D361. [PubMed: 27899662]

36. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995;57: 289–300.

37. Koh H, Blaser MJ, Li H. A powerful microbiome-based association test and a microbial taxa discovery framework for comprehensive association mapping. Microbiome 2017;5: 45. [PubMed: 28438217]

38. Perez-Chaparro PJ, Goncalves C, Figueiredo LC, et al. Newly identified pathogens associated with periodontitis: a systematic review. Journal of Dental Research 2014;93: 846–858. [PubMed: 25074492]

39. Bhatt AP, Redinbo MR, Bultman SJ. The role of the microbiome in cancer development and therapy. CA: A Cancer Journal for Clinicians 2017;67: 326–344. [PubMed: 28481406]

40. Lee WH, Chen HM, Yang SF, et al. Bacterial alterations in salivary microbiota and their association in oral cancer. Sci Rep 2017;7: 16540. [PubMed: 29184122]

41. Hsiao JR, Chang CC, Lee WT, et al. The interplay between oral microbiome, lifestyle factors and genetic polymorphisms in the risk of oral squamous cell carcinoma. Carcinogenesis 2018;39: 778–787. [PubMed: 29668903]

42. Chatterjee S, Burns TF. Targeting Heat Shock Proteins in Cancer: A Promising Therapeutic Approach. Int J Mol Sci 2017;18.

43. Schopf FH, Biebl MM, Buchner J. The HSP90 chaperone machinery. Nat Rev Mol Cell Biol 2017;18: 345–360. [PubMed: 28429788]

44. Pisani LP, Estadella D, Ribeiro DA. The Role of Toll Like Receptors (TLRs) in Oral Carcinogenesis. Anticancer Research 2017;37: 5389–5394. [PubMed: 28982847]

45. Butler T Capnocytophaga canimorsus: an emerging cause of sepsis, meningitis, and post-splenectomy infection after dog bites. European Journal of Clinical Microbiology and Infectious Diseases 2015;34: 1271–1280. [PubMed: 25828064]

46. Ciantar M, Gilthorpe MS, Hurel SJ, Newman HN, Wilson M, Spratt DA. Capnocytophaga spp. in periodontitis patients manifesting diabetes mellitus. Journal of Periodontology 2005;76: 194–203. [PubMed: 15974842]

47. Lechner M, Chakravarthy AR, Walter V, et al. Frequent HPV-independent p16/INK4A overexpression in head and neck cancer. Oral Oncology 2018;83: 32–37. [PubMed: 30098776]

Novelty and Impact:

Our findings are novel because they suggest that in patients who lack the major risk factors of oral cancer, periodontal pathogens are an independent risk factor. Our study suggests microbiome-mediated inflammation may contribute to the development of oral cancer possibly through HSP90 or triggering TLRs. Our study has important public health implications for more efficient screening and early detection of OCSCC. Such patients could be treated with antibiotics or probiotic agents to prevent cancer.
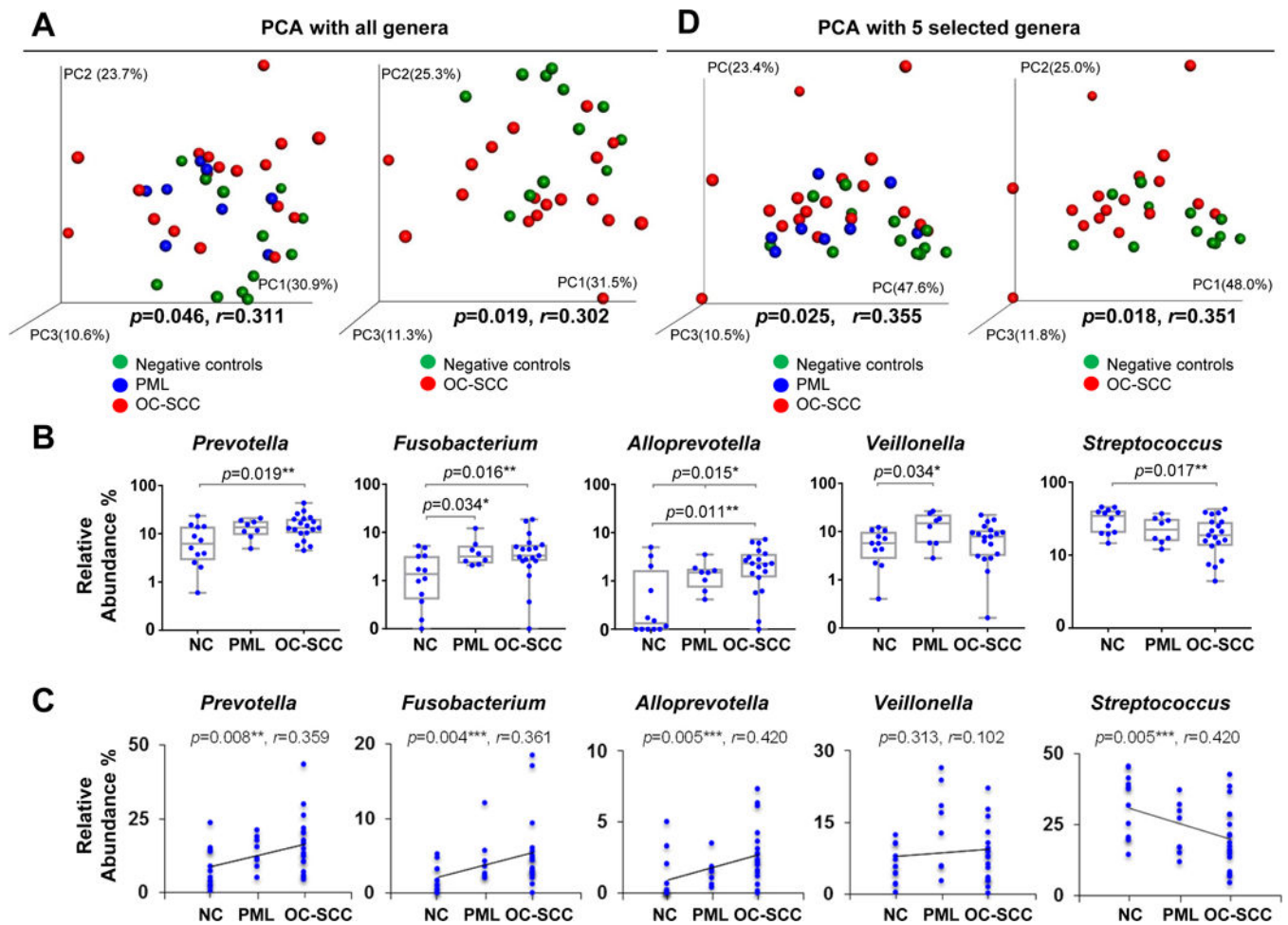
Figure 1. Selection of bacterial genera as markers to differentiate among and between negative controls, oral PML and oral cavity squamous cell carcinoma (OC-SCC).

Initial analysis was done with all genera by principal coordinate analysis for the global differences of microbiome with weighted UniFrac distances matrices. Comparisons were made among negative control, oral PML, and OC-SCC, between negative controls and OC-SCC (A). *P* values were calculated by Adonis test. Selective analyses were based on relative abundance of five genera (*Prevotella, Veillonella, Fusobacterium, Alloprevotella,* and *Streptococcus*) selected by Random Forest coupled with Support Vector Machine and statistical tests with Kruskal Wallis test for difference among three groups and Mann-Whitney test between two groups (B), *Jonckheere* trend test for trend (C) and principal coordinate analysis with weighted UniFrac distances matrices among the five genera (D). *P* value is marked with * if <0.1, ** if <0.05, or *** if <0.01 after FDR-adjustment.
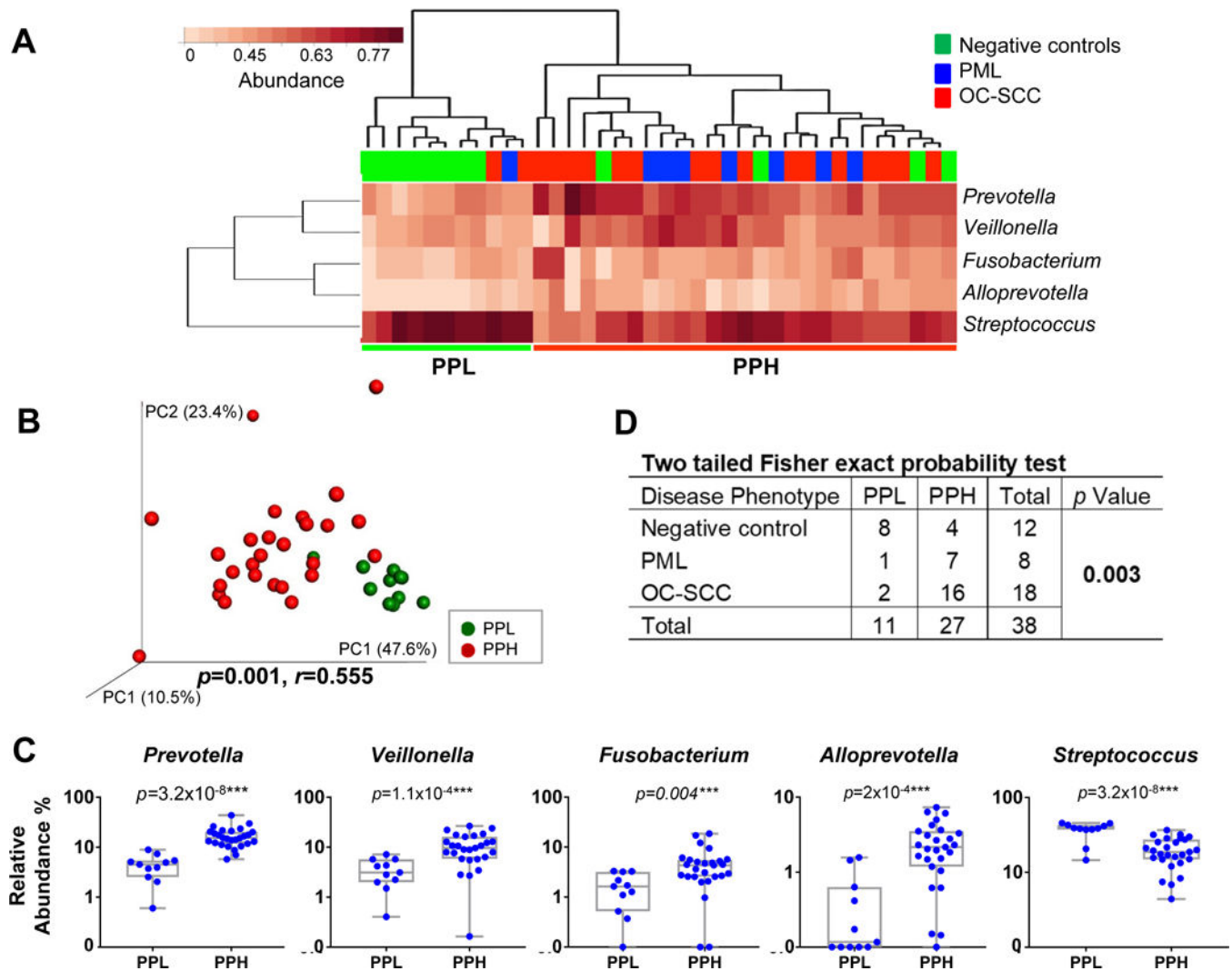
Figure 2. Classification of bacterial communities into microbiome types and their correlation with diseases.

The bacterial communities sampled were classified into two microbiome types based on the relative abundance of five genera informative of disease status as determined by statistical analyses, including *Prevotella, Veillonella, Fusobacterium, Alloprevotella,* and *Streptococcus.* A Euclidean metrics between all possible pairs of samples were calculated and used to cluster the samples by Ward's linkage algorithm (A). Major clusters generated were termed periodontal pathogen-low (PPL) and periodontal pathogen-high (PPH) types of microbiome, respectively. Separation between the two types of microbiome was further analyzed by principal coordinate analysis with weighted UniFrac distances matrices with *p* values calculated by Adonis (B). Correlation of the microbiome types with diseases was evaluated using taxonomic analysis by Mann-Whitney test (C) and two tailed Fisher exact test (D). *P* value is marked with * if <0.1, ** if <0.05, or *** if <0.01 after FDR-adjustment.
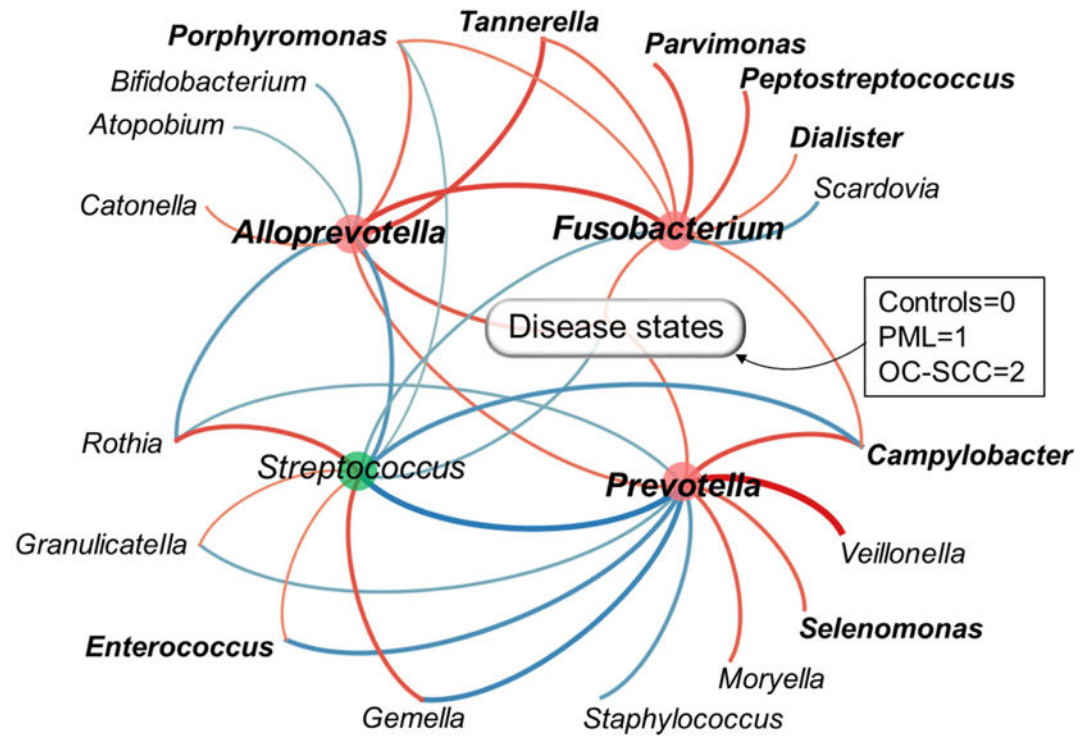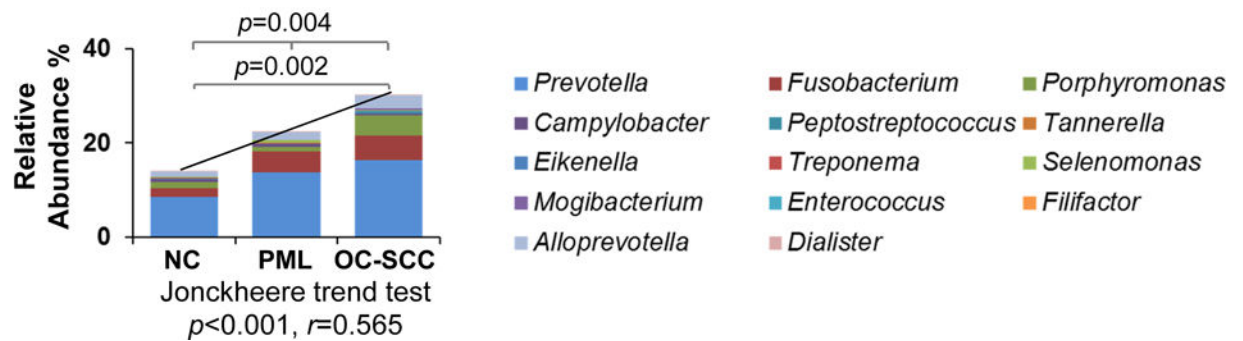
Figure 3. Co-occurrence/avoidance network among major genera in the oral microbiome and their correlations with disease states.

The four large nodes are genera that directly correlated with the disease states and the termini show genera that indirectly correlated with the disease states through the four genera (A). Connecting lines represent strong (r>0.3) and significant (p<0.05) correlations. Red lines indicate a positive (collaborative) correlation while blue lines indicate a negative (inhibitive) correlation. The thickness of a connection line is in proportion to the r value in Spearman's correlation coefficient. Known periodontal pathogens are bold-faced. Difference among negative controls, PML, and OC-SCC in the combined relative abundance of all periodontal pathogens were analyzed by with Kruskal Wallis test for difference among the three case control groups and Mann-Whitney test between two groups (B). Trend of changes in the combined relative abundance of all 14 periodontal pathogens found in the samples along the negative controls(NC)→PML→OC-SCC sequence was analyzed by Jonckheere trend test.
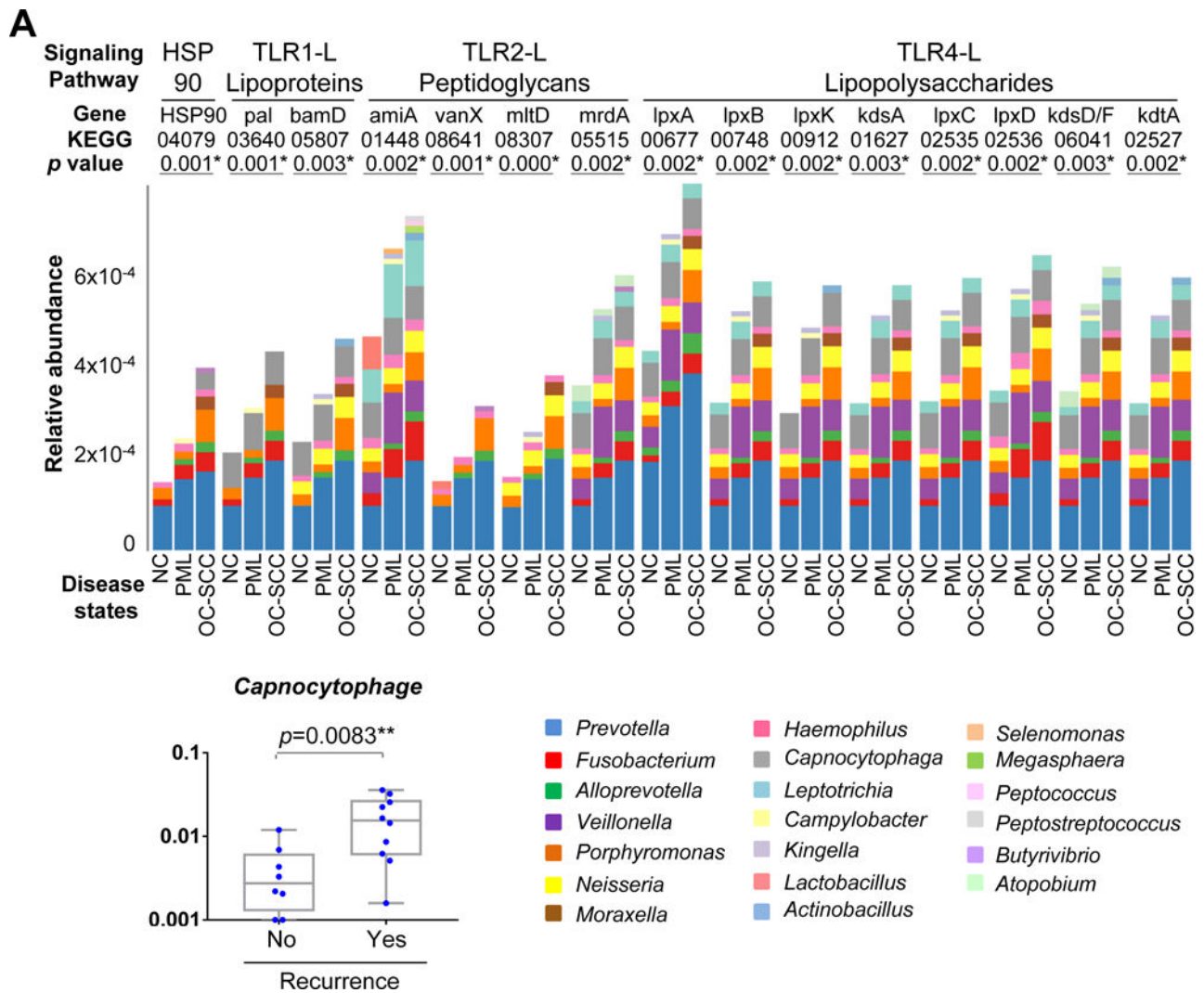
Figure 4. Progressive enrichment of proinflammatory genes/pathways along the sequence of negative controls→PML→OC-SCC and enrichment of *Capnocytophaga* in patients with OC-SCC.

Microbial gene contents in the samples were deduced from 16S rRNA gene profiles using PICRUSt (A). A gene shown is specified by name and KEGG ID. Relative abundance of each gene was stratified to source genera by color codes. Difference between patients with and without recurrence of OC-SCC after tumor resection was shown in relation to the relative abundance of *Capnocytophaga* (B). *P* value is marked with * if <0.1 or ** if <0.05 after FDR-adjustment.