Carcinogenesis, 2021, Vol. 42, No. 1, 127–135

doi:10.1093/carcin/bgaa062
Advance Access publication July 4, 2020
Original Article

ORIGINAL ARTICLE

Oral microbial dysbiosis and its performance in predicting oral cancer

Shih-Chi Su[1,2,3], Lun-Ching Chang[4], Hsien-Da Huang[5,6], Chih-Yu Peng[7,8], Chun-Yi Chuang[9,10], Yi-Tzu Chen[7,8], Ming-Yi Lu[7,8], Yu-Wei Chiu[7,8], Pei-Yin Chen[7,8] and Shun-Fa Yang[11,12,*]

[1]Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan, [2]Department of Dermatology, Drug Hypersensitivity Clinical and Research Center, Chang Gung Memorial Hospital, Linkou, Taiwan, [3]Central Research Laboratory, XiaMen Chang Gung Hospital, XiaMen, China, [4]Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, USA, [5]School of Life and Health Sciences and [6]Warshel Institute for Computational Biology, Chinese University of Hong Kong, Shenzhen, China, [7]School of Dentistry, Chung Shan Medical University, Taichung 40201, Taiwan, [8]Department of Dentistry, Chung Shan Medical University Hospital, Taichung 40201, Taiwan, [9]School of Medicine, Chung Shan Medical University, Taichung, Taiwan, [10]Department of Otolaryngology, Chung Shan Medical University Hospital, Taichung, Taiwan, [11]Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan and [12]Department of Medical Research, Chung Shan Medical University Hospital, Taichung, Taiwan

*To whom correspondence should be addressed. Institute of Medicine, Chung Shan Medical University, 110 Chien-Kuo N. Road, Section 1, Taichung 402, Taiwan. Tel: +886 4 24739595 ext. 34253; Fax: +886 4 24723229; Email: ysf@csmu.edu.tw

Abstract

Dysbiosis of oral microbiome may dictate the progression of oral squamous cell carcinoma (OSCC). Yet, the composition of oral microbiome fluctuates by saliva and distinct sites of oral cavity and is affected by risky behaviors (smoking, drinking and betel quid chewing) and individuals' oral health condition. To characterize the disturbances in the oral microbial population mainly due to oral tumorigenicity, we profiled the bacteria within the surface of OSCC lesion and its contralateral normal tissue from discovery (n = 74) and validation (n = 42) cohorts of male patients with cancers of the buccal mucosa. Significant alterations in the bacterial diversity and relative abundance of specific oral microbiota (most profoundly, an enrichment for genus *Fusobacterium* and the loss of genus *Streptococcus* in the tumor sites) were identified. Functional prediction of oral microbiome shown that microbial genes related to the metabolism of terpenoids and polyketides were differentially enriched between the control and tumor groups, indicating a functional role of oral microbiome in formulating a tumor microenvironment via attenuated biosynthesis of secondary metabolites with anti-cancer effects. Furthermore, the vast majority of microbial signatures detected in the discovery cohort was generalized well to the independent validation cohort, and the clinical validity of these OSCC-associated microbes was observed and successfully replicated. Overall, our analyses reveal signatures (a profusion of *Fusobacterium nucleatum CTI-2* and a decrease in *Streptococcus pneumoniae*) and functions (decreased production of tumor-suppressive metabolites) of oral microbiota related to oral cancer.

Introduction

Oral cancer is a common neoplasm worldwide, with a vast majority (~90%) of cases being oral squamous cell carcinoma (OSCC) (1). OSCC is known as a multifactorial malignancy where genetic factors are interwined with acquired risks to predispose to this disease (2). Diverse genetic variations that affect cell apoptosis, cell cycle and DNA repair (3) alone or in combination with environmental triggers, including cigarette smoking, alcohol drinking

Received: May 7, 2020; **Revised:** June 10, 2020; **Accepted:** June 16, 2020

Abbreviations

AUC	area under the receiver-operating characteristic curve
CI	confidence interval
OSCC	oral squamous cell carcinoma
OUT	operational taxonomic units

and betel quid chewing (4), have been recognized to mediate the pathogenesis of OSCC. In addition, other etiological parameters of oral cancer may involve but not limited to periodontitis, poor oral hygiene (5), and inflammation and chronic microbial infections (6). Despite the advances in therapeutic strategies, the death rate of OSCC has not ameliorated substantially in the past decades (7), mainly because of late diagnosis and frequent recurrences and/or metastases after primary treatment. Therefore, improvement on OSCC diagnosis, monitoring and prognosis is an unmet need.

The human mouth harbors highly populated bacterial communities (>700 bacterial species) (8) and possesses a large number of commonly shared microorganisms between unrelated subjects (9). Cumulative data have suggested a role for oral microbiome in oral cancer via production of carcinogens (e.g. cytolethal distending toxin, typhoid toxin and colibactin) and proinflammatory responses (10). Significant alterations in the levels of oral microflora (e.g. enrichment for genus *Lactobacillus* and the loss of genus *Haemophilus*, *Neisseria*, *Rothia*, *Leptotrichia* and *Aggregatibacter*) have been noted in OSCC cases (11). However, the composition of oral bacterial communities fluctuates by saliva and distinct sites of oral cavity (12). Moreover, betel quid chewing, tobacco and alcohol contain a number of chemical components, such as alkaloids, copper, catechol, reactive oxygen species (13,14), may influence the oral microbial population (15). Therefore, careful control for the distinction among oral habitats and inter-individual variations is necessary to characterize the shift in the composition of oral microbiota that is solely due to oral tumorigenicity.

Here, we aimed to properly explore OSCC-associated oral microbiome among male patients in Taiwan where both the age-standardized incidence for males and the ratio of male to female are among the highest in Asia (2,16). The bacteria within the surface of OSCC lesion and its anatomically matched contralateral region of normal tissue from each patient of both a discovery and a subsequent validation cohort were profiled through performing 16S rRNA amplicon sequencing of buccal swabs. Our analyses reveal signatures and functions of oral microbiota related to oral cancer. The clinical validity of these OSCC-associated microorganisms was studied and further validated in an independent cohort, implicating specific oral microflora as a valuable marker for prognosis monitoring of oral malignancies.

Materials and methods

Cohort recruitment and swab collection

Oral swabs of tumor lesions and their contralateral, normal regions from 74 male patients with cancers of the buccal mucosa who have been previously untreated were conducted in the Department of Otorhinolaryngology, with the approval of the institutional review board of Chung Shan Medical University Hospital, Taichung, Taiwan. The cases were recruited between 2016 and 2018. Patients with detectable periodontal inflammation, visible carious lesions or any history of diabetes mellitus or immune-associated disorders were ruled out as described previously (17). In addition, a validation cohort, including 42 OSCC patients bearing cancers of the buccal mucosa, was enrolled with the same inclusion and exclusion criteria in the Department of Oral and Maxillofacial Surgery, Chung Shan Medical University Hospital for verification. In both cohorts, participants were instructed to keep from dining, smoking, drinking or mouth hygiene activities for an hour prior to collection of buccal swabs. Swabs were collected using a modified protocol reported previously (18), placed in sterile, DNAse/RNAse-free tubes and stored at –20°C until further analysis.

16S rRNA gene amplicon sequencing and data processing

Prior to DNA isolation, each swab in 500 μl of lysis and DNA stabilization solution (LS solution, Isohelix), was vortexed and spun at 14 000 r.p.m. for 1 min to pellet the microbial populations. The detailed protocols for 16S rRNA gene sequencing and processing of sequence reads were described previously (19). In brief, bacterial genomic DNA was isolated with a Qiamp DNA Blood Mini kit (Qiagen) according to the manufacturer's instructions, and the variable region 4 (V4) of small subunit rRNA (16S rRNA) gene was PCR-amplified and used for library construction by the TruSeq DNA LT Sample Preparation Kit (Illumina). Purified libraries were quantified, normalized, pooled and applied for cluster generation and sequencing on a MiSeq instrument (Illumina).

Paired-end reads were processed (merging, quality filtering and removal of chimeric reads) using the QIIME 1.7 pipeline (20) to obtain effective tags (the processed sequencing reads), which were then clustered into operational taxonomic units (OTU) at 97% sequence identity using the UPARSE (21). Taxonomy classification was assigned according to the information retrieved from the SILVA database (22). Any sequence with one-time occurrence (singletons) or detected in only one sample was filtered out, and samples with less than 10^4 effective tags were excluded from further analyses. To evaluate the phylogenetic relationship of different OTUs, alignment of multiple sequences was conducted using the PyNAST software v.1.2 (23) against the dataset of the SILVA database, and a phylogenetic tree was generated with the FastTree (24). For estimating α diversity, species richness and evenness were evaluated by the Shannon index. For evaluating β diversity, the weighted and unweighted UniFrac parameters (25) were calculated by using the QIIME pipeline. Non-metric dimensional scaling was conducted using the weighted correlation network analysis, stat and ggplot2 packages in R software by transforming a distance matrix of weighted or unweighted UniFrac parameters among samples into a new set of orthogonal axes.

Functional composition of metagenomes was predicted from 16S rRNA data by the PICRUSt software (26), the pipeline of which is composed of two workflows, gene content prediction and metagenome prediction. A table of gene copy numbers for each gene family in each sequenced bacterial and archaeal genome based on the IMG database (27) and a phylogenetic tree from the Greengenes database (28) was precomputed for gene content prediction. Subsequently, metagenome prediction was performed through multiplying the vector of gene counts for each OTU by the abundance of that OTU in each sample, and summed across all OTUs.

Statistical analysis

Mann–Whitney *U*-test was used to compare demographic and clinical parameters between the discovery and validation cohort. Shannon index was compared by using Student's *t*-test. The discrimination in community composition between lesion and normal sites was determined by analysis of similarities of UniFrac parameters using 999 permutations in each test. Significant changes in the relative abundance of the taxa were analyzed at the genus level using Student's *t*-test and adjusted by Benjamini–Hochberg correction. Statistically significant biomarkers at OTU level were determined by the linear discriminant analysis of effect size analysis (29). Selected microbial biomarkers were trained on the discovery cohort using receiver-operating characteristic curves, and confusion matrix was constructed by the best threshold with the optimal sum of specificity and sensitivity, followed by replication of these microbial biomarkers using the validation cohort. A P value of <0.05 was considered significant.

Results

Cohort characteristics

To characterize the OSCC-associated oral microbiome, we collected and analyzed oral swabs of the lesion and its contralateral

region of normal tissue from each patient of a cross-sectional cohort with cancers of the buccal mucosa. This includes 74 male patients, with the mean age at disease onset being 53 (Supplementary Table S1, available at *Carcinogenesis* online). Of the cases enrolled in this project, 90.5%, 63.5% and 93.2% reported a history of cigarette smoking, alcohol drinking and betel quid chewing, respectively.

Oral microbial composition and diversity between the surfaces of OSCC lesions and their contralateral normal tissues

We first compared oral microbial diversity between the lesions and their contralateral region of normal tissue in patients with cancers of the buccal mucosa. An increase in within-sample bacterial richness (α diversity) was observed in those lesion sites (Figure 1A). Moreover, analyses of sample-to-sample differences in microbial community structures (β diversity) showed that the oral microbiome of the lesions clustered separately from that of the contralateral normal controls (analysis of similarity, $P = 0.001$) (Figure 1B). To characterize the oral microbial composition, we conducted a taxonomic analysis and detected a predominance of phylum *Firmicutes* (28.8%; i.e. 28.8% of the overall reads sequenced), *Proteobacteria* (25.9%), *Bacteroidetes* (24.6%) and *Fusobacteria* (11.4%), with less common presence of *Actinobacteria* (3.6%) on the surface of tumor lesions (Figure 1C). At the genus level, the most prevalent ones were *Streptococcus* (13.3%), *Neisseria* (12.3%), *Haemophilus* (7.7%), *Fusobacterium* (7.7%) and *Veillonella* (5.6%) (Figure 1D).

Bacterial taxa associated with OSCC lesions

Since disturbance in the relative abundance of oral microbiota was observed (Figure 1C and D), we sought for correlation of

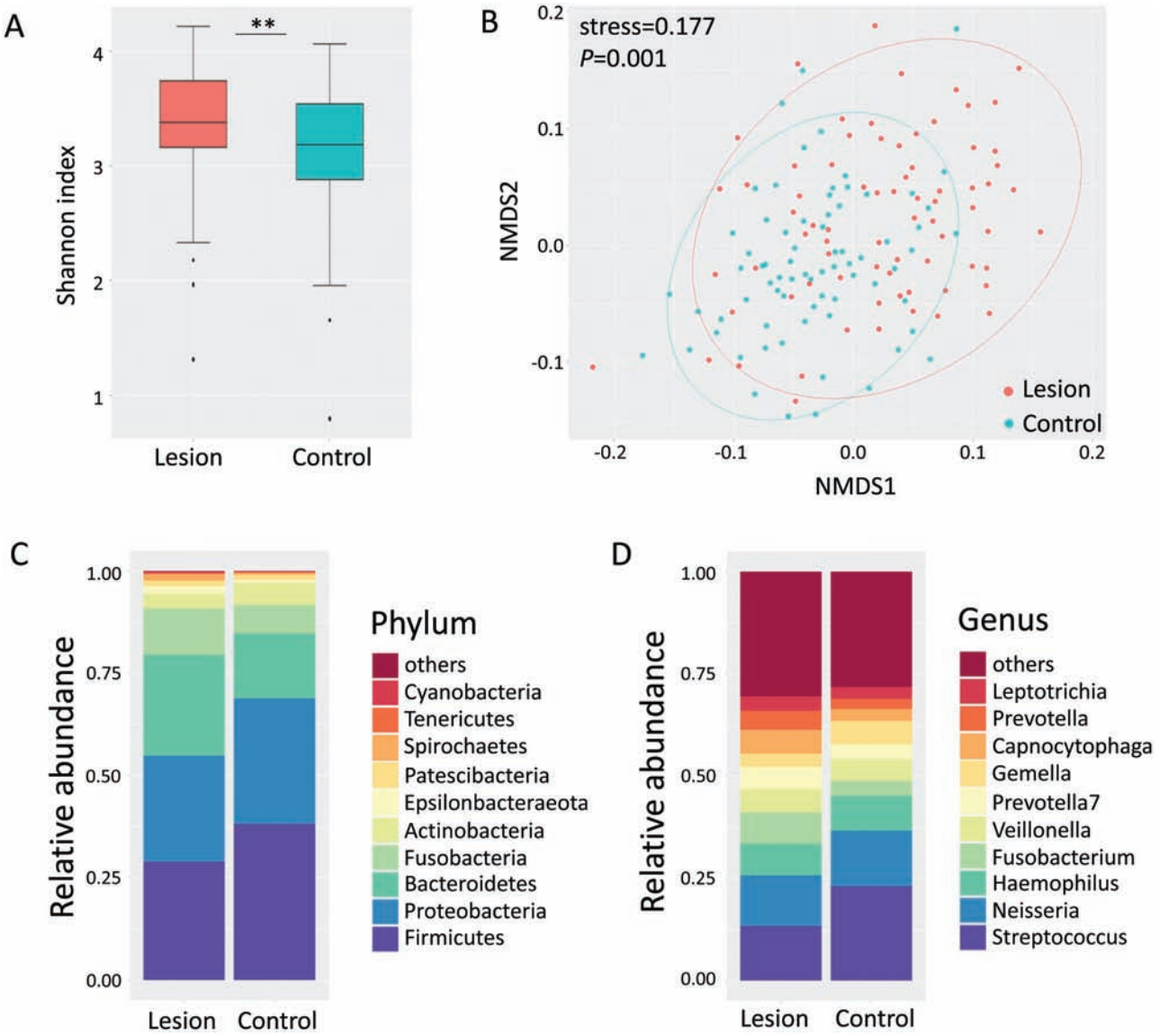

Figure 1. Comparisons of oral microbiota diversity and composition in OSCC lesions and contralateral normal tissues. **(A)** α Diversity (Shannon index) of oral microbial communities on the surface of OSCC lesions (Lesion) and their contralateral normal tissues (Control). The box-plot shows the median, the 25th percentile and the 75th percentile in each group. Shannon index was analyzed using Student's *t*-test. $^{**}P < 0.01$. **(B)** Non-metric multidimensional scaling ordination based on weighted UniFrac parameters of oral microbial communities in the lesion and control groups. Significant sample-to-sample dissimilarities refer to analysis of similarity ($P = 0.001$) test for discrimination in community composition between the lesion and control group. **(C** and **D)** The distribution of top 10 phyla (C) and top 10 genera (D) detected in the OSCC lesion and control groups.

specific oral microorganisms at the genus level with OSCC. Taking stringent criteria (higher than 0.1% in abundance and presence in more than 90% of swabs), six genera with significant differences in the levels between the OSCC lesions and their contralateral normal sites were identified as the core OSCC-associated microbiota (Figure 2). We found that tumor sites harbored higher levels of *Fusobacterium*, *Peptostreptococcus*, *Campylobacter*, *Prevotella* and *Capnocytophaga* than did their contralateral normal controls, whereas a decrease of *Streptococcus* spp. was detected in the surface of OSCC lesions. Additionally, further analyses were conducted to uncover bacterial taxa associated with oral cancer development. We predicted the biomarkers for the lesion and control group by considering both statistical significance and biological consistency with linear discriminant analysis of effect size. Among possible markers identified, specimens from the tumor sites exhibited a substantial enrichment for OTUs belonging to those genera of the core OSCC-associated microbiota aforementioned (e.g. enrichment for *Fusobacterium nucleatum CTI-2*, *Campylobacter concisus*, *Campylobacter* sp. *RM6137* and many *Peptostreptococcus* spp.) (Figure 3A). In contrast, a profusion of genus *Streptococcus* was seen in the normal tissues over the OSCC lesions. Furthermore, we utilized the Random Forests model on the overall microbiota profiles (Figure 3B) to determine OSCC-discriminatory taxa. Consistently, many OTUs assigned to genus *Streptococcus* were identified to be capable of distinguishing OSCC from the control, suggesting the predictive value of specific oral microbes for cancers of the buccal mucosa with prognostic implications.

Functional prediction of oral microbiota related to the development of OSCC

With a focus on pathways relevant to microbial metabolism, we found that pathway modules associated with metabolism of terpenoids and polyketides were differentially enriched between the OSCC lesions and their contralateral normal controls (Figure 4A). Specifically, genetic markers assigned to biosynthesis of siderophore group nonribosomal peptides, monoterpenoid biosynthesis and biosynthesis of 12-, 14- and 16-membered macrolides were less abundant in OSCC lesions when compared with the normal counterparts. Overall, our results present here reveal OSCC-associated alterations in the composition and functionality of oral microbiota.

Majority of microbial signatures replicates in an independent validation cohort

To further test the replicability, we evaluated the generality of oral microbial features identified earlier in an independent cohort of 42 patients with cancers of the buccal mucosa (Supplementary Table 1, available at *Carcinogenesis* online). Similar patterns, in terms of microbial composition and diversity, were observed between OSCC lesions and their contralateral normal tissues (Supplementary Figure 1, available at *Carcinogenesis* online). Of six genera identified as the core OSCC-associated microbiota in the discovery cohort, all trended in the same direction in the validation cohort and were false discovery rate significant, with most differential genera being *Fusobacterium* (enriched) and *Streptococcus* (decreased) (Supplementary Figure 2, available at *Carcinogenesis* online).

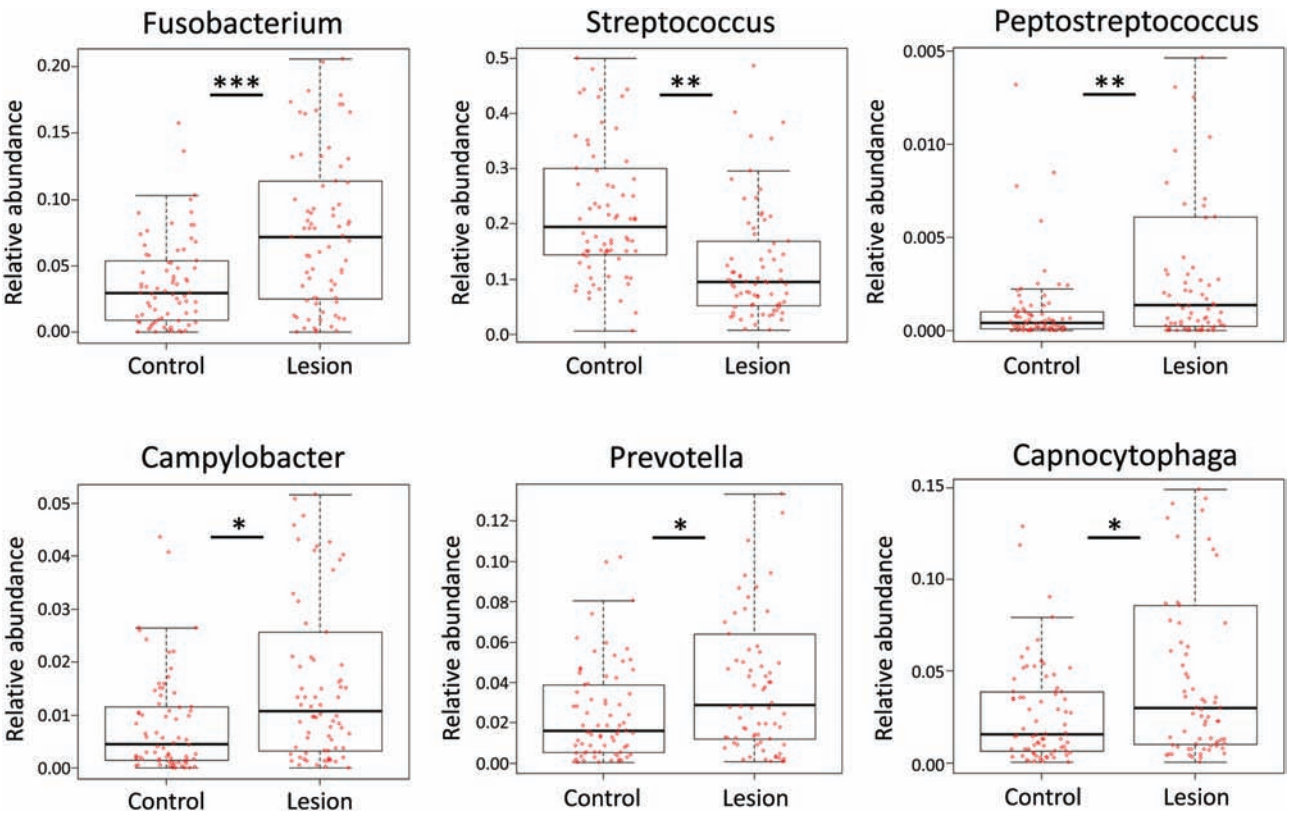

Figure 2. Differentially abundant genera in OSCC lesions and contralateral normal tissues. Relative abundance of the taxa at the genus level was compared using Student's *t*-test and adjusted by Benjamini–Hochberg correction. *$q < 0.05$; **$q < 0.01$; ***$q < 0.001$. The boxplots show the mean, the 25th percentile and the 75th percentile in each group.

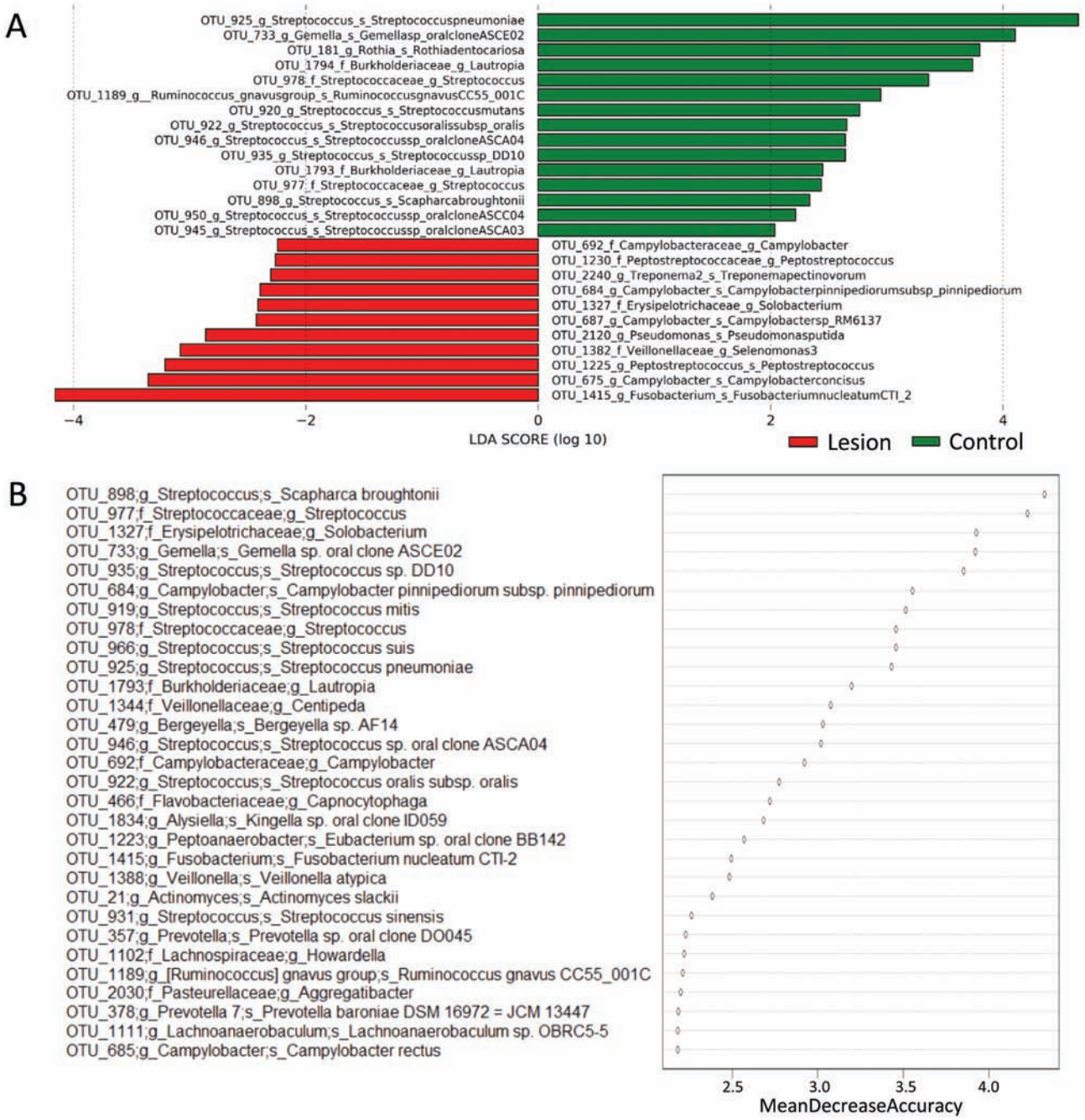

Figure 3. Determination of bacterial biomarkers in OSCC lesions compared with contralateral normal tissues. (**A**) Bacterial taxa that best characterize each group were identified by using linear discriminant analysis of effect size on OTU tables. (**B**) OSCC-discriminatory taxa were determined by applying Random Forests analysis using the overall OTU dataset against the disease status. Bacterial taxa that are most discriminatory were ranked in descending order of their importance to the accuracy of the model. Importance was determined based on the mean decrease in accuracy of microbiota prediction when the relative abundance of each taxon was randomly permuted.

Clinical validity of oral microbiota in discriminating OSCC

In spite of distinct analyses carried out above, a number of potential biomarkers consistently emerged from our exploration. The clinical validity of these OSCC-related oral microorganisms was therefore assessed by constructing receiver-operating characteristic curves for distinguishing OSCC from non-lesion controls (Table 1). Several bacterial taxa significantly separate histologically normal counterparts from OSCC lesions. Among them, fairly good area under the receiver-operating characteristic curve (AUC) in distinguishing the tumor lesions from the normal tissues was obtained for *Streptococcus* [AUC, 0.75; 95% confidence interval (CI), 0.67–0.83] and *Fusobacterium* spp. (AUC, 0.7; 95% CI, 0.61–0.78), and their performance was further replicated in the validation cohort (AUC, 0.85; 95% CI, 0.77–0.93 for the former; AUC, 0.88; 95% CI, 0.81–0.95 for the latter). Of note, at the lower taxonomical level, *Streptococcus pneumoniae* showed a satisfactory efficacy in distinguishing non-tumor controls from the lesion sites, with an overall AUC of 0.77 (95% CI, 0.71–0.83).

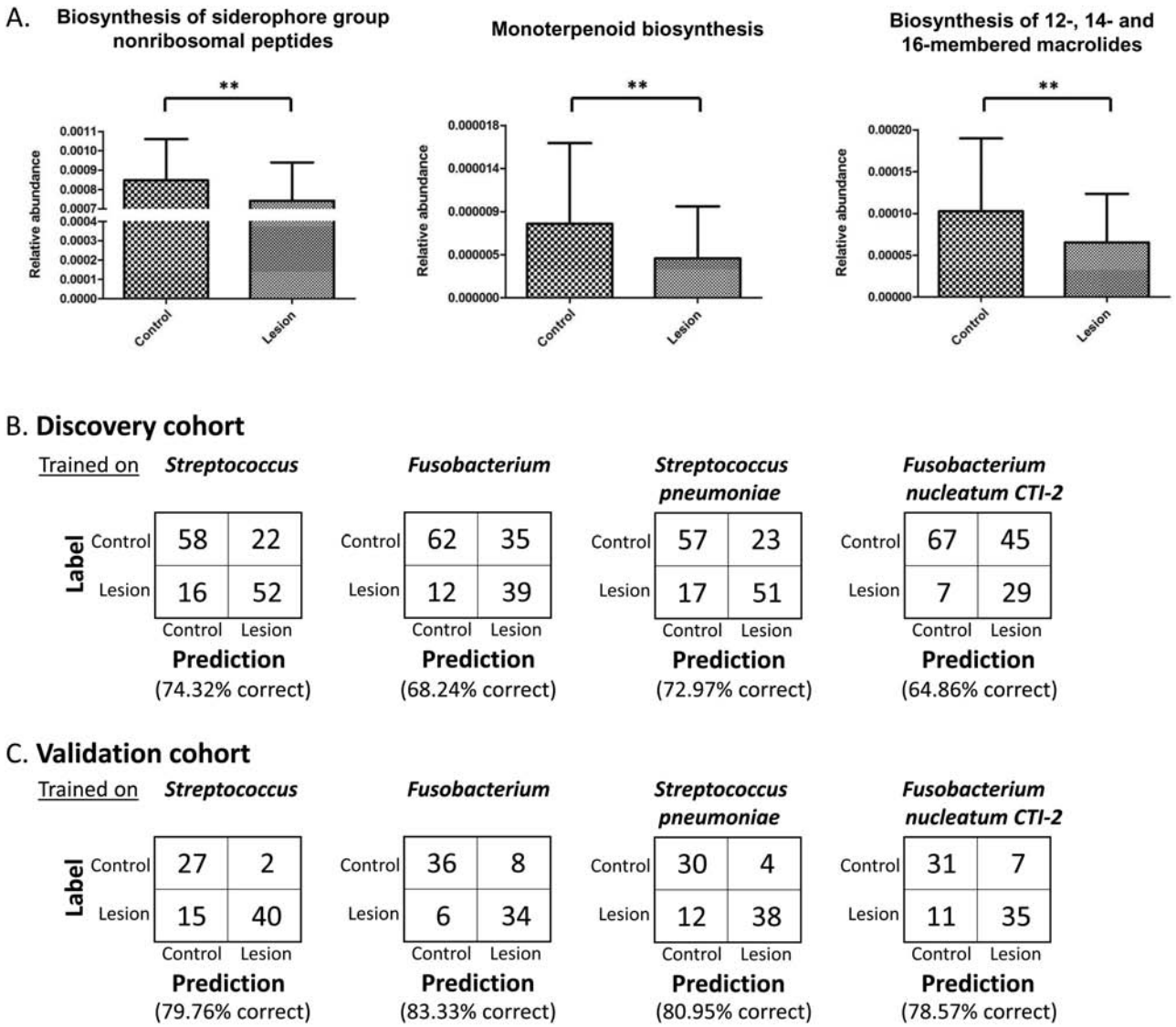

Figure 4. Prediction for microbial gene functions and OSCC status from oral microbial features. (**A**) Pathway enrichment for KEGG metabolism was inferred by PICRUSt. Difference in relative abundance of predicted microbial genes related to the metabolism between OSCC lesions (Lesion) and their contralateral normal tissues (Control) was analyzed using Student's *t*-test. $^{**}P < 0.01$. (**B**) Confusion matrix evaluations of OSCC-associated biomarkers, including genus *Streptococcus*, genus *Fusobacterium*, *Streptococcus pneumoniae* and *Fusobacterium nucleatum CTI-2*, within the discovery cohort (74 pairs of normal-lesion samples). The number in row *i* and column *j* indicates how many samples were labeled as clinical status *i* but assigned to clinical status *j*. Accuracy values below the matrices indicate the fraction of correctly classified instances. (**C**) Confusion matrix evaluations of OSCC-associated biomarkers trained on the discovery cohort and tested on the independent validation cohort (42 pairs of normal-lesion samples).

Moreover, those microbial biomarkers were evaluated within the discovery cohort for predicting OSCC (Figure 4B) and then tested on the validation cohort (Figure 4C). We found that *Fusobacterium* spp. performed reasonably well on successfully labeling the regions of non-cancerous control tissues but caused marked misclassification of OSCC lesions. On the contrary, *Streptococcus* spp. and, in particular, *Streptococcus pneumoniae* were useful for accurate classification of the lesion sites. Our findings unveil promising avenues for prognosis monitoring and diagnosis of OSCC by using specific oral microflora.

Discussion

Disturbance in oral microflora, which causes production of carcinogenic substances (e.g. nitrosamine), inflammatory responses and direct proliferative effects on cellular signaling in epithelium, may dictate the development of oral cancer (30). One of the major confounding factors in the microbiome research of OSCC relates to a number of anatomical regions that comprise such devastating disorder, while it is known that the composition of oral microbiome communities fluctuates by saliva and distinct sites of the mouth (12). In addition, many host parameters are shown to modify the oral microbiome such as sex, ethnicity, habitual exposure to OSCC risks (smoking, drinking and betel nut chewing) and individuals' oral health condition (31). Therefore, careful control for the distinction between oral habitats and inter-individual variations is necessary to explore the shift in oral microbiome populations that is solely due to oral tumorigenicity. Considering this, we performed a discovery screen by profiling the bacteria within the

Table 1. Clinical validity for potential biomarkers in discriminating OSCC from the controls

Biomarkers	Discovery cohort (n = 74 pairs)		Validation cohort (n = 42 pairs)		Combined (n = 116 pairs)	
	AUC (95% CI)	*P* value/adj. *P* value	AUC (95% CI)	*P* value/adj. *P* value	AUC (95% CI)	*P* value/adj. *P* value
Genus						
Streptococcus	0.75 (0.67–0.83)	1.1e-7/**9.0e-6***	0.85 (0.77–0.93)	4.2e-9/**2.8e-7***	0.79 (0.73–0.85)	1.6e-14/**1.2e-12***
Fusobacterium	0.7 (0.61–0.78)	3.0e-5/**0.0024***	0.88 (0.81–0.95)	7.0e-11/**4.7e-9***	0.76 (0.70–0.82)	5.3e-12/**4.0e-10***
Peptostreptococcus	0.67 (0.58–0.75)	0.00049/**0.0402***	0.85 (0.77–0.94)	2.8e-8/**1.9e-6***	0.73 (0.66–0.80)	1.2e-9/**9.2e-8***
Campylobacter	0.66 (0.58–0.75)	0.00063/0.0516	0.81 (0.72–0.91)	2.6e-7/**1.7e-4***	0.72 (0.66–0.79)	4.3e-9/**3.3e-7***
Species						
Streptococcus pneumoniae	0.74 (0.66–0.82)	3.8e-7/**7.9e-5***	0.82 (0.73–0.92)	8.2e-8/**1.5e-5***	0.77 (0.71–0.83)	6.1e-13/**1.2e-10***
Fusobacterium nucleatum CTI-2	0.66 (0.57–0.75)	0.00078/0.1638	0.84 (0.75–0.93)	1.9e-8/**3.5e-6***	0.72 (0.66–0.79)	5.4e-9/**1.0e-6***

P values were adjusted (adj.) by using Bonferroni's correction (discovery: n = 210, for species; n = 82, for genus; validation: n = 187, for species; n = 68, for genus; combined: n = 201, for species; n = 77, for genus). *Bold values represent adj. *P* Value < 0.05.

surface of OSCC lesion and its contralateral normal tissue from each of 74 male patients with cancers of the buccal mucosa, the most common site for oral cancer (~40%) among tumors developed in all distinct anatomical locations of the mouth in Taiwan (16). Significant alterations in the bacterial diversity and relative abundance of specific oral microbiota (most profoundly, an enrichment for genus *Fusobacterium* and the loss of genus *Streptococcus* in the tumor sites) were identified. The clinical validity of these OSCC-associated bacteria was evaluated and then validated in an independent cohort of 42 male patients with cancers of the buccal mucosa. In addition to the compositional changes, prediction of oral microbiome functionality unveiled that bacterial genes associated with the metabolism of terpenoids and polyketides were differentially enriched between the control and tumor lesion group. Overall, our analyses reveal signatures and functions of oral microbiota related to oral cancer.

Consistent with a previous report where buccal swabs were analyzed (32), we observed an increase in microbial richness (α diversity) in the surface of OSCC lesions. This is incompatible with the findings from the studies using tissue biopsies (33) or saliva samples (11), indicating that careful interpretation of oral microbiome data is needed in OSCC studies as different sampling methods were employed or tumors located at different habitats of oral cavity were investigated. In addition to the bacterial diversity, bacterial biomarkers related to OSCC were identified in the present study, with most differential genera being *Fusobacterium* (enriched in OSCC) and *Streptococcus* (decreased in OSCC). The clinical validity of genus *Fusobacterium* and *Streptococcus* was further assessed and verified by using an independent cohort. Such shift in the levels of these two genera could favor a more inflammatory state of oral epithelium, as *Streptococcus* spp. have been shown to impair *F.nucleatum*-induced inflammation in oral epithelial cells (34). Moreover, with initial colonization of *Streptococcus* on oral epithelium, *Fusobacterium* spp. elicit coaggregation with other bacterial genera to assemble the complex oral biofilm (35). The formation of biofilm structure is found to promote the invasiveness of *F.nucleatum* into oral mucosa (36). The presence of *F.nucleatum* was demonstrated to protect tumor cells from immune cell attack (37) and to stimulate OSCC development by engagement with oral epithelium via Toll-like receptors (38). These findings support an oncogenic role of this dysbiotic signature in oral cancer development.

Other than the two most differential genera mentioned above, genus *Peptostreptococcus*, *Prevotella* and *Capnocytophaga* detected as OSCC-associated biomarkers in the present study were also reported elsewhere (15,33,39). As the potential role of these differentially abundant microbes in OSCC progression remains inconclusive, one might speculate that these microorganisms have a propensity for adherence or propagation in the malignant oral tissues. Of note, we demonstrated that *Campylobacter* spp. were significantly enriched in oral cancer (Figure 2 and Supplementary Figure 2, available at *Carcinogenesis* online) and showed a fairly good discernibility in discriminating the controls from OSCC lesions (Table 1). A *Campylobacter*-derived genotoxin, cytolethal distending toxin, was found to induce DNA double-strand breaks and facilitate colorectal tumorigenesis (40). It is also documented that *Campylobacter* spp. were elevated in oral leukoplakia, a condition that presents as white patches of the oral mucosa and probably transforms into malignancies (41). Moreover, in concordance with the observations from OSCC biopsies (42), we have seen a particular enrichment for *C.concisus* in the surface of OSCC lesions (Figure 3A). These findings, together with our data, indicate that *Campylobacter* spp. may be involved in oral carcinogenesis via the induction of DNA damage and potentially beneficial for early diagnosis of OSCC.

The functional capacity of oral bacteria upon oral cancer development may include generation of carcinogenic substances, chronic inflammation, alteration in tissue stem-cell homeostasis and direct manipulation of host cell biology resulting in elevated proliferation or attenuated apoptosis (30,31). It is demonstrated that microbes and their products activate fibroblasts and immune cells to produce reactive oxygen species that trigger DNA damage in epithelial cells (43). In addition, bacterial genotoxins (cytolethal distending toxin, typhoid toxin and colibactin) or metabolites (hydrogen sulfide, ammonia and fatty acids) may directly target to DNA and elicit mutations. Here, we found that microbial pathway modules associated with metabolism of terpenoids and polyketides, including biosynthesis of siderophore group nonribosomal peptides, monoterpenoid biosynthesis and biosynthesis of 12-, 14- and 16-membered macrolides, were less abundant in the tumor lesions when compared with the controls. Such functional characteristics were replicated in the validation cohort and combined samples (data not shown). Terpenoids and polyketides are secondary metabolites produced by certain microorganisms and possess potent pharmaceutical activities against cancer (44,45). Specifically, monoterpenoids and macrolides belong to the terpenoid and polyketide class, respectively, both of which exhibit cancer-chemopreventive activities (46,47). These open a novel aspect of

OSCC etiology and restate a functional role of oral microbiome in formulating a tumor microenvironment via attenuated biosynthesis of secondary metabolites with anti-cancer effects.

To characterize OSCC-associated oral microorganisms, extra efforts are needed to deal with many limitations of this study. One caveat is the lack of a cohort without OSCC for additional comparisons. Also unavailable was the information regarding eating behaviors and diet composition. Collecting samples from patients' non-cancerous family members for analyses may improve this caveat. Another issue is that the functionality of microbial communities was inferred from the sequences of the 16S rRNA gene. Further characterizing the network of numerous pathway modules employed in OSCC-associated oral microflora and improving the resolution of taxonomical classification at the species level will require a deep shotgun metagenomics analysis in the following study. Collectively, we demonstrated that oral microbiota is compositionally and functionally associated with the development of oral cancer. Our study implicates specific oral microbes as a potential biomarker for early diagnosis and prognosis monitoring of this deadly malignancy.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Funding

This study was supported by research grants from Health and Welfare Surcharge of Tobacco Products, Ministry of Health and Welfare, Taiwan (MOHW108-TDU-B-212-124025, MOHW109-TDU-B-212-134025) and from Chung Shan Medical University Hospital, Taiwan (CSH-2020-E-001-Y3) to S.F.Y. and from Chang Gung Memorial Hospital (CMRPG2F0421) to S.C.S.
Conflict of Interest Statement: None declared.

References

1. Jemal, A. *et al.* (2008) Cancer statistics, 2008. *CA Cancer J. Clin.*, 58, 71–96.
2. Krishna Rao, S.V. *et al.* (2013) Epidemiology of oral cancer in Asia in the past decade – an update (2000–2012). *Asian Pac. J. Cancer Prev.*, 14, 5567–5577.
3. Scully, C. *et al.* (2000) Genetic aberrations in oral or head and neck squamous cell carcinoma (SCCHN): 1. Carcinogen metabolism, DNA repair and cell cycle control. *Oral Oncol.*, 36, 256–263.
4. Scully, C. *et al.* (2009) Oral squamous cell carcinoma overview. *Oral Oncol.*, 45, 301–308.
5. Tezal, M. *et al.* (2009) Chronic periodontitis and the incidence of head and neck squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.*, 18, 2406–2412.
6. Hooper, S.J. *et al.* (2009) Exploring the link between microorganisms and oral cancer: a systematic review of the literature. *Head Neck*, 31, 1228–1239.
7. Zini, A. *et al.* (2010) Oral cancer over four decades: epidemiology, trends, histology, and survival by anatomical sites. *J. Oral Pathol. Med.*, 39, 299–305.
8. Dewhirst, F.E. *et al.* (2010) The human oral microbiome. *J. Bacteriol.*, 192, 5002–5017.
9. Zaura, E. *et al.* (2014) Acquiring and maintaining a normal oral microbiome: current perspective. *Front. Cell. Infect. Microbiol.*, 4, 85.
10. Wang, L. *et al.* (2014) The oral microbiome and oral cancer. *Clin. Lab. Med.*, 34, 711–719.
11. Guerrero-Preston, R. *et al.* (2016) 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. *Oncotarget*, 7, 51320–51334.
12. Zhou, Y. *et al.* (2013) Biogeography of the ecosystems of the healthy human body. *Genome Biol.*, 14, R1.
13. Jeng, J.H. *et al.* (2001) Role of areca nut in betel quid-associated chemical carcinogenesis: current awareness and future perspectives. *Oral Oncol.*, 37, 477–492.
14. Jeng, J.H. *et al.* (2000) Areca nut extract up-regulates prostaglandin production, cyclooxygenase-2 mRNA and protein expression of human oral keratinocytes. *Carcinogenesis*, 21, 1365–1370.
15. Lee, W.H. *et al.* (2017) Bacterial alterations in salivary microbiota and their association in oral cancer. *Sci. Rep.*, 7, 16540.
16. Dhanuthai, K. *et al.* (2018) Oral cancer: a multicenter study. *Med. Oral Patol. Oral Cir. Bucal*, 23, e23–e29.
17. Yang, S.F. *et al.* (2018) Compositional and functional variations of oral microbiota associated with the mutational changes in oral cancer. *Oral Oncol.*, 77, 1–8.
18. Schmidt, B.L., *et al.* (2014) Changes in abundance of oral microbiota associated with oral cancer. *PLoS One*, 9, e98741.
19. Wu, I.W. *et al.* (2020) Gut microbiota as diagnostic tools for mirroring disease progression and circulating nephrotoxin levels in chronic kidney disease: discovery and Validation Study. *Int. J. Biol. Sci.*, 16, 420–434.
20. Caporaso, J.G. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, 7, 335–336.
21. Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*, 10, 996–998.
22. Quast, C., *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*, 41, D590–596.
23. Caporaso, J.G. *et al.* (2010) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26, 266–267.
24. Price, M.N. *et al.* (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.*, 26, 1641–1650.
25. Lozupone, C. *et al.* (2011) UniFrac: an effective distance metric for microbial community comparison. *ISME J.*, 5, 169–172.
26. Langille, M.G. *et al.* (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.*, 31, 814–821.
27. Markowitz, V.M., *et al.* (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.*, 40, D115–122.
28. DeSantis, T.Z. *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, 72, 5069–5072.
29. Segata, N. *et al.* (2011) Metagenomic biomarker discovery and explanation. *Genome Biol.*, 12, R60.
30. Chattopadhyay, I. *et al.* (2019) Role of oral microbiome signatures in diagnosis and prognosis of oral cancer. *Technol. Cancer Res. Treat.*, 18, 1533033819867354.
31. Lim, Y. *et al.* (2017) Oral microbiome: a new biomarker reservoir for oral and oropharyngeal cancers. *Theranostics*, 7, 4313–4321.
32. Zhao, H. *et al.* (2017) Variations in oral microbiota associated with oral cancer. *Sci. Rep.*, 7, 11773.
33. Pushalkar, S. *et al.* (2012) Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiol.*, 12, 144.
34. Zhang, G. *et al.* (2011) Streptococcus cristatus modulates the *Fusobacterium nucleatum*-induced epithelial interleukin-8 response through the nuclear factor-kappa B pathway. *J. Periodontal Res.*, 46, 558–567.
35. Nobbs, A.H. *et al.* (2011) Stick to your gums: mechanisms of oral microbial adherence. *J. Dent. Res.*, 90, 1271–1278.
36. Gursoy, U.K. *et al.* (2010) Biofilm formation enhances the oxygen tolerance and invasiveness of *Fusobacterium nucleatum* in an oral mucosa culture model. *J. Periodontol.*, 81, 1084–1091.
37. Gur, C. *et al.* (2015) Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity*, 42, 344–355.
38. Binder Gallimidi, A. *et al.* (2015) Periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget*, 6, 22613–22623.

39. Mager, D.L. *et al.* (2005) The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *J. Transl. Med.*, 3, 27.
40. He, Z. *et al.* (2019) Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut*, 68, 289–300.
41. Amer, A. *et al.* (2017) The microbiome of potentially malignant oral leukoplakia exhibits enrichment for *Fusobacterium*, *Leptotrichia*, *Campylobacter*, and *Rothia* species. *Front. Microbiol.*, 8, 2391.
42. Perera, M. *et al.* (2018) Inflammatory bacteriome and oral squamous cell carcinoma. *J. Dent. Res.*, 97, 725–732.
43. Schwabe, R.F. *et al.* (2013) The microbiome and cancer. *Nat. Rev. Cancer*, 13, 800–812.
44. Huang, M. *et al.* (2012) Terpenoids: natural products for cancer therapy. *Expert Opin. Investig. Drugs*, 21, 1801–1818.
45. Gomes, E.S. *et al.* (2013) Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Braz. J. Microbiol.*, 44, 1007–1034.
46. Crowell, P.L. (1999) Prevention and therapy of cancer by dietary monoterpenes. *J. Nutr.*, 129, 775S–778S.
47. Kanoh, S. *et al.* (2010) Mechanisms of action and clinical application of macrolides as immunomodulatory medications. *Clin. Microbiol. Rev.*, 23, 590–615.