

# SCIENTIFIC REPORTS



OPEN

## Variations in oral microbiota associated with oral cancer

Hongsen Zhao<sup>1</sup>, Min Chu<sup>1</sup>, Zhengwei Huang<sup>1</sup>, Xi Yang<sup>2</sup>, Shujun Ran<sup>1</sup>, Bin Hu<sup>2</sup>, Chenping Zhang<sup>3</sup> & Jingping Liang<sup>1</sup>

Individual bacteria and shifts in microbiome composition are associated with human disease, including cancer. To unravel the connections underlying oral bacterial dysbiosis and oral squamous cell carcinoma (OSCC), cancer lesion samples and anatomically matched normal samples were obtained from the same patients. We then profiled the bacteria within OSCC lesion surface samples at the species level using next-generation sequencing to comprehensively investigate bacterial community composition and functional genes in these samples. Significantly greater bacterial diversity was observed in the cancer samples than in the normal samples. Compared with previous studies, we identified many more taxa demonstrating remarkably different distributions between the groups. In particular, a group of periodontitis-correlated taxa, including *Fusobacterium*, *Dialister*, *Peptostreptococcus*, *Filifactor*, *Peptococcus*, *Catonella* and *Parvimonas*, was significantly enriched in OSCC samples. Additionally, several operational taxonomic units (OTUs) associated with *Fusobacterium* were highly involved in OSCC and demonstrated good diagnostic power. Our study revealed drastic changes in surface bacterial communities of OSCC. The findings enrich knowledge of the association between oral bacterial communities and oral cancer.

The human body is inhabited by over 100 trillion microbial cells living in symbiosis with their host<sup>1</sup>. Bacteria at certain body sites have long been believed to be involved in immune modulation, disease development, and health maintenance. The term microbiome was coined to describe “the collective genomes and gene products of all microbes residing within an organism”<sup>2</sup>. With the advent of high-throughput, next-generation sequencing (NGS), there has been a surge of interest in studying the human microbiome in the context of disease. Recent studies have demonstrated the importance of the gut microbiota in digestion, fat storage, angiogenesis, immune system development and responses, resistance to colonization, epithelial architecture<sup>3,4</sup>, and dysbiosis, which is believed to contribute to the pathogenesis of local and systemic diseases, including inflammatory bowel disease, diabetes, and colorectal cancer<sup>5</sup>. Located at the beginning of the aerodigestive tract, approximately 700 prokaryote species have been detected in the human oral cavity. These species belong to 185 genera and 12 phyla, of which approximately 54% are officially named, 14% unnamed (but cultivated) and 32% known only as uncultivated phylotypes<sup>6</sup>. This oral bacterial flora plays an essential role in maintaining a normal oral physiological environment and is associated with host health<sup>7</sup>. In contrast to traditional views, recent analyses suggest the involvement of a consortium of microbes, rather than a single species, as causing disease<sup>8</sup>, a phenomenon that has been well characterized for periodontal diseases<sup>9</sup>.

Oral cancer, primarily oral squamous cell carcinoma (OSCC) deriving from the oral mucosa, is a disease that arises from both host genetics and environmental factors; tobacco and alcohol consumption, betel quid chewing, and human papillomavirus infection are well-known risk factors<sup>10</sup>. The incidence of oral cancer is increasing, and this disease continues to be a major global health problem. Furthermore, approximately 15% of oral cancer cases cannot be attributed to the aforementioned major risk factors, resulting in the need to explore other potential risk factors<sup>11</sup>. A plethora of bacteria, the proverbial bacterial biofilm, coat each surface of the oral cavity<sup>12</sup>, and groups inhabiting the mucosal surface might constitute the bulk of the tumor microenvironment. To date, various microbes and changes in different bacteria have been associated with several types of cancer<sup>13</sup>. Cancer-associated

<sup>1</sup>Ninth People's Hospital, School of Medicine, Shanghai Jiao Tong University, Department of Endodontics and Operative Dentistry, Shanghai Key Laboratory of Stomatology, Shanghai, 200011, China. <sup>2</sup>Ninth People's Hospital, School of Medicine, Shanghai Jiao Tong University, Department of Prosthodontics, Shanghai, 200011, China. <sup>3</sup>Ninth People's Hospital, School of Medicine, Shanghai Jiao Tong University, Department of Oral Maxillofacial-Head and Neck Oncology, Shanghai, 200011, China. Hongsen Zhao and Min Chu contributed equally to this work. Correspondence and requests for materials should be addressed to B.H. (email: [hub1093@2m9h.net](mailto:hub1093@2m9h.net)) or C.Z. (email: [zhang.chenping@hotmail.com](mailto:zhang.chenping@hotmail.com)) or J.L. (email: [liangjpdentist@126.com](mailto:liangjpdentist@126.com))

changes in the oral microbiome have been assessed in several early studies employing culture-based or molecular techniques<sup>14–19</sup>, but a consensus has not been reached due to the limited number of strains/clones that it is feasible to test. However, the emergence of NGS allows microbial communities to be profiled at an unprecedented depth and coverage.

To date, several studies have employed NGS to assess bacterial profiles associated with OSCC. Pushalkar *et al.* evaluated the diversity and relative abundance of bacteria in the saliva of subjects with OSCC; however, only three OSCC cases and two healthy controls were included<sup>20</sup>. Later, a larger-scale study analyzed swabs of lesion surfaces and contra-lateral normal mucosae from 18 OSCC patients, and significant decreases in the abundances of the genera *Firmicutes* and *Actinobacteria* were observed in cancer samples<sup>21</sup>. In another report, saliva bacterial communities in six OSCC patients were elucidated by performing pyrosequencing, and paired taxa within the family *Enterobacteriaceae* together with the genus *Oribacterium* were suggested to distinguish OSCC samples from oropharyngeal squamous cell carcinoma (OPSCC) and normal samples<sup>22</sup>. Regardless, the studies above failed to provide bacterial composition at the species level, even though specific species or even strains are usually involved in disease. Al-hebshi successively profiled bacterial communities within 23 OSCC tissue samples from Yemeni patients at the species level, providing the first epidemiological evidence for associations of *Fusobacterium nucleatum* and *Pseudomonas aeruginosa* with OSCC<sup>23,24</sup>. Given the limited number of OSCC samples included, the significance of the findings from these studies is unclear. Accordingly, more studies are warranted to validate these results.

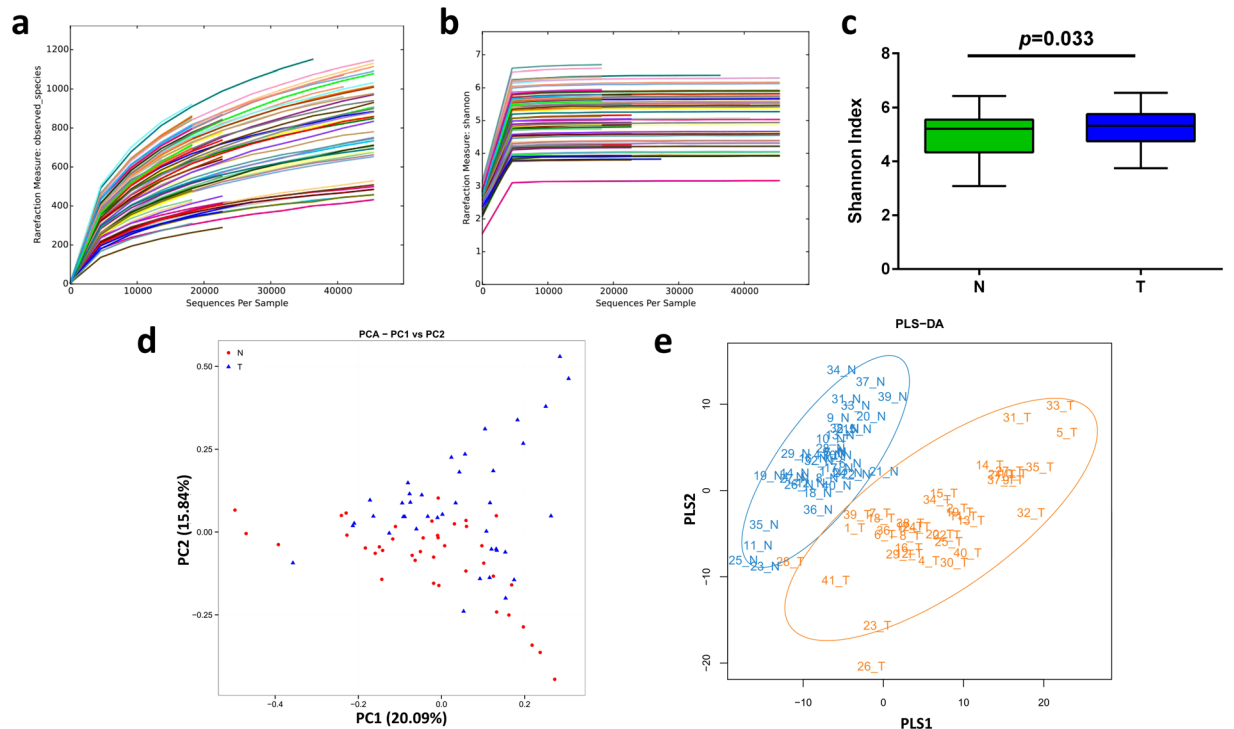
In the current study, cancer lesion samples and matched controls were procured from 40 Chinese subjects with OSCC. Bacterial profiles within the samples were characterized at the species level. Shifts in bacterial composition and gene functions associated with OSCC were described and analyzed. In particular, we detected a group of periodontitis-related taxa that was significantly enriched in OSCC samples. Our findings may contribute to further clarification of the connection between OSCC and oral bacteria.

## Results

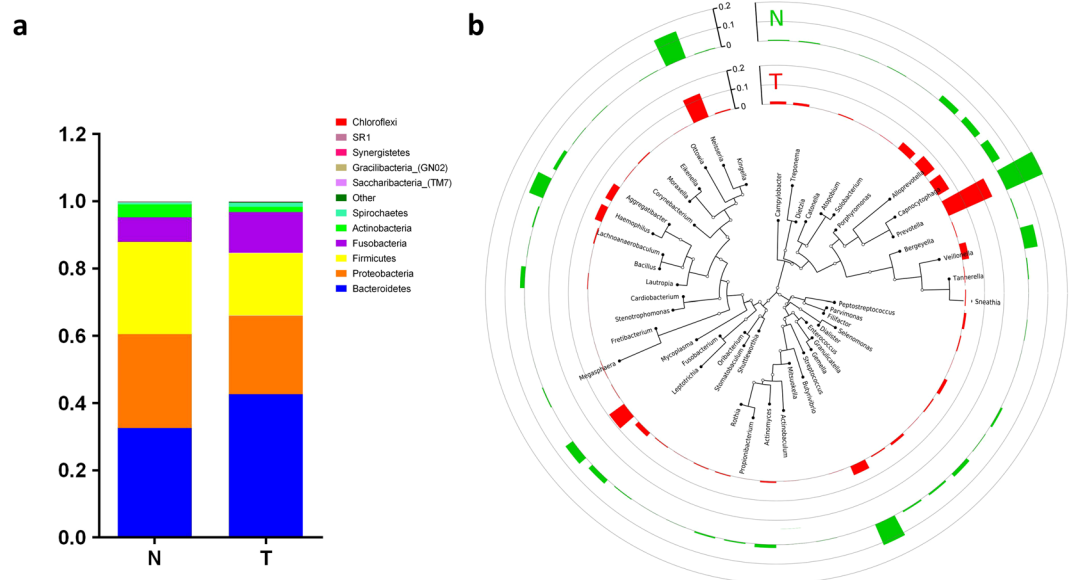
**Overall structure of bacterial communities across samples.** In the current study, 80 samples were sequenced using an Illumina MiSeq system, and a total of 5,075,391 raw sequences were generated. After quality trimming and chimera checking, 4,075,169 high-quality sequences with an average length of 391 bp were recovered for downstream analysis, with an average of 50,940 reads (ranging from 19,353 to 117,244 reads) per sample. After alignment in the HOMD, unique representative sequences were classified into 2,334 operational taxonomic units (OTUs) at a 97% similarity level, from which 11 phyla, 130 genera and 389 species were detected. Good's estimator of coverage was 99.26%, indicating the 16 S rRNA sequences identified in this study likely represent the majority of bacterial sequences present in the samples. Different indexes (Shannon, Simpson, Abundance-based Coverage Estimator (ACE), and Chao 1) were employed to estimate the  $\alpha$ -diversity of the bacterial community. Although rarefaction curves of numbers of observed OTUs per sample suggested new phylotypes would be expected with additional sequencing (Fig. 1a), the rarefaction curves for the Shannon diversity index for each sample reached plateaus, indicating that the majority of the diversity was already procured (Fig. 1b). As revealed by the Shannon diversity index (Fig. 1c), the diversity of the bacterial community in the cancer samples was significantly increased compared with that of the corresponding clinically normal control samples. A similar trend was observed when employing other diversity indexes (Simpson, ACE, Chao 1), although without statistical significance (Supplementary Fig. S1).

To evaluate the extent of the similarity of the bacterial communities, unweighted UniFrac Principal Component Analysis (PCA) at the OTU level was employed and indicated no obvious separation between groups (Fig. 1d). Then, Partial Least Squares Discriminant Analysis (PLS-DA), a supervised analysis suitable for high-dimensional data, was performed (Fig. 1e). The bacterial communities in the cancer samples and the matched controls clustered separately, suggesting the overall structures of the bacterial communities in the groups were significantly different. Spots representing the cancer samples presented more dispersed distribution patterns than those of the controls, aligning with the increased level of bacterial diversity found in the cancer samples. A nonparametric multivariate analysis of variance (Adonis) and an analysis of similarities (ANOSIM) based on UniFrac distances were performed, and the calculated P values ( $P = 0.002$  for Adonis,  $P = 0.001$  for ANOSIM) further demonstrated the remarkable differences between the bacterial communities in the groups.

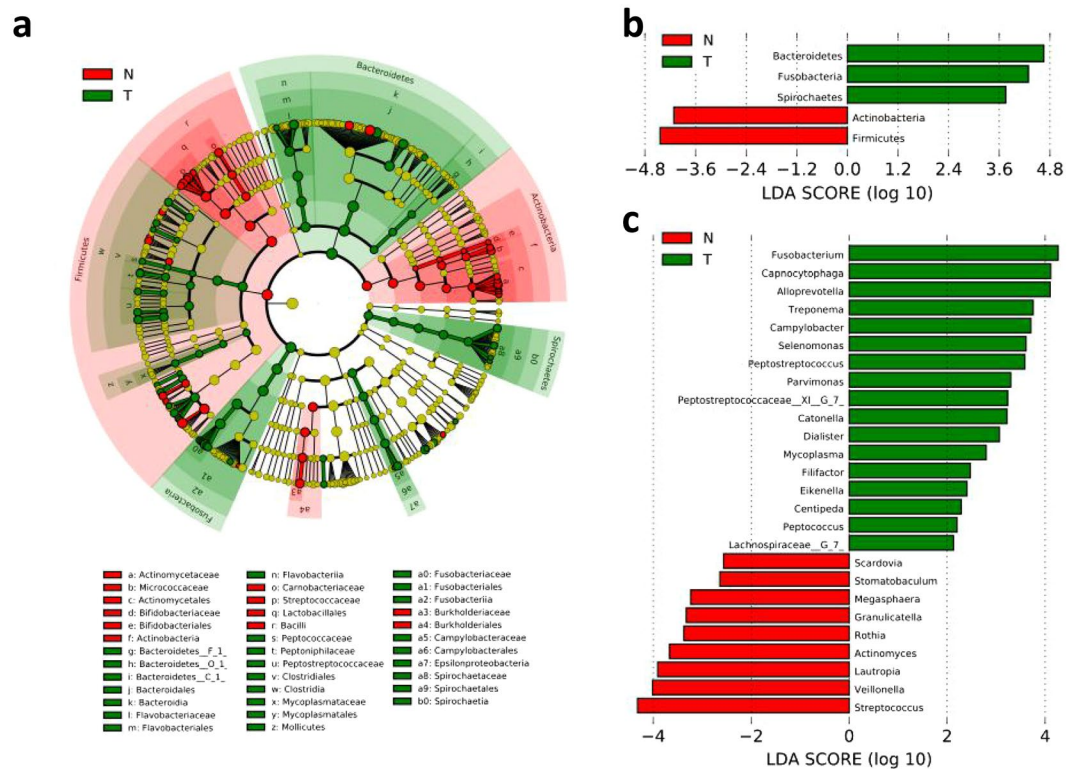
**Common and distinct bacterial taxa in the analyzed groups.** The bacterial communities in the cancer lesions and the controls were analyzed at different taxonomic levels (Fig. 2 and Supplementary Fig. S2). *Bacteroidetes*, *Proteobacteria*, *Firmicutes*, *Fusobacteria*, *Actinobacteria*, the top five most abundant phyla, together comprised 98.62% of all sequences (Fig. 2a). *Bacteroidetes* was the most abundant phylum, accounting for 37.6% of sequences. In contrast, the abundances of the other detected phyla, including *Synergistetes*, *SRL1*, and *Chloroflexi*, were less than 0.1%. At the genus level, *Prevotella*, *Neisseria*, *Streptococcus*, *Fusobacterium*, and *Haemophilus* were the five most abundant genera, comprising 22.46%, 13.67%, 8.17%, 6.95%, and 5.74% of sequences, respectively (Fig. 2b). Of all genera detected, 18 taxa were found in all samples; in addition to the five genera mentioned above, these included *Capnocytophaga*, *Veillonella*, *Alloprevotella*, *Porphyromonas*, *Leptotrichia*, *Aggregatibacter*, *Selenomonas*, *Campylobacter*, *Granulicatella*, *Actinomyces*, *Gemella*, *Lachnoanaerobaculum*, and *Bergeyella*. The shared genera collectively represented over 80.0% of all detected sequences. At the species level, an average of approximately 200 species were detected per sample. The relative abundances of eleven species each surpassed 2%; specifically, *Neisseria flavescens*, *Prevotella melaninogenica*, *Fusobacterium periodonticum*, *Streptococcus oralis*, *Prevotella intermedia*, *Veillonella atypica*, *Haemophilus parahaemolyticus*, *Porphyromonas sp. \_oral\_taxon\_279*, *Capnocytophaga leadbetteri*, *Alloprevotella sp. \_oral\_taxon\_473* and *Haemophilus parainfluenzae* together accounted for a total of 53.8% of sequences (Supplementary Fig. S2). *Neisseria flavescens* was the most abundant species in both groups, with a relative abundance of 10.73% in cancer lesions and 12.19% in normal



**Figure 1.** Comparison of oral microbiota structures in the N and T groups. (a) Rarefaction analysis of bacterial 16S rRNA gene sequences was performed to evaluate whether further sequencing would likely detect additional taxa, indicated by a plateau. Different colors represent different samples. (b) Shannon index curves were constructed to evaluate the numbers of samples likely required to identify additional taxa, indicated by a plateau. Different colors represent different samples. (c) Box plots depict differences in bacterial diversity between the N and T groups according to the Shannon index. (d) PCA at the OTU level. (e) Partial least square discriminant score plot of oral microbiota between the N and T groups. N, clinical normal samples; T, oral cancer samples.



**Figure 2.** Composition of bacterial communities across samples at the phylum and genus levels. (a) Relative abundance of bacterial phyla among the N and T groups. (b) Classification tree of the 50 most abundant genera across all samples. The outer cycle colored bars represent the relative abundances of taxa in each group. N, clinical normal samples; T, oral cancer samples.



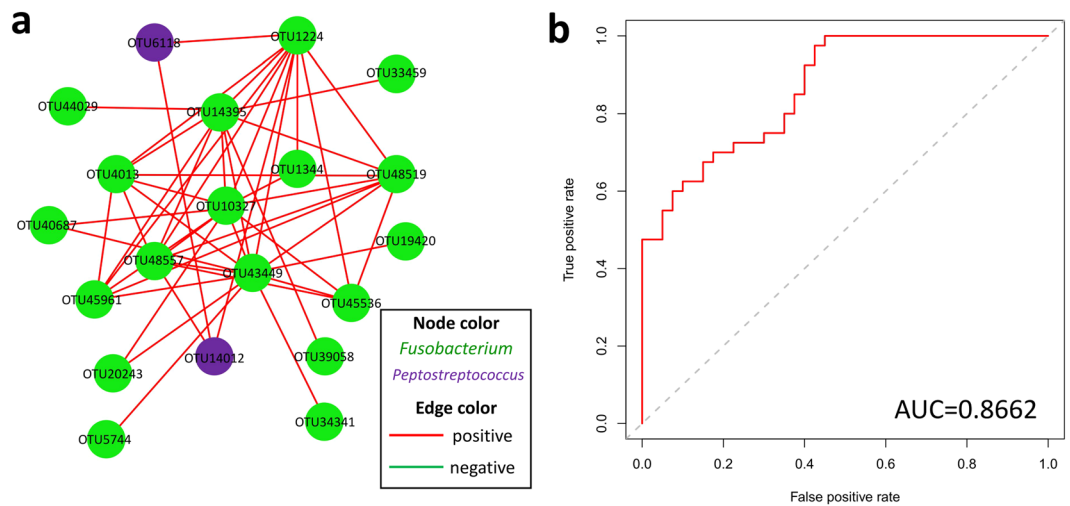
**Figure 3.** Distinct taxa identified in the N and T groups using LEfSe analysis. **(a)** Cladogram constructed using the LEfSe method to indicate the phylogenetic distribution of bacteria that were remarkably enriched in the N and T groups. **(b)** LDA scores showed significant bacterial differences within groups at the phylum level. **(c)** LDA scores showed significant bacterial differences within groups at the genus level. N, clinical normal samples; T, oral cancer samples.

controls. Despite significant inter-individual variation, 14 species were detected across all samples, including *Actinomyces odontolyticus*, *Prevotella melaninogenica*, *Prevotella scopos*, *Capnocytophaga gingivalis*, *Gemella sanguinis*, *Granulicatella adiacens*, *Streptococcus oralis*, *Streptococcus salivarius*, *Lachnoanaerobaculum umeaense*, *Veillonella atypica*, *Fusobacterium periodonticum*, *Neisseria elongata*, *Neisseria flavescens*, and *Haemophilus parainfluenzae*, and constituted the oral mucosal core bacteriome of the OSCC patients.

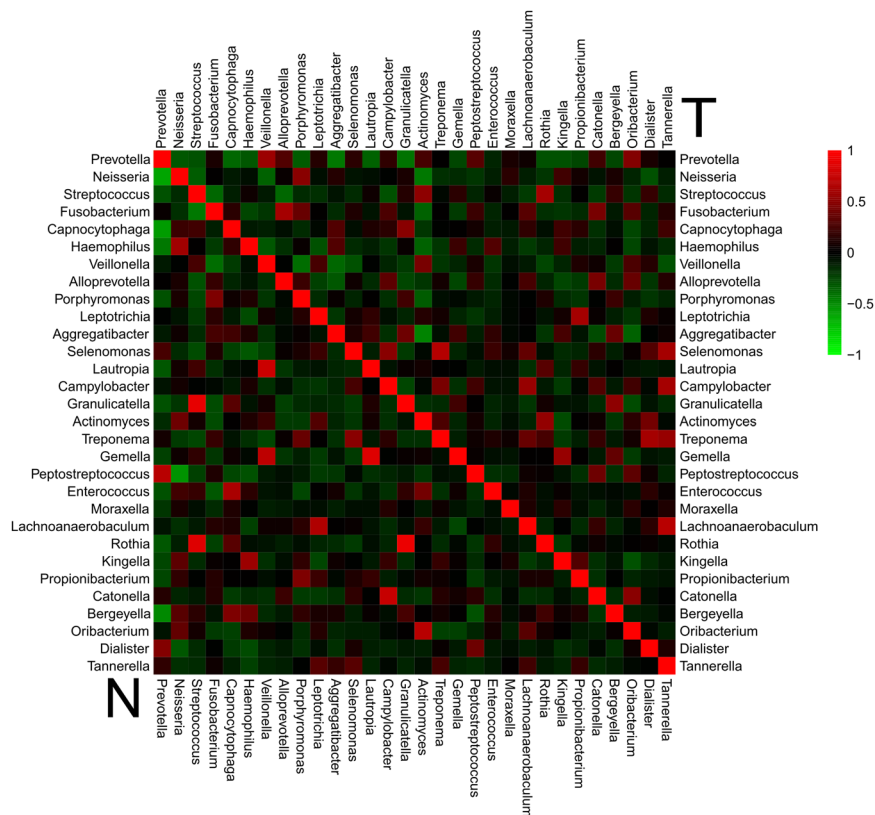
The bacterial composition of the cancer samples varied from that of the controls. There were 6 phyla and 68 genera with higher abundances in the cancer samples than in the controls. To identify the distinguishing taxa within the groups, the linear discriminant analysis (LDA) effect size (LEfSe) method was implemented (Fig. 3). At the phylum level, *Spirochaetes*, *Fusobacteria*, and *Bacteroidetes* were significantly enriched in diseased samples, while *Firmicutes* and *Actinobacteria* were significantly decreased (Fig. 3b). At the genus level, 17 taxa exhibited significantly higher abundances in the cancer samples than in the controls, including *Mycoplasma*, *Treponema*, *Campylobacter*, *Eikenella*, *Centipeda*, *Lachnospiraceae\_G\_7*, *Alloprevotella*, *Fusobacterium*, *Selenomonas*, *Dialister*, *Peptostreptococcus*, *Filifactor*, *Peptococcus*, *Catonella*, *Parvimonas*, *Capnocytophaga*, and *Peptostreptococcaceae\_XI\_G\_7*. The taxa *Megasphaera*, *Stomatobaculum*, *Granulicatella*, *Lautropia*, *Veillonella*, *Streptococcus*, *Scardovia*, *Rothia*, and *Actinomyces* were remarkably prevalent in the controls (Fig. 3c). *Prevotella* and *Neisseria* were the predominant genera in both groups: there was no significant difference in distribution between the two groups (Fig. 2b). At the species level, 39 species were significantly increased, while 28 species were significantly decreased in diseased samples compared to those levels in the controls (Supplementary Fig. S3).

**Co-occurrence network analysis and function predictions.** To predict the ecological relationships across different bacterial communities, spatial Pearson's correlations between bacterial species were visualized and then analyzed (Supplementary Fig. S4). In the network, pairwise relationships were represented by edges connecting two nodes. There were 1,574 associations among 306 nodes in the controls and 1,794 associations among 366 nodes in the lesions at the OTU level, indicating increased network complexity in the latter group. The genera *Prevotella* and *Neisseria* comprised the two densest clusters in both groups. The hub OTU with the most associations in the clusters, OTU50445 (belonging to *Prevotella melaninogenica*), primarily represented the genus *Prevotella* in both groups, while the hub OTUs in the clusters primarily representing the genus *Neisseria* were OTU52958 in the controls and OTU46085 in the cancer samples, both of which belonged to *Neisseria flavescens*. There was a highly connected bacterial cluster of *Fusobacterium* comprising OTUs (OTU48557, 45961, 43449, 14395, 1224, 10327, and 4013) that were heavily involved in the bacterial ecology structure of the diseased samples compared with the controls, suggesting that *Fusobacterium* may play a critical role in the development





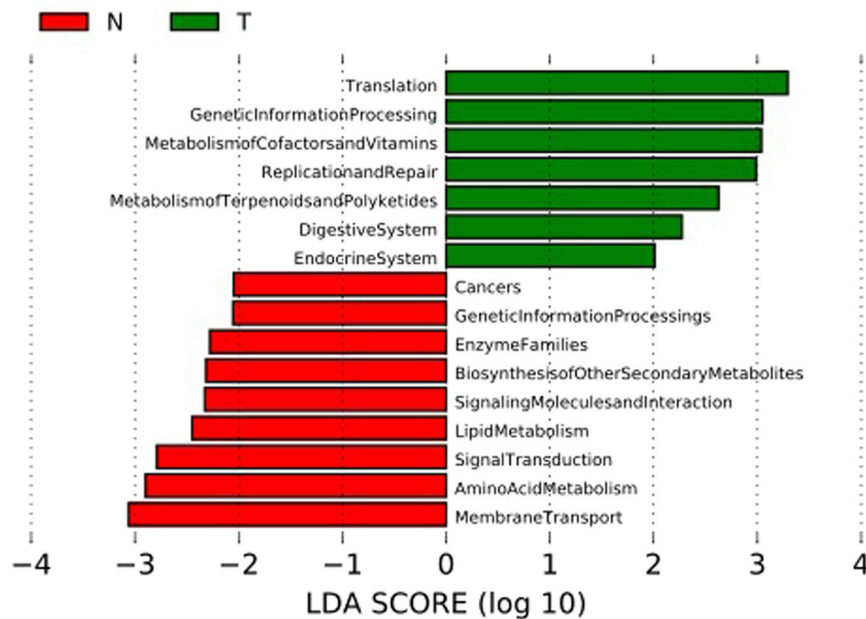
**Figure 4.** Co-occurrence network of *Fusobacterium* comprising OTUs and the diagnostic power of selected OTUs. (a) Each node represents an OTU colored for its genus-level phylotypes, and each edge represents a significant co-occurrence relationship colored according to its association (red: positive, green: negative). (b) ROC curves for selected *Fusobacterium* OTUs were constructed to predict diagnostic power.



**Figure 5.** Co-occurrence and co-exclusion analysis of bacterial genera. Pearson correlations among the top 30 most abundant bacterial genera were calculated and analyzed; groups are shown on the left and right. Correlation values ranged from  $-1.00$  (green) to  $1.00$  (red).

of oral cancer (Fig. 4a). The predictive power of the seven described OTUs was further assessed by constructing a receiver operating characteristic (ROC) curve, a function of the true-positive rate (TPR or sensitivity) and false-positive rate (FPR or 1-specificity). The area under the ROC curve (AUC) reached 0.866, indicating good diagnostic performance (Fig. 4b).

The co-occurrence patterns of the 30 most abundant bacterial genera in each group were investigated in detail (Fig. 5). In matched normal controls, *Tannerella* and *Lachnoanaerobaculum* were the most positively correlated



**Figure 6.** LDA scores predict gene function enriched in different groups using PICRUSt. N, clinical normal samples; T, oral cancer samples.

( $\rho = 0.681$ ), whereas *Actinomyces* and *Aggregatibacter* were the most negatively correlated ( $\rho = -0.462$ ). In the cancer samples, the most positively correlated genera were *Rothia* and *Granulicatella* ( $\rho = 0.903$ ), and *Neisseria* and *Prevotella* were the most negatively correlated ( $\rho = -0.618$ ). Additionally, opposing co-occurrence patterns at the genus and species level were observed. For example, the genera *Actinomyces* and *Neisseria* correlated positively ( $\rho = 0.383$ ) in the lesions but were negatively correlated ( $\rho = -0.429$ ) in normal controls. *Prevotella pallens* and *Prevotella intermedia* demonstrated negative correlations ( $\rho = -0.222$ ) in lesions but positively correlated in the control group ( $\rho = 0.342$ ) (Supplementary Fig. S5).

Finally, the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) algorithm was employed to predict bacterial functions in the two groups. The LEfSe outputs showed a series of metabolic pathways presenting significantly different distributions in each group (Fig. 6). Pathways related to Genetic Information Processing were remarkably enriched in cancer lesions.

## Discussion

Interest in a possible relationship between bacteria and different stages of cancer development has been increasing since the classification of *Helicobacter pylori* as a definite carcinogen by the World Health Organization<sup>25</sup>. Although microorganisms have been implicated in 15.4% of human malignancies<sup>26</sup>, there is a paucity of knowledge regarding the role of bacteria in the progression of OSCC. Because bacterial composition at the species level within OSCC samples has rarely been reported, we attempted to add knowledge regarding this aspect.

Compared with the pyrosequencing applied in previous studies<sup>20–23</sup>, Illumina Miseq is a popular established platform of NGS due to its reduced run times and lower cost of reagents, with an overall error rate below 1%, and the  $2 \times 300$  bp read length is flexible for sequencing of small genomes<sup>27,28</sup>. However, the different sequencing regions (V4–V5<sup>20</sup>, V4<sup>21</sup>, V3–V5<sup>22</sup>, V1–V3<sup>23,24</sup>) targeted between studies, along with differences in sample types, selection of control tissues, and number of samples included, may produce inconsistent results. In this study, the V4–V5 region was chosen for sequencing, as it is proposed to be an ideal target for bacteria in 16S rRNA-based analyses<sup>29</sup>. In the current study, we obtained many more raw sequences and OTUs than those obtained in earlier studies<sup>21,23</sup>, making it possible to comprehensively profile bacterial structure in cancer lesions. Then, we compared the complexity of the bacterial communities present at mucosal sites in both groups, and our results unraveled critical yet pronounced bacterial characteristics associated with disease. Different from previous reports in which tissue biopsies or swabs were analyzed<sup>17,19,21,24</sup>, greater phylogenetic diversity was observed in the OSCC lesion surface in our study, as suggested by rarefaction curves for the Shannon diversity index. With regard to saliva samples, Guerrero-Preston reported a significant loss in richness and diversity of oral bacterial species in oral cancer patients compared to controls<sup>22</sup>, whereas another study revealed much greater diversity of bacterial communities in OSCC samples<sup>30</sup>. Moreover, Adonis and ANOSIM analyses further corroborated the significant differences between bacterial communities in oral cancer lesions and those in matched controls. Thus, oral bacteria dysbiosis appears to be present during OSCC development.

Overall, five of the most abundant phyla detected in our study were consistent with those found in previous studies, specifically, *Bacteroidetes*, *Proteobacteria*, *Firmicutes*, *Fusobacteria*, and *Actinobacteria*, while the less abundant phyla detected in different surveys varied. Although *Firmicutes* was the most abundant phylum in previous studies<sup>20–22</sup>, *Bacteroidetes* was detected as the most abundant phylum in the current study. Of the highly

abundant genera in this report, *Fusobacterium*, *Neisseria*, *Prevotella*, *Streptococcus*, *Leptotrichia*, *Veillonella* and *Capnocytophaga* were previously shown to be the most abundant genera in samples from periodontitis patients<sup>31</sup>. As opposed to *Streptococcus*, which was predominant in Pushalkar's study<sup>17</sup>, *Prevotella* was the most abundant genus across all samples in our analysis, accounting for 22.5% of abundance, which is comparable to previous findings<sup>21</sup>. The use of HOMD enabled the assignment of many more OTUs at the species level. *Neisseria flavescens* and *Fusobacterium periodonticum* are consistently predominant in cancer lesions<sup>23</sup>.

Compared with other studies, our study identified larger numbers of distinguishing taxa at each level using the LEfSe method. At the phylum level, *Firmicutes* and *Actinobacteria* presented with the same patterns reported by Schmidt<sup>21</sup> and were remarkably decreased in cancer lesions, while significant increases in *Fusobacteria* was also observed, consistently<sup>24</sup>. In line with the previous studies, genera *Streptococcus* and *Rothia* were significantly decreased in cancer lesions<sup>21,24</sup>. Intriguingly, the majority of these significantly enriched genera in lesions are involved in periodontal disease, including *Fusobacterium*, *Dialister*, *Peptostreptococcus*, *Filifactor*, *Peptococcus*, *Catonella*, and *Parvimonas*<sup>31</sup>. Consistent with previous findings, remarkable enrichment of *Peptostreptococcus* and *Parvimonas* was observed in cancer samples<sup>17,22</sup>. Additionally, *Veillonella* was significantly decreased in cancer lesions, a finding that was previously reported in 73% of oral cancer patients after treatment<sup>17</sup>, indicating *Veillonella* correlates with a healthy status. Of the distinguishing species identified across the groups, forty species were highly abundant in cancer lesions, including *Porphyromonas endodontalis*, *Filifactor alocis* and *Dialister pneumosintes*, which are newly recognized periodontal pathogens<sup>32</sup>. Of all oral bacteria, *Porphyromonas gingivalis* and *Fusobacterium nucleatum* might possess the greatest potential to be correlated with oral cancer, as both have been implicated in pancreatic and colorectal cancers. Recently, a report by Gallimidi indicated *P. gingivalis* and *F. nucleatum* promote oral cancer progression via direct interactions with oral epithelial cells through Toll-like receptors<sup>33</sup>. However, *P. gingivalis* did not differ in abundance between groups. *Fusobacterium*, comprising the species *periodonticum*, *naviforme*, and *nucleatum\_subsp*, was significantly enriched in lesions, accounting for 8.33%, 0.103%, and 0.297% of sequences in the cancer group, respectively. In another study, *F. periodonticum*, *F. naviforme*, and *F. nucleatum\_subsp* were reported to account for 4.08%, 0.01% and 11.67% of sequences in cancer samples, respectively<sup>24</sup>. Thus, the different prevalences of *Fusobacterium* species detected in OSCC samples between studies may largely be due to differences in sample types, races and geographic regions of the subjects recruited. Further evidence is needed to verify these findings. A higher abundance of several *Treponema* species was observed in cancer lesions. *T. denticola*, a member of the periodontal "red complex" involved in pancreatic cancer<sup>34</sup>, was not included. In the literature, *Bacteroides fragilis* has been linked to colon cancer<sup>35</sup>, but it was not observed in our study, although it was detected in OSCC tissues in another report<sup>23</sup>. *Capnocytophaga* levels were significantly higher in the saliva of lung cancer patients<sup>36</sup> than in healthy controls, and *Capnocytophaga gingivalis* was previously suggested to be a potential salivary biomarker of oral cancer<sup>37</sup>. In this study, *C. gingivalis* was detected at higher levels in control samples without any significance, while *C. leadbetteri* and *C. sp\_oral\_taxon\_902* were remarkably overabundant in lesions. Members of the genus *Selenomonas* have been repeatedly associated with periodontal disease, although the *Selenomonas* species detected in this study did not correlate with known diseases<sup>38</sup>. Several species of *Peptostreptococcus* and *Parvimonas* were extensively enriched in cancer samples, including *Peptostreptococcus stomatis* and *Parvimonas micra*, both of which are reportedly related to colorectal cancer<sup>39</sup>. *Eikenella corrodens*, a fastidious gram-negative facultative anaerobic bacillus, was also detected in another study<sup>24</sup>. The genus *Eikenella* is significantly overrepresented in colorectal cancer<sup>5</sup> and is associated with HPV-negative head and neck squamous cell carcinoma samples<sup>22</sup>. Given its documented history of pathogenicity, further investigation of the potential role of *E. corrodens* in the etiology of OSCC is warranted. In our design, paired lesion and control samples were procured from one individual, eliminating inter-individual variation. Therefore, even slight differences in the bacterial profiles between groups may be closely correlated with OSCC. Although several of the distinguishing taxa were present in relatively tiny proportions, their role in the development of OSCC should not be ignored.

Bacteria coexist in complex interaction webs, and interactions within these webs affect the species involved, while perturbations may contribute to disease<sup>40</sup>. As revealed in our network analysis, bacterial communities in OSCC samples presented with more complex webs depicting ecological relationships, consistent with the extensive bacterial diversity detected in the samples. The genera *Prevotella* and *Neisseria* clustered, forming two of the densest interaction webs in both groups. *Prevotella* and *Neisseria* may play key roles in maintaining the stability of the oral bacterial community across samples. Conversely, an association network centered around *Fusobacterium* arose in the cancer group, indicating that the genus *Fusobacterium* was implicated in the development of OSCC in accordance with its significant increase in the cancer group. *Fusobacterium* tends to co-adhere with other species in oral biofilms by forming bridges between early and late colonizers<sup>41,42</sup>. Thus, it was reasonable to infer a critical role for *Fusobacterium* in increasing OSCC bacterial diversity. Further evaluation of the role of *Fusobacterium* in OSCC may require more study. It was observed that the same paired taxa showed absolute opponent relationships within the groups, implicating that some drastic changes in the bacterial symbiotic relationships occurred during the oral carcinogenesis.

Functions of bacterial communities were inferred using PICRUST based on 16S rRNA sequence information, and differences between controls and diseased samples were further analyzed using the LEfSe method. Overall, the most abundant gene categories were replication and repair, membrane transport, amino acid metabolism, carbohydrate metabolism, and translation, likely reflecting the fundamental requirements for bacterial life in the oral mucosal habitat. As illustrated above, several gene functions exhibited remarkably different distributions within the groups. Notably, among the functions that were significantly decreased in cancer lesions, those related to membrane transport, amino acid metabolism, signal transduction, and lipid metabolism were also under-represented in periodontitis samples<sup>43</sup>. Periodontitis is suggested to be an independent risk factor for OSCC<sup>44</sup>, and bacterially induced chronic inflammation has been anticipated as being involved in oral carcinogenesis<sup>45</sup>. The findings observed in our study, such as several predicted functions in OSCC presenting

same patterns as periodontitis and periodontitis-related taxa being significantly enriched in OSCC, indicate the pro-inflammatory potential of the bacterial communities of OSCC samples, which is consistent with a recent study<sup>24</sup>. Nonetheless, to obtain further details regarding the changes in gene functions of bacterial communities presenting in lesions, whole-metagenome sequencing is warranted in future studies.

In conclusion, bacterial dysbiosis was observed within OSCC surface lesion samples in our study, with drastic changes in bacterial composition and bacterial gene functions compared to controls. In particular, a group of periodontitis-correlated taxa, including *Fusobacterium*, *Dialister*, *Peptostreptococcus*, *Filifactor*, *Peptococcus*, *Catonella* and *Parvimonas*, was found to be significantly enriched in OSCC samples. In addition, several OTUs belonging to *Fusobacterium* were inferred to be heavily involved in OSCC and demonstrated good diagnostic power. The oral microbiota towards OSCC actually is considered like comorbidity factor. According to the current design and observation, it is difficult to determine whether bacterial dysbiosis changed the local microenvironment and then drove carcinogenesis or cancerization in bacterial habitats allowed bacteria suitable for a tumor microenvironment to thrive, resulting in shifts in bacterial communities. More investigation is needed.

## Methods

**Subject recruitment and sample collection.** Subjects with OSCC with a median age of 62 (60% male and 40% female) were recruited from the Department of Oral and Maxillofacial-Head and Neck Oncology of the Ninth People's Hospital (Shanghai, China). All subjects consented to clinical examination and sampling. Subjects did not have detectable periodontal inflammation, visible carious lesions, oral mucosal diseases, or any severe systemic disorders (such as diabetes, immune compromise, or genetic diseases). Moreover, they had not received treatment for OSCC or taken antibiotics at least two weeks prior to sampling. This study was approved by the Ethics Committee of Shanghai Ninth People's Hospital affiliated with the Shanghai Jiao Tong University School of Medicine. Written informed consent was obtained from all subjects. All experiments were performed in accordance with the approved guidelines.

According to a well-defined clinical protocol, swabs of oral lesions and anatomically matched normal sites were collected. Subjects were prevented from drinking and eating for at least 2 h before sampling. All samples were transported to the laboratory on ice within 2 h of collection and were stored at  $-80^{\circ}\text{C}$  before subsequent processing.

**DNA extraction.** Metagenomic DNA was individually extracted from swabs using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The quantity and quality of the isolated DNA was measured with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and by performing agarose gel electrophoresis, respectively. DNA samples were frozen at  $-20^{\circ}\text{C}$  for further analysis.

**Illumina sequencing and bioinformatics analysis of 16S rRNA gene amplicons.** Gene sequencing of 16S rRNA was conducted on an Illumina MiSeq platform according to a previously described protocol<sup>46</sup>. PCR amplifications were performed with the 515 F (5'-GTGCCAGCMGCCGCGGTAA-3') and 926 R (5'-CCGTCAATTTTTRAGTTT-3') primers, which targeted the V4-V5 hypervariable region of the 16S rRNA gene. DNA was amplified following a previously described protocol<sup>47</sup>. Pairs of reads from the original DNA fragments were merged using fast length adjustment of short reads (FLASH) software<sup>48</sup>, and sequences were analyzed using quantitative insights into microbial ecology (QIIME) software<sup>49</sup>. Sequences were assigned to OTUs at 97% similarity; a representative sequence was selected for each OTU, and the RDP classifier was employed to assign taxonomic data to each representative sequence.

Representative sequences were assigned at different taxonomic levels (from phylum to species) to the Human Oral Microbiome Database following the Bayesian approach with a 97% cutoff value. Bacterial diversity was determined by performing a sampling-based OTU analysis and was displayed as a rarefaction curve. Bacterial richness and diversity across samples were assessed using the following  $\alpha$  indexes, which were estimated at a distance of 3%: Chao 1, ACE, Simpson, Shannon and Good's coverage<sup>50</sup>. Student's t-test was used to compare bacterial diversity. GraphPad Prism V.6.0 (San Diego, CA, USA) was used for graph preparation. PCA using unweighted UniFrac distance metrics was carried out, and the R package was used to visualize interactions among bacterial communities in different samples. In addition, PLS-DA, nonparametric analysis of Adonis distance matrices and ANOSIM were also performed to compare bacterial composition between samples<sup>51</sup>. LEfSE (<http://huttenhower.sph.harvard.edu/galaxy/>) was employed to identify distinguishing taxa between the two groups at multiple levels and to visualize the results using taxonomic bar charts and cladograms<sup>52</sup>. Network structures in the bacterial communities of the samples were defined by the Molecular Ecological Network Analysis Pipeline<sup>53</sup> and visualized using Cytoscape<sup>54</sup>. A ROC curve was constructed to determine the diagnostic values of *Fusobacterium* OTUs for OSCC. Co-occurrence patterns for the 30 most abundant taxonomic groups across samples were explored by calculating Pearson correlation coefficients. The results were clustered and visualized using the MeV package<sup>55</sup>. Functional compositions of the bacterial communities were predicted using PICRUSt according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) dataset<sup>56</sup>.

**Data Availability.** The sequencing data from this study have been deposited in the GenBank Sequence Read Archive under accession number SRP097643.

## References

1. Kudo, Y. *et al.* Oral environment and cancer. *Genes Environ* **38**, 13 (2016).
2. Turnbaugh, P. J. *et al.* The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* **449**, 804–810 (2007).



3. Flint, H. J., Duncan, S. H., Scott, K. P. & Louis, P. Interactions and competition within the microbial community of the human colon: links between diet and health. *Environ Microbiol* **9**, 1101–1111 (2007).
4. Burcelin, R. Gut microbiota and immune crosstalk in metabolic disease. *Mol Metab* **5**, 771–781 (2016).
5. Allali, I. *et al.* Gut microbiome compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain. *Gut Microbes* **6**, 161–172 (2015).
6. Dewhirst, F. E. *et al.* The human oral microbiome. *J Bacteriol* **192**, 5002–5017 (2010).
7. Gholizadeh, P. *et al.* Role of oral microbiome on oral cancers, a review. *Biomed Pharmacother* **84**, 552–558 (2016).
8. Darveau, R. P. Periodontitis: a polymicrobial disruption of host homeostasis. *Nat Rev Microbiol* **8**, 481–490 (2010).
9. Chen, H. *et al.* A Filifactor alovis-centered co-occurrence group associates with periodontitis across different oral habitats. *Sci Rep* **5**, 9053 (2015).
10. Ng, J. H., Iyer, N. G., Tan, M. H. & Edgren, G. Changing epidemiology of oral squamous cell carcinoma of the tongue: A global study. *Head Neck* **39**, 297–304 (2017).
11. Perera, M., Al-Hebshi, N. N., Speicher, D. J., Perera, I. & Johnson, N. W. Emerging role of bacteria in oral carcinogenesis: a review with special reference to perio-pathogenic bacteria. *J Oral Microbiol* **8**, 32762 (2016).
12. Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**, 5721–5732 (2005).
13. Bultman, S. J. Emerging roles of the microbiome in cancer. *Carcinogenesis* **35**, 249–255 (2014).
14. Nagy, K. N., Sonkodi, I., Szoke, I., Nagy, E. & Newman, H. N. The microflora associated with human oral carcinomas. *Oral Oncol* **34**, 304–308 (1998).
15. Hooper, S. J. *et al.* Viable bacteria present within oral squamous cell carcinoma tissue. *J Clin Microbiol* **44**, 1719–1725 (2006).
16. Hooper, S. J. *et al.* A molecular analysis of the bacteria present within oral squamous cell carcinoma. *J Med Microbiol* **56**, 1651–1659 (2007).
17. Pushalkar, S. *et al.* Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiol* **12**, 144 (2012).
18. Bebek, G. *et al.* Microbiomic subprofiles and MDR1 promoter methylation in head and neck squamous cell carcinoma. *Hum Mol Genet* **21**, 1557–1565 (2012).
19. Wang, H. *et al.* Microbiomic differences in tumor and paired-normal tissue in head and neck squamous cell carcinomas. **9**, 14 (2017).
20. Pushalkar, S. *et al.* Microbial diversity in saliva of oral squamous cell carcinoma. *FEMS Immunol Med Microbiol* **61**, 269–277 (2011).
21. Schmidt, B. L. *et al.* Changes in abundance of oral microbiota associated with oral cancer. *PLoS One* **9**, e98741 (2014).
22. Guerrero-Preston, R. *et al.* 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, Human Papilloma Virus infection and surgical treatment. *Oncotarget* (2016).
23. Al-Hebshi, N. N., Nasher, A. T., Idris, A. M. & Chen, T. Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: application to oral carcinoma samples. *J Oral Microbiol* **7**, 28934 (2015).
24. Al-Hebshi, N. N. *et al.* Inflammatory bacteriome featuring *Fusobacterium nucleatum* and *Pseudomonas aeruginosa* identified in association with oral squamous cell carcinoma. **7**, 1834 (2017).
25. Khatoun, J., Rai, R. P. & Prasad, K. N. Role of *Helicobacter pylori* in gastric cancer: Updates. *World J Gastrointest Oncol* **8**, 147–158 (2016).
26. Plummer, M. *et al.* Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Health* **4**, e609–616 (2016).
27. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012**, 251364 (2012).
28. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol Cell* **58**, 586–597 (2015).
29. Sun, D.-L., Jiang, X., Wu, Q. L. & Zhou, N.-Y. Intragenomic Heterogeneity of 16S rRNA Genes Causes Overestimation of Prokaryotic Diversity. *Applied and Environmental Microbiology* **79**, 5962–5969 (2013).
30. Hu, X., Zhang, Q., Hua, H. & Chen, F. Changes in the salivary microbiota of oral leukoplakia and oral cancer. *Oral Oncol* **56**, e6–8 (2016).
31. Li, Y. *et al.* Phylogenetic and functional gene structure shifts of the oral microbiomes in periodontitis patients. *Isme j* **8**, 1879–1891 (2014).
32. Goncalves, C. *et al.* Association of three putative periodontal pathogens with chronic periodontitis in Brazilian subjects. *J Appl Oral Sci* **24**, 181–185 (2016).
33. Binder Gallimidi, A. *et al.* Periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget* **6**, 22613–22623 (2015).
34. Ogrendik, M. Oral bacteria in pancreatic cancer: mutagenesis of the p53 tumour suppressor gene. *Int J Clin Exp Pathol* **8**, 11835–11836 (2015).
35. Toprak, N. U. *et al.* A possible role of *Bacteroides fragilis* enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect* **12**, 782–786 (2006).
36. Yan, X. *et al.* Discovery and validation of potential bacterial biomarkers for lung cancer. *Am J Cancer Res* **5**, 3111–3122 (2015).
37. Mager, D. L. *et al.* 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 (2005).
38. Nagpal, D., Prakash, S., Bhat, K. G. & Singh, G. Detection and comparison of *Selenomonas sputigena* in subgingival biofilms in chronic and aggressive periodontitis patients. *J Indian Soc Periodontol* **20**, 286–291 (2016).
39. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).
40. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat Rev Microbiol* **10**, 538–550 (2012).
41. Nobbs, A. H., Jenkinson, H. F. & Jakubovics, N. S. Stick to your gums: mechanisms of oral microbial adherence. *J Dent Res* **90**, 1271–1278 (2011).
42. Abed, J. *et al.* Fap2 Mediates *Fusobacterium nucleatum* Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe* **20**, 215–225 (2016).
43. Wang, J. *et al.* Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Sci Rep* **3**, 1843 (2013).
44. Tezal, M. *et al.* Chronic periodontitis and the incidence of head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* **18**, 2406–2412 (2009).
45. Hooper, S. J., Wilson, M. J. & Crean, S. J. Exploring the link between microorganisms and oral cancer: a systematic review of the literature. *Head Neck* **31**, 1228–1239 (2009).
46. Zhang, G., Wei, L., Chang, C. C., Zhang, Y. & Wei, D. Molecular Biological Methods in Environmental Engineering. *Water Environ Res* **88**, 930–953 (2016).
47. Xiao, C., Ran, S., Huang, Z. & Liang, J. Bacterial Diversity and Community Structure of Supragingival Plaques in Adults with Dental Health or Caries Revealed by 16S Pyrosequencing. *Front Microbiol* **7**, 1145 (2016).
48. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
49. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336 (2010).

50. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**, 7537–7541 (2009).
51. Warton, D. I., Wright, S. T. & Wang, Y. Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* **3**, 89–101 (2012).
52. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60 (2011).
53. Deng, Y. *et al.* Molecular ecological network analyses. *BMC Bioinformatics* **13**, 113 (2012).
54. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
55. Mar, J. C., Matigian, N. A., Quackenbush, J. & Wells, C. A. attract: A method for identifying core pathways that define cellular phenotypes. *PLoS One* **6**, e25445 (2011).
56. Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**, 814–821 (2013).

## Acknowledgements

This research was financially supported by Shanghai Summit & Plateau Disciplines. We acknowledge Shanghai Personal Biotechnology Co., Ltd. for their kind help with gene sequencing and bioinformatics analysis.

## Author Contributions

Conception and design of the experiments: J.L., Z.H. and B.H. Performed the experiments: H.Z., M.C. and S.R. Data processing and analysis: H.Z. and Z.H. Volunteer recruitment and sample collection: H.Z., C.Z., M.C. and X.Y. Paper writing: H.Z. Revision of the manuscript: J.L., C.Z., B.H., Z.H., M.C. and H.Z.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-11779-9

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017