Oral Microbial Species and Virulence Factors Associated with Oral Squamous Cell Carcinoma

Manolito G. Torralba¹ · Gajender Aleti¹ · Weizhong Li¹ · Kelvin Jens Moncera¹ · Yi-Han Lin¹ · Yanbao Yu² · Michal M. Masternak³ · Wojciech Golusinski⁴ · Pawel Golusinski^{4,5} · Katarzyna Lamperska⁶ · Anna Edlund¹ · Marcelo Freire¹ · Karen E. Nelson¹

Received: 20 April 2019 / Accepted: 1 September 2020 / Published online: 6 November 2020
© The Author(s) 2020, corrected publication 2020

Abstract

The human microbiome has been the focus of numerous research efforts to elucidate the pathogenesis of human diseases including cancer. Oral cancer mortality is high when compared with other cancers, as diagnosis often occurs during late stages. Its prevalence has increased in the USA over the past decade and accounts for over 40,000 new cancer patients each year. Additionally, oral cancer pathogenesis is not fully understood and is likely multifactorial. To unravel the relationships that are associated with the oral microbiome and their virulence factors, we used 16S rDNA and metagenomic sequencing to characterize the microbial composition and functional content in oral squamous cell carcinoma (OSCC) tumor tissue, non-tumor tissue, and saliva from 18 OSCC patients. Results indicate a higher number of bacteria belonging to the Fusobacteria, Bacteroidetes, and Firmicutes phyla associated with tumor tissue when compared with all other sample types. Additionally, saliva metaproteomics revealed a significant increase of *Prevotella* in five OSCC subjects, while *Corynebacterium* was mostly associated with ten healthy subjects. Lastly, we determined that there are adhesion and virulence factors associated with *Streptococcus gordonii* as well as from known oral pathogens belonging to the *Fusobacterium* genera found mostly in OSCC tissues. From these results, we propose that not only will the methods utilized in this study drastically improve OSCC diagnostics, but the organisms and specific virulence factors from the phyla detected in tumor tissue may be excellent biomarkers for characterizing disease progression.

Keywords Oral cancer · Sequencing · Microbiome · Oral microbiome · Oral squamous cell carcinoma · *Fusobacterium*

Introduction

Oral cancer (OC) includes any cancer that affects head and neck tissues including mouth cancer, tongue cancer, tonsil

cancer, and throat cancer. Over 40,000 individuals in the USA are diagnosed with a form of OC each year, and more than 481,000 new patients are diagnosed annually worldwide [1, 2]. The etiology of OC is multifactorial, and the incidence

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00248-020-01596-5>) contains supplementary material, which is available to authorized users.

✉ Manolito G. Torralba
mtorralb@jcv.org

¹ Department of Genomic Medicine, J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, CA 92037, USA

² Department of Genomic Medicine, J. Craig Venter Institute, 9605 Medical Center Drive Suite 150, Rockville, MD 20850, USA

³ Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, 6900 Lake Nona Blvd, Central Florida Blvd, Orlando, FL 32827, USA

⁴ Department of Head and Neck Surgery, Poznan University of Medical Sciences, The Greater Poland Cancer Centre, Garbary 15, 61-866 Poznan, Poland

⁵ Department of Otolaryngology and Maxillofacial Surgery, University of Zielona Gora, Podgórze 50, 65-246 Zielona Góra, Poland

⁶ Laboratory of Cancer Genetics, Greater Poland Cancer Centre, 15th Garbary Street, room 5025, 61-866 Poznan, Poland

is increasing. Based on the available evidence, etiological factors include tobacco usage, excess consumption of alcohol, betel quid usage, HIV, HPV, periodontal diseases, exposure to sunlight, and ethnicity [3–5]. For most countries, 50% of individuals with cancers of the tongue, oral cavity, or oropharynx present 5-year survival rates. In case of late diagnosis, survival rates drop to 15%. Though there have been limited improvements in OC treatment, the mortality rate remains high at 43%, and the 5-year survival rate is discouraging at 56%. These bleak mortality numbers, however, are in part due to a majority of the diagnoses occurring during the late stages of OC development [1, 6]. The common occurrence of these late diagnoses can be attributed to a lack of consistent OC prevention programs, the inability to identify early-stage OC in asymptomatic individuals, and the lack of consistent biological markers to use for diagnostic purposes [7–9]. Fortunately, according to the Surveillance, Epidemiology, and End Results (SEER) database and the National Cancer Institute (NCI), OC has an 80–90% relative survival rate when found early compared with the overall population [10]. In oral squamous cell carcinoma (OSCC), the 5-year survival rate of 80% is reduced to 20–40% if the cancer is diagnosed at later stages (T3 and T4). One major factor behind the mechanism of high mortality in OSCC is the lack of early-stage molecular markers. Fortunately, to improve early diagnosis, the use of biofilm and saliva for the detection of OC has started to come to fruition with promising results [5, 11–15]. However, since there are no clear and consistent microbial biomarkers associated with OSCC progression, it is likely that using multi-omic approaches such as molecular characterization of OSCC as outlined in our study can lead to improved diagnostics, prevent late cancer stage detection, as well as increasing the potential for novel therapeutics.

Advancements in genomics approaches have allowed researchers to characterize entire microbial communities to understand complex interactions between the host and its microbiome. Conserved genomic markers such as the 16S rDNA gene (16S) have been used to characterize microbial diversity in these communities to further understand how their interactions with the host may contribute to health or pathogenesis. Although previous studies that utilized culture-based approaches, PCR, and sequencing of the oral microbiome revealed abundances and diversity of pathogens, there is a lack of multi-omic approaches to map the bacterial virulence associated with the taxonomic changes that are likely shaping disease phenotypes. Additional genomics approaches including metagenomics and transcriptome sequencing can also be used to understand these interactions. Efforts such as the Human Microbiome Project (HMP) have pioneered these approaches, which continue to have a significant impact on increasing our understanding of host–microbiome interactions [16]. These efforts, when applied to focus on specific interactions between the host and their microbial communities, are

critical to understanding the pathology of complex diseases such as cancer. OC is one cancer where these approaches can provide novel insight into understanding pathology and most importantly into improving diagnostic approaches.

Previous studies have shown various associations between specific microorganisms and cancer [4, 17–19]. Most notable are the associations of *Helicobacter pylori* and gastric cancers as well as human papillomavirus (HPV) and cervical, genital, mouth, and throat cancers [20–24]. Studies have shown an association of various species of *Streptococcus* with OSCC using Sanger sequencing, Roche 454 pyrosequencing, and DGGE profiling in oral swabs and saliva from healthy and OSCC patients [5, 25–27]. These results, however, are perplexing since there are species of *Streptococcus* that are classified as commensal bacteria such as *Streptococcus salivarius*, *S. sanguinis*, *S. oralis*, and *S. mitis*, given their association with healthy oral microbial communities [28, 29] contrary to other species such as *S. mutans*, *S. intermedius*, and *S. constellatus*, which are known to be associated with the development of caries, dental plaque, and periodontal disease [28, 30, 31]. Additionally, when analyzing broader taxonomic differences, at the genus and phylum level, results from these studies are often inconsistent. For example, Schmidt and colleagues used 16S rRNA amplicon sequencing to show a reduction in *Streptococcus* and *Rothia* in OC, whereas Guerro-Preston et al. showed the opposite [32, 33]. Wang et al. also used 16S rRNA amplicon sequencing and showed little variation in alpha diversity between tumor and non-tumor tissues along with demonstrating the low abundance of *Actinomyces* and high abundance of *Parvimonas* in tumor tissues while having no mention of *Streptococcus* abundance in their study [34]. Lastly, Schmidt et al. (2014) showed decreased relative abundance of *Streptococcus* and *Rothia* in oral swabs from tumor lesions when compared with contralateral swab controls and healthy individuals [32]. The inconsistencies in existing studies clearly suggest a new approach to understanding the microbial community as it relates to OSCC pathology is necessary.

Though there have been a number of attempts to characterize the microbial communities associated with OSCC [5, 13, 35], these associations have yet to be fully determined. In the study presented here, we validated our findings by using high throughput sequencing on tumor and contralateral non-tumor oral tissues in conjunction with proteomics approaches on saliva derived from OSCC patients. More importantly, analyses of tissues allowed for a thorough assessment of the microbial community particularly bacteria such as *Fusobacterium* that have adherent and invasive characteristics and have been implicated in colorectal cancer [15, 36, 37] and leukoplakia [38]. More recent transcriptomics analysis by Yost et al. (2018) supports the role that *Fusobacterium* may have in the development of OSCC, showing a higher number of fusobacterial transcripts in tumor adjacent samples [39]. In the present study, we compare 16S rRNA amplicon and

metagenomics sequences of tumor and healthy tissues of OSCC patients. From this dataset, we have been able to characterize candidate virulence factors from known oral pathogens that appear to be exclusive to tumor tissues. Additionally, we have profiled the host proteome of salivary OSCC samples [14] and further evaluated the metaproteomic composition of these samples to validate our sequencing data. To the best of our knowledge, this study is the first oral microbiome study that utilizes deep Illumina sequencing in conjunction with proteomics approaches to link microbial community structure and function in oral tissues with OSCC.

Methods

Sample Collection

Samples including non-tumor tissue ($n = 18$), tumor tissue ($n = 18$), and saliva ($n = 18$) were collected from OSCC patients enrolled in the study at the Greater Poland Cancer Centre (GPCC) Poznan, Poland. Of the samples collected, seven tumors were excised from the palatine tonsil, three were excised from the throat, three were excised from the bottom of the oral cavity, and five were excised from the tongue. Matched contralateral non-tumor tissues were also collected. Additional saliva samples ($n = 5$) collected from other members of the same cohort were randomly selected for proteomics analysis to further support data generated from the first 18 participants. Saliva samples from the original 18 samples were exhausted after DNA extraction and prior to proteomics analysis resulting in the use of saliva samples from other members of the cohort. Inclusion criteria included patients who only had surgery as a primary treatment. All patients with the previous diagnosis of any cancer, history of cancer treatment, including radiation and chemotherapy, and HPV-positive tumors were excluded from this study. Tumor tissue (TT) and contralateral non-tumor tissue (NT) was surgically excised from each patient in the outpatient clinic of the head and neck surgery department at the GPCC. Saliva was collected from OSCC patients following standard operating procedures for saliva sample collection as outlined by the Human Microbiome Project (HMP) [40]. Relevant clinical metadata including age, gender, and tumor location was collected. All samples were flash-frozen in liquid nitrogen and stored in ultra-low freezers ($-80\text{ }^{\circ}\text{C}$) after collection. Samples were later transported on dry ice to the J. Craig Venter Institute (JCVI). All authors declare that all methods in this study followed the protocol approved by the Institutional Review Board of Poznan University of Medical Sciences in Poznan, Poland, under IRB# 412/18. All experiments were performed in accordance with relevant guidelines and regulations. Informed consent for participation in the study was obtained from all patients included in the study.

DNA Extraction, 16S rRNA PCR, and Sequencing

DNA from 18 patient TT, NT, saliva, and blank negative sample controls was extracted using lysozyme and proteinase K enzymatic digests followed by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. We utilized 20–30 mm diameter of tissue for each extraction along with blank extraction negative controls. Extracted DNA was quantified using Sybr Gold (Thermo Fisher, Waltham, MA) via TECAN assay (Tecan Systems, Inc., San Jose, CA) in preparation for 16S PCR. Samples, extraction negative controls, and no template controls were amplified using Platinum Taq DNA polymerase (Thermo Fisher) and barcode and adaptor-ligated 16S primers [41] targeting the V4 region of the 16S rRNA gene fragment [42, 43]. Cycling conditions included $95\text{ }^{\circ}\text{C}$ initial denaturing step for 5 min, followed by 35 cycles of $95\text{ }^{\circ}\text{C}$ for 30 s, $57\text{ }^{\circ}\text{C}$ for 30 s, and $72\text{ }^{\circ}\text{C}$ for 30 s, followed by a final extension step of $72\text{ }^{\circ}\text{C}$ for 5 min and $4\text{ }^{\circ}\text{C}$ hold. Amplicons were then purified using the QIAquick PCR purification kit (Qiagen Inc., Hilden, Germany) in order to remove residual dNTPs and primer dimers according to the manufacturer's specification. Purified amplicons were then quantified via TECAN, normalized, and pooled in preparation for 16S sequencing using the MiSeq platform (Illumina, La Jolla, CA) with V2 chemistry 500 cycles, $2 \times 250\text{ bp}$ dual index format using standard manufacturer's specifications. Sequences generated from 16S amplicons are available in Genbank under NCBI project PRJNA666891 stored as SRP ID SRP286018 in the Short Read Archive (SRA).

Metagenomics Sequencing

We selected genomic DNA (gDNA) extracted from healthy and tumor tissue from eight patients based on tumor location type. Eight TT samples were selected according to their anatomical location: palatine tonsil ($n = 2$), bottom of the oral cavity ($n = 2$), throat ($n = 2$), and tongue ($n = 2$), in addition to contralateral NT samples from the same patients to serve as controls. TT and NT gDNA extracts were enriched for microbial DNA using the NEB Microbiome Enrichment kit using standard manufacturer's specifications (New England Biolabs, Ipswich, MA). Metagenomic libraries were prepared from 100 ng of enriched microbial DNA using the NEBNext Ultra II DNA library preparation kit for Illumina sequencing (New England Biolabs). Libraries were prepared by targeting 300–400 bp fragments for adapter ligation and amplification. Six PCR cycles were utilized to generate fragment libraries in preparation for sequencing. PCR reactions were purified according to specifications provided by NEB. Libraries were then checked for quality control using High Sensitivity DNA Lab on a Chip (Agilent Technologies, Santa Clara, CA) and the KAPA Library quantification kit according to manufacturer's specification. The libraries that passed quality control were then normalized and pooled in preparation for

sequencing on the Illumina NextSeq platform. Libraries were sequenced using the NextSeq 500, Mid Output kit, 300 cycles 2×150 bp, using manufacturer's specifications (Illumina Inc.). Sequences that did not pass quality filtering were discarded, while high-quality sequences were binned according to Illumina indexes utilized during library preparation. Metagenomic sequences were deposited into Genbank under NCBI project PRJNA666891 stored as SRP ID SRP286018.

Illumina 16S Sequence Processing and Data Analysis

Unprocessed sequences were demultiplexed using CASAVA v1.8.2 (Illumina Inc., La Jolla, CA) to produce individual .fastq files for each pair of corresponding dual indexes. The DNA sequences were processed to ensure that only sequences with quality scores $>Q30$ were applied to the mothur pipeline in addition to utilizing stringent settings to ensure that no barcode mismatches were permitted among the demultiplexed reads [30]. Additional sequence filtering was used by applying the screen.seqs function of mothur to remove all sequences shorter than 220 bp [44]. Other QC steps were implemented, and the sequences were aligned against the SILVA database version 132 [45] to confirm the orientation of noise-filtered sequences along with ensuring the correct positioning of the reads with respect to which variable regions were amplified and sequenced. The sequences passing QC were then checked for chimeras using chimera slayer in mothur [44]. Sequences were classified taxonomically using the RDP classifier, and hits matching mitochondria, chloroplast, archaea, unknown, and eukaryote were eliminated to avoid noise on the data [46]. Sequence reads were then clustered at various taxonomic levels including 97% rDNA sequence similarity (OTU), genus, and phylum level. OTU table was imported in R and rarefied with a minimum library size of 4611 reads using vegan package version 2.5 [47]. Significant differences between sample groups were tested using pairwise permutation multivariate analysis of variance (PERMANOVA) with 999 permutations using ADONIS within RVAideMemoire package. Taxonomic differences between sample types were calculated using the SIMPER function in vegan. Core microbiome was computed using the script compute_core_microbiome.py in Qiime [48]. OTUs present in at least 65% of samples in each group were considered as core OTUs. Functional profiles were predicted based on 16S rRNA data using the Tax4fun software package and SILVA database [49]. Random Forest analysis was applied to the 16S rRNA data using machine learning methods developed by Leo Breiman to identify the most significant microbial features of our dataset. Features with a minimum prevalence of 10% across samples were included, and those with >0.005 accuracy were considered significant. Data was further transformed to centered log ratio (CLR) before applying the Random Forest classification algorithm.

Functional Potential of Non-Tumor and Tumor Tissue-Associated Metagenomes

Sequences were filtered for low-quality reads and human sequence contaminants using KneadDATA version 0.5.4 (available at <http://huttenhower.sph.harvard.edu/kneaddata>). Reads were scanned with a four-base wide sliding window and trimmed when the average base Phred score dropped below 20. Further, reads that were shorter than 70 nucleotides were discarded. Human genome assembly version hg38 (available at <https://www.ncbi.nlm.nih.gov/grc/human>) was used as a reference for removal of human contaminant sequences from the sequencing data. Taxonomy profiling of metagenomic sequences after quality filtering and human reads removal was performed by mapping the reads to a JCVI in-house microbial reference genome database with Centrifuge [50]. The reference database includes 27,115 representative genomes covering bacteria, archaea, viruses, fungi, and microbial eukaryotes species selected from NCBI RefSeq genomes. Species with relative abundance at least $1e-4$ reported by Centrifuge were kept in this study.

Mapping the biosynthetic gene cluster (BGC) abundances for functional annotations was performed as described previously [51]. Briefly, both forward and reverse quality filtered FASTQ reads were aligned using BWA-MEM [52], Burrows–Wheeler aligner-maximum exact matches, with default parameters against the oral BGC database encompassing 4915 BGCs from oral bacteria [51]. For each gene, we summarized the counts of mapped reads from SAM alignment files using custom Perl script. We then applied the awk command to merge the individual count files generated from several mapping events. Count files were normalized using the DESeq2 pipeline [53] in R version 3.4.0 (<https://www.r-project.org/>), and size factors were determined using the median ratio method “poscounts,” which compensates for geometric means in samples where all genes have a value of zero. For significance testing, we performed 2-way ANOVA (Fisher's least significant difference) on the DESeq normalized count data since negative binomial model-based Wald significance test (implemented by DESeq2) is more stringent for low count genes.

A virulence factor database was constructed by downloading 4,071,128 open reading frames (ORFs) from 490 well-curated oral bacterial genomes, available via the expanded Human Oral Microbiome Database (eHOMD) (available at http://www.homd.org/ftp/HOMD_prokka_genomes/). By using BLASTN, these ORFs were searched against the virulence factor database (VFDB) version 2019 (available at <http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi>), containing 3204 experimentally verified virulence genes of pathogenic bacteria. Closest matching amino acid sequences were selected with a minimum sequence identity cutoff at 75% and with a coverage of 75%. We adapted the above read

mapping approach to evaluate relative abundances of these virulence genes in TT-, NT-, and saliva-associated metagenomes [54]. In order to quantify the abundances of antibiotic resistance genes (ARGs) in NT and TT metagenomes, ORFs from oral bacterial genomes were BLASTN searched against the antimicrobial resistance database MEGARes [55], containing a collection of 4000 curated ARG sequences. Subsequently, we mapped metagenome reads against potential oral ARGs to evaluate the abundances and performed enrichment analysis using 2-way ANOVA on the DESeq2 normalized counts as described above. In addition, for in-depth analysis of *Fusobacterium* functional potential, we aligned metagenomes against the 132-putative virulence and antibiotic resistance gene sequences collected mainly from multiple strains of *F. nucleatum* and *F. polymorphum* genomes. We applied the same read mapping approach as described above to analyze the differential enrichment of genes in tissue samples from healthy and cancer sites. In order to distinguish disease-associated genes, we further clustered DESeq2 normalized gene abundances based on Pearson's correlation coefficient.

Metaproteomics Analysis of OSCC Saliva

Five OSCC saliva samples from the same cohort were randomly selected along with five control saliva samples from a non-OSCC cohort and were subjected to metaproteomics analyses as previously described [14]. In brief, whole saliva samples were first thawed on ice and an aliquot of similar protein amount (20–30 µg) from each sample after SDS PAGE estimation [56] was digested using suspension trapping (STrap) approach with in-house (JCVI)-made glass fiber filters (Whatman GF/F, 0.7 µm). The protein digests were then analyzed by a nanoLC and Q-Exactive MS system (ThermoFisher Scientific, Waltham, MA) following a 150 min gradient on a 19-cm reversed phase column, particle size 3.0 µm, ReproSil-Pur C18-AQ media (Dr. Maisch GmbH, Ammerbuch-Entringen Germany). Global protein identifications were obtained by searching tandem mass spectra against a combined database of the Human Oral Microbiome Database (<http://www.homd.org/>, 1,079,626 protein sequences) and the UniProt human proteins (<https://www.uniprot.org/proteomes/UP000005640>, 20,349 reviewed sequences) using the Proteome Discoverer software (version 2.2, Thermo Scientific) and Sequest-HT algorithm [57].

Results

Sequencing Data Statistics

Post sequencing trimming and quality control of raw 16S sequences resulted in 772,366 sequence reads, with each

sample averaging approximately 14,300 reads. Sequence reads are available at the Short Read Archive (SRA) under accession number (*will update once available*). OTU table was rarefied (normalized) to an even depth of 4611 reads. All samples generated an adequate number of reads for 16S analysis. Negative controls generated minimal sequence data and were not included in our analysis. Metagenomic sequencing on the resulted in approximately 101 M sequence reads, including both human and bacterial, with each sample averaging 6.36 M sequence reads after trimming and QC. After human sequences were removed, 5–16% (~390,000 to ~1.5 M) of the reads belonged to bacteria. Five metagenomic samples: P05-TT, P19-NT, P19-TT, P20-NT, and P20-TT, did not produce adequate sequence coverage and were omitted from our metagenomic analysis.

Microbial Composition Associated with Tumor Tissue, Non-Tumor Tissue, and Saliva

The 16S sequence data showed that the microbial composition varied between TT and NT, while saliva samples mostly clustered with NT (Fig. 1). Using the *adonis* function in the *vegan* R package, we determined that the microbial composition associated with TT was significantly distinct when compared with contralateral NT ($p = 0.001$, $R^2 = 0.0757$). SIMPER analysis [58] indicated that the top five discriminating genera between the TT and NT sample types included *Leptotrichia*, *Peptostreptococcus*, *Megasphaera*, *Rothia*, and one species of *Prevotella*. Other genera of interest included *Fusobacterium*, other species of *Prevotella*, and *Actinomyces*, which were in higher relative abundance in the TT when compared with the NT from OSCC samples (Fig. 2). Additionally, the sequence data showed that *Streptococcus*, *Veillonella*, and *Rothia* were higher in abundance in the NT than TT (Fig. 2). Random Forest analysis also indicated that TT was also shown to be highly abundant in *Fusobacterium*, *Parvimonas*, and *Mogibacterium* (Fig. 3). Sequencing analysis of saliva collected from matched TT and NT patient samples suggests that saliva samples appeared to be more similar to NT than TT (Fig. 1). Comparing the most abundant and core microbial taxa between the three sample types revealed three discernable communities in the oral cavity. Log2 fold changes in abundance of specific taxa was evident between the sample types (Fig. 4a–d). Taxa revealed to be in common between sample types as well as exclusive to each sample type is highlighted in Fig. 4e. Network analysis showed that co-occurrence between microbes within the sample types was revealed to be of high incidence in the TT and NT while low in saliva (Fig. 4e).

Fig. 1 Principal coordinate analysis (PCoA) of samples based on the Bray-Curtis dissimilarity matrix. Each filled circle here represents a microbial community associated with non-tumor tissue (red), tumor tissue (blue), and saliva (green) with 95% confidence ellipses showing mean values of groups. A separation between groups was tested by comparing the principal coordinates using PERMANOVA statistical test (p value <

0.001). Further pairwise comparisons displayed significant differences in bacterial community composition among tumor and non-tumor ($p = 0.00075$) and saliva and tumor ($p = 0.00075$). Gaussian distribution of multivariate analysis of variance was tested to verify PERMANOVA assumptions (p value > 0.07)

Abundance of Oral Pathogens in Tumor Tissue and Non-Tumor Tissue

Analysis of 16S and metagenomic data revealed a relative abundance of microbial and fungal communities in all three sample types that illustrate a higher abundance of known oral pathogens exclusive to TT (Figure S1). TT microbial composition was higher in the abundance of known and opportunistic oral pathogens while having a decreased amount of known oral commensal bacteria when compared with NT. Taxonomic analysis revealed that a substantial percentage of sequence data belonging to genera known to contain oral pathogens or opportunistic oral pathogens such as *Prevotella* (17%), *Streptococcus* (17%), *Parvimonas* (8%), *Fusobacterium* (8%), *Peptostreptococcus* (2%), and *Porphyromonas* (2%) was present in TT. Our metagenomics data indicated that fusobacterial species in high abundance in TT included *Fusobacterium* sp. oral taxon 370 and *F. nucleatum*. *F. periodonticum* and *F. necrophorum* also appeared to be in high abundance in patients 25 and 7, respectively, while three patients, 7, 1, and 6, had higher levels of *Prevotella* species present in TT (Fig. 5). Additionally, the increased presence of *Parvimonas micra* in 63% of TTs suggests an exclusive association with OSCC (Fig. 5). Our data also showed that the percentage of sequence data belonging to

genera known to contain oral commensal bacteria such as *Streptococcus* (33%) and *Veillonella* (12%) was higher in NT. Additionally, oral pathogens and opportunistic pathogens observed in TT are demonstrated to be in much lower abundance in NT, inclusive of *Parvimonas* (0.7%), *Fusobacterium* (2%), *Peptostreptococcus* (0.4%), and *Porphyromonas* (0.6%).

Functional Potential and Association of Virulence Factors with OSCC Tissue

Analysis of functional potential of metagenome sequences showed 35 distinct BGCs, which were significantly abundant in TT samples when compared with NT samples. Relative increased abundance from the TT and NT samples are highlighted in blue and yellow, respectively (Table 1). Metagenomic sequencing also revealed several virulence factors that were significantly abundant in TT when compared with NT (Table 2). Open reading frames from our dataset searched against the VFDB revealed 9326 virulence homologs covering the broad diversity of 304 oral bacterial taxa (potential oral virulence gene sequences and their corresponding annotations are provided in the supplementary information). Additionally, we identified 3584 homologs with sequence identity > 75% and with a coverage of sequence length

Fig. 2 Box plots of the relative abundance of known oral pathogens and commensals. The top three box plots that indicate genera known to contain potentially pathogenic oral microbial species are enriched in

TT, while the bottom three box plots that indicate genera commonly associated with oral health are in higher abundance in NT

> 75% (potential ARG sequences and the annotations are provided in the supplementary information). Gene annotation showed that these virulence factors included enzymes with various functions, such as metabolic regulators, and extracellular surface protein structures. Two-way ANOVA testing showed that *Streptococcus pneumoniae* choline binding protein, *Haemophilus somnus* glycosyl transferases, along with *Acinetobacter baumannii* transcriptional regulator LysR, were enriched in TT; p values = 0.0291, 0.0004, and < 0.0001, respectively. Notably, three variants of fibronectin binding proteins were shown to be significantly abundant in TT. Two-way ANOVA testing showed that *S. gordonii* CshA and CshB fibronectin binding proteins were significantly higher in abundance in TT; p values = 0.0050 and 0.0199, respectively.

Relevance of *Fusobacterium* Virulence Factors and Antibiotic Resistance Functions in Non-Tumor and Tumor Tissue-Associated Metagenomes

Among the 31,729 reads that mapped to *Fusobacterium* virulence factors and antibiotic resistance genes, a substantial fraction of which, 99.48% (31,564 reads), mapped

to cancer metagenomes and only 0.5% (165 reads) of these sequences mapped to non-cancer metagenomes. In total, we identified 62 genes that were differentially enriched in OSCC metagenomes as compared with healthy samples highlighting their pathogenic potential (Fig. 6). These genes displayed features of adhesion, secretion, transport, resistance, and invasion. Genes encoding outer membrane proteins such as bacterial autotransporters, potential adhesion proteins including a fibronectin-binding protein, a possible autotransporter adhesion, and von Willebrand proteins were highly abundant in OSCC metagenomes. In particular, we identified a high abundance of MORN2 sequence repeats, which have been recently associated with actively invading *F. nucleatum* species [59]. In addition, we identified Type V secretion system encompassing autotransporter and two-partner secretion system genes. Also, genes associated with host immune evasion such as lipooligosaccharide sialyltransferase, which incorporates sialic acid into lipopolysaccharide biosynthesis, were more abundant in TT than NT. Additionally, endotoxin-related lipopolysaccharide biosynthetic genes were

Fig. 3 Random Forest classification model was applied on the 16S rRNA abundance data to identify most important microbial features in TT, NT, and saliva samples. Features with at least 4 reads and with a minimum prevalence of 10% across samples were included. Data was further transformed to centered log ratio (CLR) before applying the Random Forest classification algorithm

enriched in TT. Genes involved in iron acquisition such as iron ABC transporters, various iron receptors, and other siderophores were significantly in higher abundance in OSCC metagenomes. With respect to antibiotic resistance functions, we identified a high abundance of several antibiotic transporters including MOP/MATE family multi-drug efflux pumps and drug/metabolite transporters. Lastly, we identified a high abundance of genes associated with lipopolysaccharide production such as glycosyl transferases, flippases, and deacetylases in TT.

OSCC Patient Saliva Revealed Higher Concentrations of Microbial Proteins

Metaproteomics analysis resulted in the identification of 2300 protein groups (FDR < 1%) from the five OSCC saliva samples, including 1132 human proteins and 1168 from bacteria. Meanwhile, to investigate any OSCC-specific metaproteome profile, we analyzed in-parallel another five non-OSCC derived saliva samples from a separate ongoing project. These five non-OSCC samples served as a control. The number of identified unique protein groups, as well as the number of peptide-spectrum

matches (PSMs), was plotted across the 10 samples (Figure S2). On average, 294 and 78 microbial proteins with corresponding 2122 and 543 PSMs were identified from OSCC saliva and control saliva samples, respectively, which contributed to 2.1% and 6.9% of total protein mass. From one particular sample (O9), around 51% of the identified protein groups were from oral bacteria representing 79 distinct genera, suggesting a high diversity of the oral microbiota obtained by metaproteomics approach (Figure S3). However, human proteins constituted almost 81% of the total protein intensities measured in this saliva sample, indicating a possible association of saliva microbiome with abundant host proteins. In another OSCC saliva sample (O36), a significant number of microbial proteins (450) were also identified, constituting 10% of total protein mass. These data have clearly shown the in-depth coverage of saliva metaproteome by the MS-based approach. In the context of microbial diversity, *Prevotella* appeared to be the most abundant genus identified from OSCC saliva, and up to 56% of the oral microorganisms were from the genus *Prevotella* ($30.7 \pm 17.9\%$, $n = 5$). This organism showed significantly ($p < 0.05$) higher abundance in the OSCC saliva when

Fig. 4 Differentially abundant taxa and core microbiome in the tumor, non-tumor, and saliva samples. Log₂ fold changes of **a** tumor tissue vs non-tumor tissue, **b** tumor tissue vs saliva, and **c** non-tumor tissue vs

saliva. **d** Heatmap presenting the relative abundances of differentially abundant genera. **e** A Venn diagram displaying core and unique genera represented at the OTU level (97% similarity)

Fig. 5 Heatmap of microbial abundance in tumor and non-tumor tissues. Species shown on the x-axis in this heatmap are demonstrated to be in higher abundance in tumor samples when compared with non-tumor samples. The individual patient IDs are listed followed by a TT or NT, indicating tumor tissue or non-tumor tissue, respectively

compared with non-OSCC saliva (Figure S2). Other organisms such as *Corynebacterium*, *Enterococcus*, and *Mogibacterium* were shown to have significantly decreased abundance in OSCC saliva when compared with non-OSCC saliva (Figure S2).

Discussion

Mucosal homeostasis between the human microbiome and the host is the key to reducing inflammation and maintaining healthy host tissue. Virulence factors from pathobionts and pathogenic biofilms, such as those formed by *Fusobacterium*, disrupt tissue homeostasis leading to dysbiosis and disease development [60–62]. Uncontrolled mucosal tissue response is a hallmark of oral diseases, including gingivitis, mucositis, periodontitis, endodontic lesions, leukoplakia, and OC. Early stage OC, such as OSCC, has a relatively good prognosis, yet the majority of clinical cases are still diagnosed in advanced stages of the disease with a significant impact on patient prognosis and patient survival. Hence, there is an urgent need for understanding the role of oral pathobionts in OSCC development, thus facilitating the discovery of novel microbial and biofilm biomarkers dictating dysbiosis through the use of multi-omics approaches as done in our study, ultimately leading to the development of molecular diagnostics to improve early detection of OSCC and potentially novel therapeutics.

In the current study, we address the challenges associated with diagnostics of OSCC and obtained tissue samples to uncover microbial differences in tumor and non-tumor samples from each patient. We were able to discriminate our TT and NT samples by using SIMPER and Random Forest

analysis to determine the variation in relative abundance. Our SIMPER analysis results indicated that one of the top five discriminating taxa that is the most abundant in TT when compared with NT, *Leptotrichia*, may potentially contribute to pathogenesis as this genus contains a number of pathogenic species such as *Leptotrichia hofstadii* and *L. stadii*, which have been isolated from oral wounds [63]. Additionally, as a Gram-negative organism, it is likely contributing to a host inflammatory response due to the characteristic lipopolysaccharide cell wall. Review of taxonomic relative abundance also demonstrated a higher proportion of oral genera known to contain pathogenic species in TT when compared with NT. This is particularly of interest as oral pathogens such as *Fusobacterium* have been previously associated with certain cancers [64, 65] and are observed to be in higher abundance in our TT samples emphasizing the significance of this finding in OSCC. Additionally, invasive pathogens such as *Fusobacterium* can be detected in our oral tissue samples while previous studies utilized superficial swabbing sampling techniques, which may not capture invasive oral pathogens associated with OSCC. Random forest analysis supported the results demonstrating higher proportions of Gram-negative organisms such as *Fusobacterium* and *Parvimonas* in TT.

The relative microbial compositions associated with tumor tissues presented significant differences when compared with contralateral non-tumor tissues ($p = 0.001$, $R^2 = 0.0757$). Our results indicated that specific genera known to contain pathogenic and potentially pathogenic oral microbial species and strains such as *Fusobacterium*, *Prevotella*, and *Actinomyces* were detected in high abundance in tumor tissues. The potential pathogenicity of many species in each genera is of significance since oral pathogens such as *Prevotella intermedia* and

Table 1 Biosynthetic gene clusters (BGCs) reveal microbial factors associated with OSCC. List of BGCs and associated bacterial species and significance values. Rows highlighted in yellow indicate BGCs that are higher in abundance in NT, while rows highlighted in blue indicate BGCs that are higher in abundance in TT

Bacterial species	MIBiG accession	Biosynthetic gene cluster	antiSMASH annotation	Mean Diff.	Individual P Value
Leptotrichia goodii F0264	BGC0000766	Exopolysaccharide_biosynthetic_gene_cluster	saccharide	126.1 < 0.0001	
Campylobacter concisus_13826	BGC0000760	Emulsan_biosynthetic_gene_cluster	saccharide	46.42	0.0014
Fusobacterium_sp_oral_taxon_370str_F0437	BGC0000758	Capsular_polysaccharide_biosynthetic_gene_cluster	putative	103.9 < 0.0001	
Fusobacterium_necrophorum_subsp_funduliforme_1_1_36S	BGC0001391	Thusin_biosynthetic_gene_cluster	other	71.06 < 0.0001	
Leptotrichia buccalis_DSM_1135	BGC0000794	S-layer_glycan_biosynthetic_gene_cluster	saccharide	45.95	0.0015
Lactococcus_lactis_subsp_lactis_KF147	BGC0001001	Jamaicamide_biosynthetic_gene_cluster	nps-transapks	67.98 < 0.0001	
Enterococcus_saccharolyticus_30_1	BGC0000796	S-layer_glycan_biosynthetic_gene_cluster	saccharide	34.6	0.0171
Streptococcus_pyogenes_MGAS10270	BGC0000556	Streptin_biosynthetic_gene_cluster	lantipeptide	33.08	0.0226
Corynebacterium_durum_F0235	BGC0000636	Carotenoid_biosynthetic_gene_cluster	terpene	77.07 < 0.0001	
Streptococcus_oligofermentans_AS_1_3089	BGC0000206	Chartreusin_biosynthetic_gene_cluster	putative	51.62	0.0004
Leptotrichia_sp_oral_taxon_879_str_F0557	BGC0000433	Surfactin_biosynthetic_gene_cluster	saccharide	29.4	0.0427
Prevotella buccae_D17	BGC0000805	Moenomycin_biosynthetic_gene_cluster	other	31.59	0.0294
Fusobacterium_nucleatum_subsp_polymorphum_ATCC_10953	BGC0001351	Caluamycins_biosynthetic_gene_cluster	other	-86.35 < 0.0001	
Simonsiella_muelleri_ATCC_29453	BGC0000837	APE_VF_biosynthetic_gene_cluster	arylpolysene	-33.51	0.0209
Eggerthiaetenafornis_DSM_20559_strain_OT_569	BGC0000766	Exopolysaccharide_biosynthetic_gene_cluster	saccharide	-30.06	0.0382
Fusobacterium_nucleatum_subsp_animalis_F0419_strain_OT_420	BGC0000796	S-layer_glycan_biosynthetic_gene_cluster	saccharide	-30.72	0.0342
Fusobacterium_nucleatum_subsp_animalis_F0419_strain_OT_420	no_significant_hit	NA	other	-50.33	0.0005
Fusobacterium_sp_oral_taxon_370str_F0437	BGC0000053	Elaiophyllin_biosynthetic_gene_cluster	saccharide	-38.98	0.0072
Fusobacterium_nucleatum_subsp_vincentii_ATCC_49256	no_significant_hit	NA	other	-51.43	0.0004
Fusobacterium_sp_4_8	BGC0000796	S-layer_glycan_biosynthetic_gene_cluster	saccharide	-37.93	0.0089
Fusobacterium_nucleatum_subsp_animalis_F0419_strain_OT_420	BGC0000796	S-layer_glycan_biosynthetic_gene_cluster	saccharide	-43.73	0.0026
Gemella_morbilorum_M424	BGC0000717	Streptomycin_biosynthetic_gene_cluster	saccharide	-33.83	0.0197
Fusobacterium_periodonticum_1_1_41FAA	BGC0001058	Zorbamycin_biosynthetic_gene_cluster	other	-35.04	0.0157
Fusobacterium_periodonticum_1_1_41FAA	BGC0000053	Elaiophyllin_biosynthetic_gene_cluster	saccharide	-36.53	0.0118
Parvimonas_sp_oral_taxon_110_str_F0139	BGC0001418	Myxochromide_D_biosynthetic_gene_cluster	saccharide	-69.73 < 0.0001	
Parvimonas_micra_ATCC_33270	no_significant_hit	NA	saccharide	-68.58 < 0.0001	
Genella_haemolysans_ATCC_10379	BGC0000766	Exopolysaccharide_biosynthetic_gene_cluster	saccharide	-29.78	0.0401
Parvimonas_sp_oral_taxon_393_str_F0440	no_significant_hit	NA	saccharide	-41.85	0.0039
Fusobacterium_nucleatum_subsp_animalis_F0419_strain_OT_420	no_significant_hit	NA	putative	-30.24	0.0371
Parvimonas_micra_ATCC_33270	no_significant_hit	NA	putative	-58.17 < 0.0001	
Parvimonas_sp_oral_taxon_110_str_F0139	no_significant_hit	NA	saccharide	-34.37	0.0178
Parvimonas_sp_oral_taxon_393_str_F0440	no_significant_hit	NA	putative	-48.73	0.0008
Parvimonas_micra_ATCC_33270	no_significant_hit	NA	putative	-45.56	0.0017
Parvimonas_sp_oral_taxon_393_str_F0440	BGC0000344	Entolysin_biosynthetic_gene_cluster	putative	-72.31 < 0.0001	
Solobacterium_moorei_F0204	no_significant_hit	NA	putative	-37.21	0.0103

Table 2 Microbial virulence factor abundance in TT. Virulence factors found to be in higher abundance in TT when compared with contralateral NT samples. Predicted function, associated microbial taxa, and significance values are summarized below

Significantly abundant virulence factors in tumor tissue

Taxa match	Gene annotation	Individual <i>p</i> value
<i>Streptococcus pneumoniae</i> Hungary 19A-6	Choline binding protein CVF122	0.0291
<i>Treponema denticola</i> ATCC35405	Major outer sheath protein (Fibronectin binding protein AI239)	< 0.0001
<i>Mycoplasma penetrans</i> HF-2	p35 lipoprotein homolog	0.0005
<i>Acinetobacter baumannii</i> AB0057	AB57 0984, transcriptional regulator, LysR (Heme utilization)	< 0.0001
<i>Shigella dysenteriae</i> Sd197	ospC4 Mxi-Spa TTSS effector controlled by VirB CVF494	0.0169
<i>Bacillus cereus</i> ATCC 10987	O-antigen polymerase (wzy) Polysaccharide capsule CVF567	0.0013
<i>Haemophilus somnus</i> 2336	lic2A glycosyl transferase family 25 LOS CVF494	0.0004
<i>Streptococcus gordonii</i> str. Challis	csh B surface-associated protein Csh B Fibronectin binding protein AI195	0.0199
<i>Streptococcus gordonii</i> str. Challis	csh A surface-associated protein Csh B Fibronectin binding protein AI194	0.0050

Prevotella nigrescens has been implicated in contributing to periodontal disease [66–68]. Additionally, since our TT sequence data showed presence of *Porphyromonas gingivalis*, a known oral pathogen associated with periodontal disease and more recently OSCC [69], the known association of

P. intermedia with *P. gingivalis* [68] provides additional evidence that these pathogens may be contributing to OSCC pathogenesis as periodontal disease is often associated with OC and other cancers [70–72]. Expanding on this study to include deeper metagenomics sequencing to confirm the

Fig. 6 Clustered heatmap of Pearson-correlation coefficient for significantly enriched virulence and antibiotic resistance genes from *Fusobacterium*, representing non-tumor tissue and tumor tissue

metagenomes. Clustering based on the DESeq2 generated normalized abundances for these genes clearly illustrates their associations with cancer

presence of these species would certainly support our findings. Contrary to the increased abundance of specific genera known to contain oral pathogens associated within tumor tissues, we have also shown associations of several oral bacteria known to be found in healthy oral microbiomes. These genera including *Streptococcus*, *Veillonella*, and *Rothia* were observed in higher abundance in the NT samples when compared with TT samples. This finding is significant and further strengthens the assumption that OSCC pathology may be caused by oral pathogens.

The predominance and coexistence of the *Rothia* in the adjacent NT community, when compared with TT, were expected as *Rothia* species are part of the normal flora in the human oral cavity and pharynx [73, 74] and have also been isolated from the human gut microbiomes with moderate frequency [75, 76]. Additionally, *Rothia* is commonly found in homeostasis in oral communities of specific animal models including dogs [77]. The high abundance of *Streptococcus* in NT was also expected as this organism is also commonly seen in healthy oral microbiomes [28, 29]. Our results are consistent when compared with other studies that demonstrate that, in healthy humans, sequence analysis of oral communities showed the predominant taxa as *Corynebacterium*, *Rothia*, and *Actinomyces* [78] while the most prevalent species correlated with oxidative stress markers in an inter-individual specificity study were *Rothia* and *Streptococcus parasanguinis* [79].

Though our metagenomics sequencing was at shallow depth, we were able to detect several interesting findings. Our dataset showed TT community composition and suggested that there was coexistence in OSCC tissues, where *S. gordonii* and *Fusobacterium* genera were highly associated. Our metagenomics results showed that *S. gordonii* CshA and CshB fibronectin binding protein were significantly in higher abundance in TT; p values = 0.0050 and 0.0199, respectively, suggesting that association of *S. gordonii* to TT has some correlation to the development of OSCC. This association is significant since *S. gordonii* has been demonstrated as an early to mid-colonizer in oral biofilms [80–82] and may be suitable as a potential biomarker for diagnostics. Detecting increasing levels of *S. gordonii* fibronectin binding proteins over time in oral biopsy or deep swab samples may provide insight into a patient's predisposition to the development of OSCC and could potentially revolutionize diagnostics of this disease. Additional experiments utilizing metagenomics as outlined in our study in addition to transcriptomics analysis will confirm our predictions and further support our results.

As inflammatory pathogens such as *Fusobacterium* [83, 84], *Porphyromonas* [84, 85], and *Parvimonas* [86] are likely to adhere to known commensal bacteria and, more importantly, to host epithelial cells, we justified the use of tissue samples in our study. Sample collections in regard to screening patients to determine their predisposition to OSCC would

likely be improved if the collection of gum tissue biopsies or deep swabbing were utilized. We suspected that utilizing tissue samples over superficial swab samples as used in previous studies would provide a much more comprehensive description of community composition particularly invasive pathogens that may be deep in oral tissues. Two such invasive organisms, *Fusobacterium* and *Porphyromonas*, were frequently found in our data to be collectively in high abundance in TT. This is not entirely surprising since this pair of pathogens has also been isolated simultaneously in other periodontal studies in animal models [87] and human infections [86, 88]. These studies indicated enhanced pathogenicity of the mixed inocula in comparison with the individual confrontation. It is highly likely that frequency and natural coexistence of specific pathogens as demonstrated in our study may contribute to microbial synergism and virulence.

After evaluation of our sequencing data, we identified a number of potential biomarkers using various genomics approaches. However, since our metagenomics sequencing coverage was shallow due to host DNA contamination, we further evaluated the role of other microbial virulence factors in TT compared with the adjacent NT using predicted functional approaches to analyze the 16S sequences. Tax4fun analysis demonstrated predicted functional factors from microbial communities based on our 16S rRNA sequences, indicating that nucleotide metabolism, terpenoid and polyketide biosynthesis, cofactors and vitamins, and carbohydrate metabolism were enriched in TT, providing additional potential biomarkers to target to improve OSCC diagnostics. While most dynamics of virulence factors are still unknown in complex communities, it is probable that they are contributing to the survival of pathogenic bacteria associated with OSCC, immune evasion, inflammation, and disease progression.

Our results provide significant insight into OSCC pathology due to evidence and association of *Fusobacteria* with TT samples. *Fusobacteria* have long been associated with colon cancer [15, 36] and other types of cancers [37], and their pathogenic phenotype is classified as oncobacterium [89]. We found that significantly enriched virulence factors and antibiotic resistance genes from potentially oncogenic *Fusobacterium* are present in tumor metagenomes. Lipopolysaccharide biosynthesis, type V secretion proteins, outer membrane proteins, butyrate fermentation, and iron metabolism were further investigated. While butyrate production has been associated with suppression of inflammation and cancer progression, recent studies suggest that butyrate may contribute to cancer progression as pathogenic *Fusobacterium* utilizes a distinct amino acid metabolism pathway associated with the release of harmful by-products such as ammonia [90]. Clustering based on the DESeq2 generated normalized abundances for these genes illustrated their associations with OSCC metagenomes, especially outer membrane proteins

and iron metabolism (Fig. 6). This illustrates the presence, abundance, and metabolic functions of *Fusobacteria* in inflammatory processes and dysbiosis and agrees with other studies that have been performed to date on OSCC.

In addition to exploring bacterial metagenomes obtained from tissue samples, we have also validated our findings through saliva analysis. Krona plots produced radial space-filling charts displaying the mean relative abundances of bacterial taxa based on 16S gene sequencing, as well as bacterial and fungal taxa based on the metagenomic sequencing. Figure S1 displays taxonomic hierarchy with genus and species levels at the outermost circle, characterized using 16S sequencing and metagenomics. An interactive version of these charts is available in Supplementary Figure S1. Our saliva results appeared to resemble the non-tumor tissue, as evidenced by Krona and PCoA plots (Fig. 1; Figure S1). Beyond deep sequencing of DNA, we utilized other approaches such as mass spectrometry (MS)-based metaproteomics as these methods have emerged as a complementary tool to transcriptomic analysis, allowing for simultaneous measurement of proteins derived from the host and microbiome. Since we have previously identified several human salivary proteins that have high prediction accuracy as biomarkers for OSCC [14], we further characterized the metaproteomic composition of the same set of saliva samples by taking advantage of the robust STrap sampling approach [14]. This method utilizes MS with high resolution accurate mass (HR/AM) capacity and with our previous metaproteomics experience [91] to study the host and microbial composition of saliva derived from OSCC patients. Our data demonstrated the significantly increased abundance of *Prevotella* in OSCC saliva when compared with non-OSCC saliva, thus validating our sequencing data while in agreement with previous findings [92, 93]. Our proteomics results indicated that specific bacteria such as *Corynebacterium*, *Enterococcus*, and *Mogibacterium* were mostly associated with non-OSCC saliva samples and in significantly decreased abundance in the OSCC saliva samples. *Corynebacterium* and other medically relevant Gram-positive rods have been found to be present in healthy plaque samples in addition to other parts of the body including nasal tissues [94], skin, and gut [95]. Such differentiations were not observed from the sequence data of the tissue samples, which showed that saliva was more similar to NT than TT. Our results further suggested that saliva may not reflect the microbial community of TT and is unlikely to be a suitable target specimen for OSCC diagnostics when using sequencing only. Proteomics results, however, suggest that this approach may be critical for the improvement of diagnostics and would provide a sample target that is less invasive than using deep swabbing or gum tissue biopsies. These data highlighted the importance of sampling sublocation for microbiome-associated biomarker discovery [96]. Combining multi-omics approaches with multiple sample sources would better elucidate disease signatures. In the context of total microbial mass in saliva metaproteome, although greater numbers of

microbial proteins were identified in OCSS-derived saliva, as evidenced by the normalized spectrum counts, the statistical significance was low ($p > 0.05$) between OSCC and control subjects due to the small sample size. The key finding in our results is that the microbial diversity in OSCC saliva appears to have a higher abundance of microbial peptides when compared with non-OSCC saliva. We anticipate that analyzing a larger cohort would confirm our findings and reveal additional microbial markers.

Multi-omic approaches have been adapted to study the functional composition of the microbial community, each of which has strengths and weaknesses [97]. While our “genomic-proteomic” approach provided a comprehensive approach to generating information in regard to microbial virulence in TT, future mechanistic studies are needed to investigate host dynamics to validate the exact role of virulence factors in OSCC pathogenesis. Altogether, our results showed that virulence factors from *S. gordonii* and *Fusobacterium* species were found mostly associated with TT when compared with NT and saliva. Though our use of tissues may not be practical in the clinical diagnostics environment, it was necessary to achieve a comprehensive characterization of the microbial community in OSCC. Additionally, we determined that saliva samples may provide clinicians with a less invasive approach to determining a patient’s predisposition to development of OSCC. We were also able to identify specific signatures that can be further utilized in diagnostics to determine a patient’s susceptibility to OSCC through analysis of tissue biopsies or perhaps aggressive oral tissue swabs. Expanding our study to generating data from patients at each stage of OSCC development is critical to fully understand the oral microbiome’s role in OSCC pathology. Furthermore, understanding the metabolic function of inflammatory virulence genes has great potential in diagnostics and treatment of OSCC as it could potentially shed light on their functions related to the host immune system to clear microbial biofilms and respond to the dysbiotic microbiome causing the disease state and/or the exacerbated host’s inflammatory response causing tissue collateral damage.

Conflict of Interest The authors declare no conflict of interests.

Compliance with Ethical Standards

All authors declare that all methods in this study followed the protocol approved by the Institutional Review Board of Poznan University of Medical Sciences in Poznan, Poland, under IRB# 412/18. All experiments were performed in accordance with relevant guidelines and regulations. Informed consent for participation in the study was obtained from all patients included in the study.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. The Oral Cancer Foundation. The Oral Cancer Foundation. In: The Oral Cancer Foundation [Internet]. 2015 [cited 16 Nov 2015]. Available: <https://oralcancerfoundation.org/>
2. Bråmer GR (1988) International statistical classification of diseases and related health problems. Tenth revision. World Health Stat Q 41:32–36
3. Wamakulasuriya S (2009) Global epidemiology of oral and oropharyngeal cancer. Oral Oncol 45:309–316
4. Wang F, Zhang H, Xue Y, Wen J, Zhou J, Yang X, Wei J (2017) A systematic investigation of the association between HPV and the clinicopathological parameters and prognosis of oral and oropharyngeal squamous cell carcinomas. Cancer Med 6:910–917
5. Pushalkar S, Ji X, Li Y, Estilo C, Yegnanarayana R, Singh B, Li X, Saxena D (2012) Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. BMC Microbiol 12:144
6. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394–424
7. Markopoulos AK, Michailidou EZ, Tzimagiorgis G (2010) Salivary markers for oral cancer detection. Open Dent J 4:172–178
8. Ellison MD, Campbell BH (1999) Screening for cancers of the head and neck: addressing the problem. Surg Oncol Clin N Am 8:725–734 vii
9. Schantz SP (1993) Biologic markers, cellular differentiation, and metastatic head and neck cancer. Eur Arch Otorhinolaryngol 250:424–428
10. Website. [cited 26 Mar 2019]. Available: Noone AM, Howlander N, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER cancer statistics review, 1975–2015, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER web site, April 2018
11. Wong DT (2006) Salivary diagnostics for oral cancer. J Calif Dent Assoc 34:303–308
12. Al-hebshi NN, Li S, Nasher AT (2016). Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jjc.30068> Exome sequencing of oral squamous cell carcinoma in users of Arabian snuff reveals novel candidates for driver genes. J Cancer 139:363–372
13. Zhao H, Chu M, Huang Z, Yang X, Ran S, Hu B, Zhang C, Liang J (2017) Variations in oral microbiota associated with oral cancer. Sci Rep 7:11773
14. Lin Y-H, Eguez RV, Torralba MG, Singh H, Golusinski P, Golusinski W et al (2019) Self-assembled STRap for global proteomics and salivary biomarker discovery. J Proteome Res 18:1907–1915. <https://doi.org/10.1021/acs.jproteome.9b00037>
15. Yu J, Chen Y, Fu X, Zhou X, Peng Y, Shi L, Chen T, Wu Y (2016) Invasive *Fusobacterium nucleatum* may play a role in the carcinogenesis of proximal colon cancer through the serrated neoplasia pathway. Int J Cancer 139:1318–1326
16. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW et al (2012) The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS Biol 10:e1001377
17. Pelizzer T, Dias CP, Poeta J, Torriani T, Roncada C (2016) Colorectal cancer prevalence linked to human papillomavirus: a systematic review with meta-analysis. Rev Bras Epidemiol 19:791–802
18. Hoppe-Seyler K, Bossler F, Lohrey C, Bulkescher J, Rösl F, Jansen L, Mayer A, Vaupel P, Dürst M, Hoppe-Seyler F (2017) Induction of dormancy in hypoxic human papillomavirus-positive cancer cells. Proc Natl Acad Sci U S A 114:E990–E998
19. Meurman JH (2010) Oral microbiota and cancer. J Oral Microbiol 2. <https://doi.org/10.3402/jom.v2i0.5195>
20. Wauters GV, Ferrell L, Ostroff JW, Heyman MB (1990) Hyperplastic gastric polyps associated with persistent *Helicobacter pylori* infection and active gastritis. Am J Gastroenterol 85 Available: <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=00029270&AN=16089690&h=crR%2FRxmCHYExDWQOJkrsLw03FfdIPdFgtbW1Y4MisjYB2sYVka808910%2FSHQjmNHm%2BPubpNGzakRt1NXFbJevw%3D%3D&crl=c>
21. Loffield RJL, Willems I, Flendrig JA, Arends JW (1990) *Helicobacter pylori* and gastric carcinoma. Histopathology. 17:537–541
22. Parsonnet J, Vandersteen D, Goates J, Sibley RK, Pritikin J, Chang Y (1991) *Helicobacter pylori* infection in intestinal- and diffuse-type gastric adenocarcinomas. J Natl Cancer Inst 83:640–643
23. Franco EL, de Sanjosé S, Broker TR, Stanley MA, Chevarie-Davis M, Isidean SD, Schiffman M (2012) Human papillomavirus and cancer prevention: gaps in knowledge and prospects for research, policy, and advocacy. Vaccine. 30(Suppl 5):F175–F182
24. Tsimplaki E, Argyri E, Xesfyngi D, Daskalopoulou D, Stravopodis DJ, Panotopoulou E (2014) Prevalence and expression of human papillomavirus in 53 patients with oral tongue squamous cell carcinoma. Anticancer Res 34:1021–1025
25. Hooper SJ, Crean S-J, Fardy MJ, Lewis MAO, Spratt DA, Wade WG, Wilson MJ (2007) A molecular analysis of the bacteria present within oral squamous cell carcinoma. J Med Microbiol 56:1651–1659
26. Saxena S, Sankhla B, Sundaragiri KS, Bhargava A (2017) A review of salivary biomarker: a tool for early oral cancer diagnosis. Adv Biomed Res 6:90
27. Furquim CP, Soares GMS, Ribeiro LL, Azcarate-Peril MA, Butz N, Roach J, Moss K, Bonfim C, Torres-Pereira CC, Teles FRF (2017) The salivary microbiome and oral cancer risk: a pilot study in Fanconi anemia. J Dent Res 96:292–299
28. Yumoto H, Hirota K, Hirao K, Ninomiya M, Murakami K, Fujii H, Miyake Y (2019) The pathogenic factors from oral streptococci for systemic diseases. Int J Mol Sci 20. <https://doi.org/10.3390/ijms20184571>
29. Burton JP, Chilcott CN, Tagg JR (2005) The rationale and potential for the reduction of oral malodour using *Streptococcus salivarius* probiotics. Oral Dis 11(Suppl 1):29–31
30. Struzycka I (2014) The oral microbiome in dental caries. Pol J Microbiol 63:127–135
31. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent Jr RL (1998) Microbial complexes in subgingival plaque. J Clin Periodontol 25:134–144
32. Schmidt BL, Kuczynski J, Bhattacharya A, Huey B, Corby PM, Queiroz ELS, Nightingale K, Kerr AR, DeLacure MD, Veeramachaneni R, Olshen AB, Albertson DG, Muy-Teck Teh (2014) Changes in abundance of oral microbiota associated with oral cancer. PLoS One 9:e98741
33. Guerrero-Preston R, Godoy-Vitorino F, Jedlicka A, Rodríguez-Hilario A, González H, Bondy J, Lawson F, Folawiyo O, Michailidi C, Dziedzic A, Thangavel R, Hadar T, Noordhuis MG, Westra W, Koch W, Sidransky D (2016) 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. Oncotarget. 7:51320–51334

34. Wang H, Funchain P, Bebek G, Altemus J, Zhang H, Niazi F, Peterson C, Lee WT, Burkey BB, Eng C (2017) Microbiomic differences in tumor and paired-normal tissue in head and neck squamous cell carcinomas. *Genome Med* 9:14
35. Al-hebshi N, Al-haroni M, Skaug N (2006) In vitro antimicrobial and resistance-modifying activities of aqueous crude khat extracts against oral microorganisms. *Arch Oral Biol* 51:183–188
36. Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, Jelinek J, Yamano HO, Sugai T, An B, Shureiqi I, Toyota M, Kondo Y, Estecio MRH, Issa JPJ (2014) *Fusobacterium* in colonic flora and molecular features of colorectal carcinoma. *Cancer Res* 74:1311–1318
37. Yusuf E, Wybo I, Piérard D (2016) Case series of patients with *Fusobacterium nucleatum* bacteremia with emphasis on the presence of cancer. *Anaerobe* 39:1–3
38. Amer A, Galvin S, Healy CM, Moran GP (2017) The microbiome of potentially malignant oral leukoplakia exhibits enrichment for *Fusobacterium*, *Leptotrichia*, *Campylobacter*, and *Rothia* species. *Front Microbiol* 8:2391
39. Yost S, Stashenko P, Choi Y, Kukuruzinska M, Genco CA, Salama A, Weinberg EO, Kramer CD, Frias-Lopez J (2018) Increased virulence of the oral microbiome in oral squamous cell carcinoma revealed by metatranscriptome analyses. *Int J Oral Sci* 10:32
40. McInnes PAMC. Manual of Procedures for Human Microbiome Project Core Microbiome Sampling Protocol A #07-001. In: HMP Initiative 1: Core Microbiome Sampling Protocol A [Internet]. 2010 [cited 2016]. Available: <http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd002854.2>
41. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120
42. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methe B, DeSantis TZ, The Human Microbiome Consortium, Petrosino JF, Knight R, Birren BW (2011) Chimeric 16S rRNA sequence formation and detection in sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21:494–504
43. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Tumbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108(Suppl 1):4516–4522
44. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541
45. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35:7188–7196
46. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267
47. Oksanen FJ et al (2017) *vegan: Community Ecology Package*. R package version 2.4-3. - references - Scientific Research Publishing. [cited 15 Apr 2019]. Available: [https://www.scirp.org/\(S\(43dyn45teexjx455qlt3d2q\)\)/reference/ReferencesPapers.aspx?ReferenceID=2134091](https://www.scirp.org/(S(43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=2134091)
48. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336
49. Aßhauer KP, Wemheuer B, Daniel R, Meinicke P (2015) Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31:2882–2884
50. Kim D, Song L, Breitwieser FP, Salzberg SL (2016) Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res* 26:1721–1729
51. Aleti G, Baker JL, Tang X, Alvarez R, Dinis M, Tran NC et al (2018) Identification of the bacterial biosynthetic gene clusters of the oral microbiome illuminates the unexplored social language of bacteria during health and disease. *bioRxiv*:431510. <https://doi.org/10.1101/431510>
52. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]*. Available: <http://arxiv.org/abs/1303.3997>
53. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550
54. Yang J, Chen L, Sun L, Yu J, Jin Q (2008) VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. *Nucleic Acids Res* 36:D539–D542
55. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KE, Morley PS, Boucher C (2017) MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res* 45:D574–D580
56. Yu Y, Suh M-J, Sikorski P, Kwon K, Nelson KE, Pieper R (2014) Urine sample preparation in 96-well filter plates for quantitative clinical proteomics. *Anal Chem* 86:5470–5477
57. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989
58. Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. *Aust Ecol* 18:117–143
59. Manson McGuire A, Cochrane K, Griggs AD, Haas BJ, Abeel T, Zeng Q et al (2014) Evolution of invasion in a diverse set of *Fusobacterium* species. *MBio* 5:e01864
60. Rubinstein MR, Baik JE, Lagana SM, Han RP, Raab WJ, Sahoo D et al (2019) *Fusobacterium nucleatum* promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO Rep*. <https://doi.org/10.15252/embr.201847638>
61. Brennan CA, Garrett WS (2018) *Fusobacterium nucleatum*—symbiont, opportunist and oncobacterium. *Nat Rev Microbiol* 1
62. Graves DT, Corrêa JD, Silva TA (2019) The oral microbiota is modified by systemic diseases. *J Dent Res* 98:148–156
63. Eribe ERK, Olsen I (2017) *Leptotrichia* species in human infections II. *J Oral Microbiol* 9:1368848
64. Zhou Z, Chen J, Yao H, Hu H (2018) *Fusobacterium* and colorectal cancer. *Front Oncol* 8:371
65. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA, Holt RA (2012) *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res* 22:299–306
66. Fukui K, Kato N, Kato H, Watanabe K, Tatematsu N (1999) Incidence of *Prevotella intermedia* and *Prevotella nigrescens* carriage among family members with subclinical periodontal disease. *J Clin Microbiol* 37:3141–3145
67. Dorn BR, Leung KL, Progulske-Fox A (1998) Invasion of human oral epithelial cells by *Prevotella intermedia*. *Infect Immun* 66:6054–6057
68. Torkko H, Asikainen S (1993) Occurrence of *Porphyromonas gingivalis* with *Prevotella intermedia* in periodontal samples. *FEMS Immunol Med Microbiol* 6:195–198
69. Lafuente Ibáñez de Mendoza I, Maritxalar Mendia X, García dela Fuente AM, Quindós Andrés G, Aguirre Urizar JM (2020) Role of

- Porphyromonas gingivalis in oral squamous cell carcinoma development: a systematic review. *J Periodontol* 55:13–22
70. Michaud DS, Fu Z, Shi J, Chung M (2017) Periodontal disease, tooth loss, and cancer risk. *Epidemiol Rev* 39:49–58
 71. Güven DC, Dizdar Ö, Akman AC, Berker E, Yekedüz E, Ceylan F et al (2019) Evaluation of cancer risk in patients with periodontal diseases. *Turk J Med Sci* 49:826–831
 72. Shin YJ, Choung HW, Lee JH, Rhyu IC, Kim HD (2019) Association of periodontitis with oral cancer: a case-control study. *J Dent Res* 98:526–533
 73. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, Lakshmanan A, Wade WG (2010) The human oral microbiome. *J Bacteriol* 192:5002–5017
 74. Tsuzukibashi O, Uchibori S, Kobayashi T, Umezawa K, Mashimo C, Nambu T, Saito M, Hashizume-Takizawa T, Ochiai T (2017) Isolation and identification methods of Rothia species in oral cavities. *J Microbiol Methods* 134:21–26
 75. Pardi G, Acevedo AM, RM DI, Perrone M (1996) Studies over Rothia dentocariosa present in patients with and without dental caries. *J Dent Res*:1321
 76. Zamakhchhari M, Wei G, Dewhirst F, Lee J, Schuppan D, Oppenheim FG, Helmerhorst EJ (2011) Identification of Rothia bacteria as gluten-degrading natural colonizers of the upper gastro-intestinal tract. *PLoS One* 6:e24455
 77. Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, Buckley CMF, Davis IJ, Bennett ML, Marshall-Jones ZV (2012) The canine oral microbiome. *PLoS One* 7:e36067
 78. Zaura E, Keijser BJF, Huse SM, Crielaard W (2009) Defining the healthy “core microbiome” of oral microbial communities. *BMC Microbiol* 9:259
 79. Džunková M, Martinez-Martinez D, Gardlík R, Behuliak M, Janšáková K, Jiménez N, Vázquez-Castellanos JF, Martí JM, D’Auria G, Bandara HMHN, Latorre A, Celec P, Moya A (2018) Oxidative stress in the oral cavity is driven by individual-specific bacterial communities. *NPJ Biofilms Microbiomes* 4:29
 80. Periasamy S, Kolenbrander PE (2009) Mutualistic biofilm communities develop with Porphyromonas gingivalis and initial, early, and late colonizers of enamel. *J Bacteriol* 191:6804–6811
 81. Kim AR, Ahn KB, Kim HY, Seo HS, Yun C-H, Han SH (2016) Serine-rich repeat adhesin gordonii surface protein B is important for Streptococcus gordonii biofilm formation. *J Endod* 42:1767–1772
 82. Walker AM, Cannata J, Dowling MH, Ritchie B, Maloney JE (1978) Sympathetic and parasympathetic control of heart rate in unanaesthetized fetal and newborn lambs. *Biol Neonate* 33:135–143
 83. He X, Hu W, Kaplan CW, Guo L, Shi W, Lux R (2012) Adherence to streptococci facilitates Fusobacterium nucleatum integration into an oral microbial community. *Microb Ecol* 63:532–542
 84. Kaplan CW, Lux R, Haake SK, Shi W (2009) The Fusobacterium nucleatum outer membrane protein RadD is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. *Mol Microbiol* 71:35–47
 85. Zheng C, Wu J, Xie H (2011) Differential expression and adherence of Porphyromonas gingivalis FimA genotypes. *Mol Oral Microbiol* 26:388–395
 86. Murphy EC, Frick I-M (2013) Gram-positive anaerobic cocci—commensals and opportunistic pathogens. *FEMS Microbiol Rev* 37:520–553
 87. Ebersole JL, Kesavaln L, Schneider SL, Machen RL, Holt SC (2008) Comparative virulence of periodontopathogens in a mouse abscess model. *Oral Dis*:115–128. <https://doi.org/10.1111/j.1601-0825.1995.tb00174.x>
 88. Noguchi N, Noiri Y, Narimatsu M, Ebisu S (2005) Identification and localization of extraradicular biofilm-forming bacteria associated with refractory endodontic pathogens. *Appl Environ Microbiol*:8738–8743. <https://doi.org/10.1128/aem.71.12.8738-8743.2005>
 89. Brennan CA, Garrett WS. Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. - PubMed - NCBI. [cited 26 Mar 2019]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/30546113>
 90. Anand S, Kaur H, Mande SS (2016) Comparative analysis of butyrate production pathways in gut commensals and pathogens. *Front Microbiol* 7:1945
 91. Yu Y, Sikorski P, Smith M, Bowman-Gholston C, Cacciabeve N, Nelson KE, Pieper R (2017) Comprehensive metaproteomic analyses of urine in the presence and absence of neutrophil-associated inflammation in the urinary tract. *Theranostics*. 7:238–252
 92. Hsiao J-R, Chang C-C, Lee W-T, Huang C-C, Ou C-Y, Tsai S-T, Chen KC, Huang JS, Wong TY, Lai YH, Wu YH, Hsueh WT, Wu SY, Yen CJ, Chang JY, Lin CL, Weng YL, Yang HC, Chen YS, Chang JS (2018) The interplay between oral microbiome, lifestyle factors and genetic polymorphisms in the risk of oral squamous cell carcinoma. *Carcinogenesis*. 39:778–787
 93. Perera M, Al-Hebshi NN, Perera I, Ipe D, Ulett GC, Speicher DJ et al (2018) Inflammatory bacteriome and oral squamous cell carcinoma. *J Dent Res* 97:725–732
 94. Ramsey MM, Freire MO, Gabriliska RA, Rumbaugh KP, Lemon KP (2016) Staphylococcus aureus shifts toward commensalism in response to Corynebacterium species. *Front Microbiol* 7:1230
 95. Cogen AL, Nizet V, Gallo RL (2008) Skin microbiota: a source of disease or defence? *Br J Dermatol* 158:442–455
 96. Starr AE, Deeke SA, Li L, Zhang X, Daoud R, Ryan J, Ning Z, Cheng K, Nguyen LVH, Abou-Samra E, Lavallée-Adam M, Figeys D (2018) Proteomic and metaproteomic approaches to understand host-microbe interactions. *Anal Chem* 90:86–109
 97. Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, Huttenhower C (2015) Sequencing and beyond: integrating molecular “omics” for microbial community profiling. *Nat Rev Microbiol* 13:360–372