Distinct Microbial Communities Colonize Tonsillar Squamous Cell Carcinoma

Burkhard Ludewig (✉ Burkhard.Ludewig@kssg.ch)

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen   https://orcid.org/0000-0002-7685-573X

Angelina De Martin

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Mechthild Lütge

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Yves Stanossek

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Céline Engetschwiler

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Jovana Cupovic

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Kirsty Brown

  University of Calgary

Izadora Demmer

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Martina A. Broglie

  UniversitätsSpital Zürich: UniversitatsSpital Zurich

Kathy McCoy

  University of Calgary

Markus Geuking

  University of Calgary

Wolfram Jochum

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Sandro J. Stoeckli

  Kantonsspital St Gallen: Kantonsspital Sankt Gallen

Research

Abstract

Background: Squamous cell carcinoma of the tonsil is one of the most frequent cancers of the oropharynx. The escalating rate of tonsil cancer during the last decades is associated with the increase of high risk-human papilloma virus (HR-HPV) infections. While the microbiome in oropharyngeal malignant diseases has been characterized to some extent, the microbial colonization of HR-HPV-associated tonsil cancer remains largely unknown.

Results: Using 16S rRNA gene amplicon amplicon sequencing, we have characterized the microbiome of human palatine tonsil crypts in patients suffering from HR-HPV-associated tonsil cancer in comparison to an age-matched control cohort of sleep apnea patients. We found an increased abundance of the phyla *Firmicutes* and *Actinobacteria* in tumor patients, whereas the abundance of *Spirochaetes* and *Synergistetes* was significantly higher in the control cohort. Furthermore, the accumulation of several genera such as *Veillonella, Streptococcus* and *Prevotella_7* in tonsillar crypts was associated with tonsil cancer. In contrast, *Fusobacterium, Prevotella* and *Treponema_2* were enriched in sleep apnea patients. Machine learning-based bacterial species analysis indicated that a particular bacterial composition in tonsillar crypts is tumor-predictive. Species-specific PCR-based validation in extended patient cohorts confirmed that differential abundance of *Filifactor alocis* and *Prevotella melaninogenica* is a distinct trait of tonsil cancer.

Conclusion: This study shows that tonsil cancer patients harbor a characteristic microbiome in the crypt environment that differs from the microbiome of sleep apnea patients on all phylogenetic levels. Moreover, our analysis indicates that profiling of microbial communities in distinct tonsillar niches provides microbiome-based avenues for the diagnosis of tonsil cancer.

Introduction

Oropharyngeal squamous cell carcinoma is one of the most prevalent tumor types within the group of head-and-neck cancers [1]. The stark rise in the incidence of high risk-human papilloma virus (HR-HPV) infections has led to a substantial increase of tonsil cancer making the tonsillar environment the most common site of oropharyngeal cancer [2, 3]. HPV preferentially infects keratinocyte progenitors located in the basal layer of the tonsillar crypt epithelium [4, 5]. While most infected individuals clear the virus within 18 months [6], viral persistence and alterations in viral gene expression with enhanced production of the oncogenes E6 and E7, initiate HPV-mediated carcinogenesis [5]. Moreover, host resistance factors, cellular composition and the general milieu of particular epithelial niches determine the outcome of HPV infection [1, 4, 7]. It is thus important to determine whether specific microenvironmental conditions in tonsillar crypts are associated with tonsil cancer.

All mucosal surfaces of the human body are colonized by a plethora of microbial organisms. Bacterial communities in the oral and oropharyngeal regions are particularly complex and contain, in addition to the gastrointestinal tract, one of the most diverse microbiomes [8, 9]. Oropharyngeal bacterial

communities are dominated by the six major phyla *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Actinobacteria*, *Spirochaetes*, and *Fusobacteria* representing 96% of all taxa found in the oral cavity including the oropharynx [8, 10]. Moreover, different oral structures and tissues such as the hard and soft palates, sub- and supragingival surfaces, teeth, lips, cheeks and tonsils are colonized by distinct and dynamically changing microbial communities [10–12]. Hence, assessment of niche-specific differences of the oral and oropharyngeal microflora is crucial for understanding the role of bacteria in oropharyngeal diseases such as tonsil cancer.

The microbiota in human tonsils show a distinct spatial distribution with enrichment of particular bacterial species in the lymphoid tissue, whereas other species are found mainly in layers adherent to the epithelium [13]. However, biased approaches using 16S rRNA gene amplicon probes for specific bacterial species [13] may introduce a technical bias leading to over-representation of particular bacterial species. In contrast, culture-independent molecular profiling based on 16S rRNA gene amplicon amplicon sequencing revealed only minor differences in the bacterial composition when different sampling sites in human tonsils and adenoids were compared [14]. Noteworthy, the majority of studies elaborating on the composition of the tonsillar microbiome have analyzed tissues from pediatric patients [14–21]. Tonsil cancer occurs only in adults and the few available datasets on the adult microbiome in tonsil cancer have not considered the inclusion of age-matched control cohorts [22, 23]. Here, we have analyzed the tonsillar crypt microbiome from two different patient cohorts. We have enrolled patients with HR-HPV-positive squamous cell carcinoma of the tonsil and have established an age-matched control cohort of patients with obstructive sleep apnea (OSA). Sampling from different tonsillar compartments (epithelial surface, crypts and lymphoid tissue parenchyma) in OSA patients revealed a homogenous microbial composition in tonsillar niches across the oropharynx within individual patients. The microbiome composition in tonsillar crypts of tonsil cancer patients was significantly different when compared to OSA patients. Random forest-based classification of the crypt microbiome data and species-specific PCR-based validation of extended cohorts revealed that certain bacterial species combinations are tumor-predictive. In sum, the data presented here suggest that the analysis of bacterial communities in tonsillar crypts and the elaboration of tumor-preditive bacterial species pattern bear the potential to aid the detection of tonsil cancer.

Results

Tonsillar microbiome diversity in healthy individuals

As a first step for the assessment of microbial alterations in tonsil cancer, we have analyzed the tonsillar microbiome of a well-defined control cohort of patients that underwent tonsillectomy due to OSA. In total, 21 OSA patients have been enrolled between May 2017 and May 2020. 16S rRNA gene amplicon sequencing has been performed on the samples of the first 14 recruited patients (Table 1), while recruitment continued to reach the final cohort size (Supplementary Table 1). To assess whether the crypt microbiome represents the global tonsillar bacterial composition, we acquired two punch biopsies each from the surface epithelium, the lymphoid tissue parenchyma and the crypts (Fig. 1A). Data from the

same patient and sampling site were regarded as technical replicates and were merged in downstream bioinformatics analyses. In total, 160 biopsies have been analyzed in the OSA cohort with 130 samples remaining after quality control. We found a uniform within-sample (α-)diversity across different collection sites with the mean number of observed species and Chao1 species richness estimators showing no significant differences between the three sampling sites (Fig. 1B). Likewise, both Shannon and Simpson indices, which denote species richness and evenness, were not significantly different (Fig. 1C). Principle coordinate analysis using Bray Curtis (Fig. 1D), UNIFRAC (Fig. 1E) and weighted UNIFRAC (Fig. 1F) distance matrices did not reveal significant differences in β-diversity between the tonsillar compartments. *Bacteroidetes*, *Fusobacteria*, and *Firmicutes* were the dominant phyla in the three different anatomical locations (Fig. 1G). The abundance of the various phyla did not vary significantly between the sampling sites suggesting that microbial colonization occurs in a homogenous fashion across the different tonsillar niches.

Table 1

Characteristics of patients with 16S rRNA gene amplicon-based analysis of the tonsillar microbiome in the tonsillar carcinoma and obstructive sleep apnea cohorts.

Cohort	Main diagnosis[a]	N	Age (mean ± SD)	Male/female ratio	HPV p16 positivity	HR-HPV DNA positivity
1	Obstructive sleep apnea	14	38.9 ± 11.5	13:1	0/14	0/14
2	SCC of the tonsil	18	66.9 ± 12.8	2.6:1	16/18	14/16[b]

[a] Main clinical diagnosis as documented in the patient information system; SCC, squamous cell carcinoma.

[b] Data of two patients could not be obtained.

To further assess to what extent microbial community composition in tonsillar niches of individual patients is conserved, we compared the microbiome in ipsi- and contralateral tonsils of OSA patients (Fig. 2A). Wilcoxon signed-rank testing did not indicate significant differences in species richness (Fig. 2B) or richness and evenness (Fig. 2C) in the microbiome of opposing tonsils. Analysis of β-diversity revealed a significant dissimilarity for paired Permanova testing based on the Bray Curtis distance (p = 0.0275) (Fig. 2D), but not for UNIFRAC (Fig. 2E) or weighted UNIFRAC distances (Fig. 2F). The assessment of individual patients in these analyses highlights the substantially higher variance between individual patients compared to the differences between ipsi- and contralateral tonsils (Fig. 2D-F). In sum, these data indicate that tonsils of OSA patients harbor diverse bacterial communities and that the microbial composition in tonsillar niches of individual patients is similar across the oropharynx.

Tonsillar crypt microbiome composition in tonsil cancer

Tissue biopsies of 18 patients suffering from squamous cell carcinoma of the tonsils have been analyzed by 16S rRNA gene amplicon sequencing with almost 90% of the patients being HR-HPV-positive

(Table 1). In addition to the HPV-status, age, sex, TNM tumor staging, nicotine and alcohol exposure and secondary carcinomas were documented (Supplementary Table 2). Two punch biopsies from the contralateral crypt were collected during the surgery and were included in the analysis (Fig. 3A). The assessment of four α-diversity metrics did not reveal statistically significant differences between the crypt microbiome of tumor-affected and contralateral tonsils (Fig. 3B and C). Principle coordinate analysis using Bray Curtis, UNIFRAC and weighted UNIFRAC estimators did not indicate statistically significant dissimilarities in the microbial community composition of tumor-affected vs. contralateral tonsils (Fig. 3D-F). Extended analyses indicated that the severity of the disease, here assessed as pathological tumor stage, had a significant effect on the β-diversity of the microbial communities in the tumor-affected tonsils (Supplementary Fig. 1A-C). At phylum level, tumor-affected tonsils exhibited elevated (p = 0.056) relative abundance of *Firmicutes* (Fig. 3G). The genera *Fusobacterim*, *Prevotella*, *Prevoella_7* and *Veillonella* displayed the highest relative abundance in tonsillar crypts of tumor patients (Fig. 3H). These data suggest that the growth of HR-HPV-associated tonsil cancer alters the microenvironment in tonsillar crypts leading to changes in microbial communities both on phylum and genus level.

Distinct tonsillar crypt microbiome composition in tonsil cancer

Next, we compared the crypt microbiome of ipsi- and contralateral tonsils of tumor patient with the crypt microbiome of OSA patients. α-diversity measures revealed equal richness (Fig. 4A) and richness and evenness (Fig. 4B) in the microbial communities. In contrast, principle coordinate analysis revealed a significant shift of the microbiome composition in tumor patients compared to OSA patients for all distance metrics (Fig. 4C-E and Supplementary Fig. 2A-C). In particular, we found a substantial dissimilarity in the microbial composition on the phylum level with significantly elevated abundance of *Firmicutes* and *Actinobacteria* and significantly reduced abundance of *Spirochaetes*, *Synergistetes* and *Fusobacteria* in tumor patients (Fig. 4F and Supplementary Fig. 2D). These results were further corroborated using projection of taxa abundance onto the phylogenetic tree. As shown in Supplementary Fig. 3, consistent patterns emerged along tree branches with significant differences between tumor and OSA patients and substantially higher abundance of the phylum *Firmicutes* and its genera *Veillonella*, *Streptococcus* and *Megasphaer*a in tumor patients. The microbial composition analysis on the genus level revealed differential abundance of several genera in tonsillar crypts of tumor vs. OSA patients (Fig. 4G and Supplementary Fig. 4). *Fusobacterium* was the most abundant in OSA tonsils, whereas *Veillonella* and *Prevotella_7* were highly abundant in OSA tonsils (Fig. 4G). The crypt microbiome of contralateral tonsils of tumor patients consistently showed an intermediate relative abundance on the genus level (Fig. 4G and Supplementary Fig. 4A). The observed shift in relative genera abundance was coherent across patients in both cohorts (Supplementary Fig. 4B). Together, these data unveil a distinct tonsillar crypt microbiome composition in patients suffering from tonsil cancer.

Identification of tonsil cancer patients based on predictive microbiome pattern

To further elaborate a potential relationship between the microbiota and tonsil cancer, we compared the crypt microbiome of tumor-affected and OSA tonsils at the species level using highly resolved taxonomic classification according to the human oral microbiome database [24]. Since bacterial operational taxonomic units typically do not follow a Gaussian distribution, differences on the species level were assessed using the DESeq2 method [25], which permits the analysis of count data based on the negative binomial distribution. The heatmap shown in Fig. 5A lists the top 45 differentially abundant bacterial species in a pattern that clearly separates tumor from OSA patients. To assess whether the abundance of certain bacterial species in tumor tonsils could be used as a predictor for the presence of tonsil tumor, we trained a machine-learning algorithm using the available data from OSA and tonsil cancer patients. The applied random forest classification method was shown previously to outperform other supervised classifiers on microbiome data [26]. Based on the complete 16S rRNA gene amplicon datasets from tumor and OSA patients, the trained model was able to detect a tumor-affected patient with high accuracy (0.89), sensitivity (0.89) and specificity (0.88) (Fig. 5B). The top ten tumor-predictive species included *Treponema denticola*, *Fusobacterium periodonticum*, *Filifactor alocis*, *Fusobacterium nucleatum subsp. vincentii*, *Megasphaera micronuciformis*, *Prevotella melaninogenica* and *Veillonella atypica* (Fig. 5B), which showed all significantly different abundances in tumor vs. OSA patients (Fig. 5C). To validate these results, we performed quantitative PCR analysis using biopsy DNA and published primers that allow for the reliable detection of bacterial species. For the ten potentially predictive bacterial species, the published 16S rRNA gene amplicon PCR primers for *F. alocis* [27] and *P. melaninogenica* [28] produced the most robust abundance measures. Using crypt DNA samples from the extended cohorts (21 OSA patients and 28 tonsil tumor patients, Supplementary Tables 1 and 2), we found a significantly higher relative abundance of *F. alocis* in crypt biopsies of OSA compared to tumor patients (Fig. 5D). As expected from the random forest analysis, *P. melaninogenica* could be detected by PCR in all samples with significantly higher relative abundance in tumor-affected tonsils compared to OSA tonsils (Fig. 5E). Collectively, these data reveal that assessment of the composition of microbial communities in tonsillar crypts can serve as diagnostic means for the detection of tonsil cancer.

Discussion

According to the recent WHO classification, tonsillar squamous cell carcinomas are now subdivided into distinct morphological groups with HPV-positive cancer arising from the tonsillar crypts and exhibiting a non-keratinising morphology [29]. The tonsil cancer patients recruited in this study suffered from squamous cell carcinoma and the majority of the patients showed infection with the high risk strain HPV16, which is consistent with the global pattern [30]. Since the oral cavity and the oropharynx provide heterogeneous niches for bacterial colonization, we have focused on the microbiome in tonsillar crypts as the most critical region for HPV-induced carcinogenesis. Since the tonsillar microbiome in the tonsil tumor patients was significantly different from the microbiome of age-matched sleep apnea patients on

all phylogenetic levels, we conclude that HPV-induced carcinogenesis profoundly alters the conditions for bacterial growth in tonsillar crypts. It appears that the dysbiosis in tumor tonsils diffused to the contralateral, non-affected tonsil. Our data thus suggest that bacterial communities change and adapt to altered conditions in the tonsillar environment due to the tumor growth and that these changes exert a dominant effect on the oropharyngeal microbiome in remote locations.

Localized chronic inflammation is thought to be one of the major drivers of carcinogenesis [31]. The Gram-negative and invasive bacteria *Fusobacterium (F.) nucleatum* and *Porphyromonas (P.) gingivalis* are frequently discussed to be causally involved in the induction of squamous cell carcinoma in the oropharynx [32, 33]. *F. nucleatum* and *P. gingivalis* have been shown to promote tumor progression in an oropharyngeal squamous cell carcinoma mouse model [34] and *F. nucleatum subspecies polymorphum* has been found to be significantly increased in oral squamous cell carcinoma biopsies [35]. Moreover, an elevated abundance in oral wash and surface swabs of oral squamous cell carcinoma has been described for the genus *Fusobacterium* [36, 37]. However, in our study with the comparison of HPV-associated tonsillar squamous cell carcinoma and age-matched sleep apnea patient, the abundance of *Fusobacterium* in crypts of tumor tonsils was reduced. Moreover, at the species level, we found that an increased abundance of *F. periodonticum* and a decreased abundance of *F. nucleatum subspecies vincentii* was part of the tumor-predictive microbiome signature according to the random forest analysis. It is conceivable that these partially contradictory results are due to the distinct bacterial colonization pattern in the tonsillar crypt microenvironment in a diverse array of oral and oropharyngeal cancers.

The particular structure of tonsillar crypts is determined by the reticulated lymphoepithelial layer that separates the oropharynx from the underlying lymphoid tissue [38]. It is most likely that the reticulated epithelium that is interspersed with lymphocytes and myeloid cells, favors virus-induced malignant transformation due to direct access of the virus to epithelial stem cells. Furthermore, it is possible that particular bacteria foster the access of HPV particles to the basal epithelial cell layer. Indeed, *Streptococcus*-derived furin like endopeptidases have been shown to promote infection of epithelial cells by HPV16 pseudovirus suggesting that carcinogenesis could be enhanced by particular bacterial products [39]. It will be important to assess in future studies whether the bacterial species that have been identified here can alter the crypt microenvironment and thereby change HR-HPV infectivity and carcinoma induction.

It is most likely that the composition of bacterial communities in the tonsillar crypt not only affects tumor growth, but that once the carcinogenesis is progressing, the altered microenvironmental conditions in and around the tumor affect the local microbiome. Indeed, in this study tonsillar crypts of tumor-affected and contralateral, non-affected tonsils of cancer patients harbored highly similar microbial communities. These results are in line with the study of Wang et al. [23] who found more similarities than differences between the overall microbiomes from frozen tumor and adjacent normal tissues in a heterogeneous patient cohort on head and neck squamous cell carcinoma. Moreover, it appears that bacterial communities change during oral cancer progression in general [36] and in tonsil cancer as described here. Some overarching trends in oral and tonsil cancer-assocated microbiome alterations emerge: elevated

Rothia and lowered *Fusobacterium* abundance [40], elevated *Streptococcus* and *Veillonella* [41], reduced *Porphyromas* [36, 42]. Clearly, tumor-associated and -predictive changes in the microbial composition need to be detailed and validated on the species level to facilitate further delineation of pathogenic principles and the elaboration of diagnostic procedures.

Current conclusions about the association of the microbiome with oralpharyngeal cancer have been mainly drawn from studies using saliva or oral rinse microbiomes [41, 43, 44]. Using tonsillar crypt biopsies, our study has identified a set of bacterial species that appears to be tonsil cancer-predictive. The elevated abundance of *P. melaninogenica* in tonsil cancer is consistent with the heightened presence of this bacterial species in saliva of oral cancer patients [45]. The latter study of Mager et al. [45] has used DNA hybridization to detect 40 distinct bacterial species in saliva specimen. A later study based on 16S rRNA gene amplicon sequencing has found a higher abundance of *F. nucleatum subspecies polymorphum* and *Porphyromonas aeruginosa* in archived frozen tissue of oral squamous cell carcinoma patients, whereas bacterial species such as *Streptococcus mitis*, *Rothia mucilaginosa* and *Haemophilus parainfluenzae* were more frequent in epithelial swabs of healthy controls [35]. These previous studies together with the data presented here indicate that the assessment of bacterial dysbiosis with the quantification of differentially abundant bacterial "indicator" species could guide the diagnosis of oropharyngeal malignancies. Clearly, future studies will be required to further validate the set of bacterial species for diagnosis of tonsil cancer elaborated here and to define niche-specific microbiome alterations in other oropharyngeal cancer entities.

Conclusions

The incidence of HR-HPV-associated tonsil cancer has been steadily increasing in the recent decades. The comparison of the tonsillar crypt microbiome of patients suffering from tonsil cancer with an age-matched OSA patient cohort shows that the microbial ecosystem in the crypt environment changes during cancer development. Characteristic changes in the microbial composition could be detected on all phylogenetic levels. Machine learning-based analysis of the 16S rRNA gene amplicon sequencing data and PCR-based validation indicated that particular bacterial species combinations in the tonsillar crypt microenvironment can aid the diagnosis of tonsil cancer.

Methods

Study population and sample collection

The study protocol has been reviewed and approved by the Ethikkommission Ostschweiz (EKOS), permission number 2017-00051. All study participants provided written informed consent in accordance with the Declaration of Helsinki and the International Conference on Harmonization Guidelines for Good Clinical Practice. All regulations were followed according to the Swiss authorities and according to the clinical protocols. The sample size calculation for the study was not performed because of the exploratory nature of the study design.

Patients suffering from obstructive sleep apnea (OSA) were enrolled in control cohort 1 with a total of 21 participants that underwent tonsillectomy between May 2017 and May 2020. Exclusion criteria were as follows: (a) acute or chronic tonsillitis, (b) any disease with immunosuppression, (c) immunosuppressive treatment and (d) intake of antibiotics up to 4 weeks before surgery. After excision, tonsils were placed on a sterile surface and 4 mm punch biopsies (Stiefel® Biopsy punch) were taken using a sterile single-use stainless steel instrument immediately in the operation theater using separate sterile pairs of scissors and forceps to avoid cross-contamination between samples. Two biopsies were punched out of the tonsil surface epithelium and two biopsies were taken from tonsillar crypts. Next, tonsils were cut in half with a sterile surgical blade and two lymphoid tissue samples were harbested by punch biopsy.

Patients diagnosed with a primary, untreated and histologically confirmed squamous cell carcinoma of the tonsils (n = 28) were included into the study cohort 2 between August 2017 and February 2020. Patients with previous radiotherapy or surgery due to oropharyngeal cancer were excluded. Two punch biopsies from tonsillar crypts were taken directly after removal of the tumor-affected tonsil and two crypt biopsies were acquired from the contralateral tonsil during surgery. In cases when tumor tonsil were not resected, crypt biopsies were acquired in situ by punch biopsy as for of the contralateral tonsil. Each biopsy was acquired using a fresh sterile biopsy punch and a separate sterile pairs of scissors and forceps. Tissue biopsies were placed in sterile tubes on ice and stored at -80°C within 30 min until further analysis.

DNA extraction and 16S rRNA gene amplicon sequencing

Bacterial genomic DNA was extracted and purified using QIAamp DNA Stool Mini Kit (Qiagen) with minor modifications. Lysing Matrix D tubes (MPBio) were used for tissue disruption and an additional extraction step was introduced after incubation of bacteria with a lysis buffer (20 mg/ml lysozyme – Sigma 62970-5G-F, 2 mM EDTA, 1.2% Triton, 20 mM Tris-HCL) for an optimal yield of bacterial DNA. PCR amplification of the V4 region of bacterial 16S rRNA gene amplicon was performed with barcoded primers following a dual-index paired-end sequencing approach [46]. PCR was performed using KAPA HiFi polymerase (Roche) under the following cycling conditions: initial denaturation at 98°C (2 min), 25 – 40 cycles of 98°C (30 sec), 55°C (30 sec), 72°C (20 sec) and final elongation at 72°C (7 min). PCR products including a negative control were electrophoresed on a 1% agarose gel to ensure high-fidelity amplification. NucleoMag® NGS (Macherey-Nagel) was used for PCR clean up and size selection. Individual libraries were normalized with SequalPrep™ Normalization Plate Kit (ThermoFisher) according to the manufacturer's instruction and combined into a pooled library. The combined library was qualitatively and quantitatively assessed with a High Sensitivity D1000 ScreenTape device (Agilent) and on a Qubit fluorometer (ThermoFisher). 16S rRNA V4 amplicon sequencing was performed on the Illumina MiSeq v2 platform. Miseq data was converted to fastq format and demultiplexed using bcl2fastq software (Illumina).

Read processing and taxonomic analysis

Cutadapt was used for quality trimming and removal of the primer and adaptor sequences [47]. Further processing and analysis was run in R version 3.6.1. Trimmed reads were filtered and truncated based on quality scores and read length using the R/bioconductor package dada2 v1.14.0 [48] with 230 bp as length for forward reads and 210 bp for reverse reads. After dereplication error rates were estimated based on a parametric error model as implemented in the dada2 package and sequence variants were generated using the 'dada' function. Chimeric reads were removed using the 'removeBimeraDenovo' function. Finally, taxonomy was assigned based on the RDP trainset 16/release 11.5 [49] and the HOMD v.15.1 [24] database by applying the ,assignTaxonomy' function from the dada2 package that utilizes the RDP Naive Bayesian Classifier described by Wang et al [50], with kmer size = 8, and 100 bootstrap replicates. In order to further infer phylogenetic relationships R/bioconductor packages msa v.1.18.0 [51] and phangorn v.2.5.5 [52] were used to build multiple sequence alignments and construct a phylogenetic tree. Before downstream analysis samples were rarefied to an even sampling depth of 500 reads using the R/bioconductor package phyloseq v.1.32.0 [53]. Samples with less than 500 reads were removed from further analysis resulting in a final dataset of 192 samples. The observed low read counts compared to typical feces samples were expected, as the DNA was isolated from tissue biopsies.

Statistical analysis

Rarefied samples were analyzed for alpha diversity based on shannon, simpson and inverse simpson metrices and beta diversity based on Bray-Curtis, UNIFRAC, and Weighted UNIFRAC distances. All diversity measures were computed running the 'diversity' function from the R/Bioconductor package vegan v.2.5-7 [54] and Kruskal-Wallis and Wilcoxon tests were used to test for significant differences in alpha diversity measures. Beta diversity was visualized by principal coordinate analysis and permutational ANOVA was run using the 'adonis' function from R/bioconductor package vegan v.2.5-7 [54] to test for significant differences between patient and sample groups. Samples that were collected from the same patient and location were regarded as technical replicates and all statistical tests were run on mean values from technical replicates. Differences in relative abundances between groups on phylum, family and genus level were tested using Kruskal-Wallis and Wilcoxon-Mann-Whitney tests to compute statistics. In order to further examine the hierarchical relationship of taxa with differences in their relative abundance, a tree-based visualization technique was used by running the 'heat_tree' function from the metacoder R packages v.0.3.3 [55]. Finally, differentially abundant taxa between groups were identified utilizing the Bioconductor packages phyloseq v.1.32.0 [53] and DESeq2 v.1.28.1 [25] following standard protocols for differential abundance analysis. Statistical analyses of quantitative PCR data were performed with Prism 8.4.3 (GraphPad). Graphs depict mean ± s.d. and differences between two groups were evaluated using Wilcoxon-Mann-Whitney tests. Results were considered statistically significant with $p < 0.05$.

Random Forest analysis

Random Forest analysis was performed on species abundance data from control and tumor crypt samples using the randomForest R package v.4.6-14 [56]. Data pre-processing included removal of rare OTUs based on mean relative abundances followed by re-normalization and scaling of the remaining 405 OTUs. The random forest model was fitted by running the 'randomForest' function with ntree = 5001 and default values for mtry. Model significance was accessed by running the 'rf.significance' function from the rfUtilities R package v.2.1-5 [57] with 1,000 permutations. Model performance was tested by running a leave-one-out cross-validation utilizing the 'train' function from the caret R package v.6.0-86 [58] with identical mtry and ntree parameters as before. Finally, the mean decrease in accuracy (MDA) was computed as a measure for the importance of each taxa for the overall model and used to rank the taxa.

Quantitative PCR

Extracted DNA was used as a template for bacterial qPCR amplification and quantification using a real-time PCR machine (Quant Studio 5) and SYBR Green PCR technology (Applied Biosystems, Thermo Fisher Scientific). A pair of ubiquitous bacterial 16S rDNA-directed primers (forward 5'-ACTCCTACGGGAGGCAGCAGT-3' and reverse 5'-ATTACCGCGGCTGCTGGC-3') was used to assess the total amount of bacteria in the samples [59]. Cycling conditions included a denaturation step at 95°C for 10 min followed by 30 cyles of 95°C (10 sec), 63°C (20 sec) and 72°C (20 sec). Using the CT-values of eubacteria abundances of *F.alocis* and *P.melaninogenica* were calculated. *F.alocis* was quantified using species specific primers targeting the 16S rRNA gene amplicon that have been established and validated [27]. 16S rRNA gene amplicon-directed forward 5'-CAG GTG GTT TAA CAA GTT AGT GG-3' and reverse 5'-CTA AGT TGT CCT TAG CTG TCT CG-3' primers and a cycling protocol of denaturation at 95°C (2 min) followed by 40 cycles of 95°C (15 sec), 60°C (1 min) and 72°C (1 min) were applied. *P.melaninogenica* specific 16S rRNA gene amplicon directed forward 5'-GTGGGATAACCTGCCGAAAG-3' and reverse 5'-CCCATCCATTACCGATAAATCTTTA-3' primers have been described and tested before [28]. The cycling protocol used was 95°C (10 min) and 40 cycles of 95°C (15 sec), 60°C (1 min) and 72°C (1 min). PCR amplification of each samples was performed in duplicates and the 20 ul qPCR mixture consisted of 10 ul of 2X SYBR Green MasterMix (Applied Biosystems, Thermo Fisher Scientific) 1 ul of each primer, 7 ul PCR-grade water and 1ul of extracted DNA

Declarations

Consent for publication

Not applicable.

Data and materials availability

The datasets generated and/or analysed during the current study are available in the SRA repository, with accession number PRJNA699728. Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Burkhard Ludewig (burkhard.ludewig@kssg.ch).

Competing interests

The authors declare to competing financial intests.

Funding

Author contributions

B.L., S.J.S. and M.A.B. designed the study; B.L., A.D.M. and M.L. discussed data and wrote the paper; A.D.M., M.L., Y.S., C.E., J.C., and I.D. conducted experiments and discussed data; M.L., K.B. performed bioinformatics analyses and discussed data; K.MC. discussed data and provided reagents; M.G. discussed data; S.J.S., W.J., M.A.B., discussed data and provided patient material.

Acknowledgements

References

1. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR: Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers* 2020, 6(1):92.
2. Leemans CR, Snijders PJF, Brakenhoff RH: The molecular landscape of head and neck cancer. *Nat Rev Cancer* 2018, 18(5):269-282.
3. Weatherspoon DJ, Chattopadhyay A, Boroumand S, Garcia I: Oral cavity and oropharyngeal cancer incidence trends and disparities in the United States: 2000–2010. *Cancer Epidemiology* 2015, 39(4):497-504.
4. Egawa N, Egawa K, Griffin H, Doorbar J: Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. *Viruses* 2015, 7(7):3863-3890.

5. Faraji F, Zaidi M, Fakhry C, Gaykalova DA: Molecular mechanisms of human papillomavirus-related carcinogenesis in head and neck cancer. *Microbes Infect* 2017, 19(9-10):464-475.

6. Giuliano AR, Lu B, Nielson CM, Flores R, Papenfuss MR, Lee J-H, Abrahamsen M, Harris RB: Age-Specific Prevalence, Incidence, and Duration of Human Papillomavirus Infections in a Cohort of 290 US Men. *The Journal of Infectious Diseases* 2008, 198(6):827-835.

7. Wakabayashi R, Nakahama Y, Nguyen V, Espinoza JL: The Host-Microbe Interplay in Human Papillomavirus-Induced Carcinogenesis. *Microorganisms* 2019, 7(7).

8. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, Lakshmanan A, Wade WG: The human oral microbiome. *J Bacteriol* 2010, 192(19):5002-5017.

9. Structure, function and diversity of the healthy human microbiome. *Nature* 2012, 486(7402):207-214.

10. Huse SM, Ye Y, Zhou Y, Fodor AA: A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One* 2012, 7(6):e34242-e34242.

11. Martinez-Guryn K, Leone V, Chang EB: Regional Diversity of the Gastrointestinal Microbiome. *Cell Host & Microbe* 2019, 26(3):314-324.

12. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE: Defining the Normal Bacterial Flora of the Oral Cavity. *Journal of Clinical Microbiology* 2005, 43(11):5721-5732.

13. Swidsinski A, Göktas O, Bessler C, Loening-Baucke V, Hale LP, Andree H, Weizenegger M, Hölzl M, Scherer H, Lochs H: Spatial organisation of microbiota in quiescent adenoiditis and tonsillitis. *J Clin Pathol* 2007, 60(3):253-260.

14. Johnston J, Hoggard M, Biswas K, Astudillo-García C, Radcliff FJ, Mahadevan M, Douglas RG: Paired analysis of the microbiota between surface tissue swabs and biopsies from pediatric patients undergoing adenotonsillectomy. *Int J Pediatr Otorhinolaryngol* 2018, 113:51-57.

15. Johnston J, Hoggard M, Biswas K, Astudillo-García C, Waldvogel-Thurlow S, Radcliff FJ, Mahadevan M, Douglas RG: The bacterial community and local lymphocyte response are markedly different in patients with recurrent tonsillitis compared to obstructive sleep apnoea. *Int J Pediatr Otorhinolaryngol* 2018, 113:281-288.

16. Johnston J, Hoggard M, Biswas K, Astudillo-García C, Radcliff FJ, Mahadevan M, Douglas RG: Pathogen reservoir hypothesis investigated by analyses of the adenotonsillar and middle ear microbiota. *Int J Pediatr Otorhinolaryngol* 2019, 118:103-109.

17. Choi DH, Park J, Choi JK, Lee KE, Lee WH, Yang J, Lee JY, Park YJ, Oh C, Won H-R *et al*: Association between the microbiomes of tonsil and saliva samples isolated from pediatric patients subjected to tonsillectomy for the treatment of tonsillar hyperplasia. *Experimental & Molecular Medicine* 2020, 52(9):1564-1573.

18. Galli J, Calò L, Posteraro B, Rossi G, Sterbini FP, Paludetti G, Sanguinetti M: Pediatric oropharyngeal microbiome: Mapping in chronic tonsillitis and tonsillar hypertrophy. *Int J Pediatr Otorhinolaryngol* 2020, 139:110478.

19. Jensen A, Fagö-Olsen H, Sørensen CH, Kilian M: Molecular mapping to species level of the tonsillar crypt microbiota associated with health and recurrent tonsillitis. *PLoS One* 2013, 8(2):e56418.

20. Kim KS, Min HJ: Correlation between adenotonsillar microbiome and clinical characteristics of pediatrics with snoring. *Clin Exp Otorhinolaryngol* 2020.

21. Watanabe H, Goto S, Mori H, Higashi K, Hosomichi K, Aizawa N, Takahashi N, Tsuchida M, Suzuki Y, Yamada T *et al*: Comprehensive microbiome analysis of tonsillar crypts in IgA nephropathy. *Nephrology Dialysis Transplantation* 2017, 32(12):2072-2079.

22. Bahig H, Fuller CD, Mitra A, Yoshida-Court K, Solley T, Ping Ng S, Abu-Gheida I, Elgohari B, Delgado A, Rosenthal DI *et al*: Longitudinal characterization of the tumoral microbiome during radiotherapy in HPV-associated oropharynx cancer. *Clinical and Translational Radiation Oncology* 2021, 26:98-103.

23. Wang H, Funchain P, Bebek G, Altemus J, Zhang H, Niazi F, Peterson C, Lee WT, Burkey BB, Eng C: Microbiomic differences in tumor and paired-normal tissue in head and neck squamous cell carcinomas. *Genome Med* 2017, 9(1):14-14.

24. F. Escapa I, Huang Y, Chen T, Lin M, Kokaras A, Dewhirst FE, Lemon KP: Construction of habitat-specific training sets to achieve species-level assignment in 16S rRNA gene amplicon datasets. *Microbiome* 2020, 8(1):65.

25. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014, 15(12):550.

26. Knights D, Costello EK, Knight R: Supervised classification of human microbiota. *FEMS Microbiology Reviews* 2011, 35(2):343-359.

27. Siqueira JF, Jr., Rôças IN: Detection of Filifactor alocis in endodontic infections associated with different forms of periradicular diseases. *Oral Microbiol Immunol* 2003, 18(4):263-265.

28. Martin FE, Nadkarni MA, Jacques NA, Hunter N: Quantitative microbiological study of human carious dentine by culture and real-time PCR: association of anaerobes with histopathological changes in chronic pulpitis. *Journal of clinical microbiology* 2002, 40(5):1698-1704.

29. Westra WH, Lewis JS, Jr.: Update from the 4th Edition of the World Health Organization Classification of Head and Neck Tumours: Oropharynx. *Head Neck Pathol* 2017, 11(1):41-47.

30. Näsman A, Du J, Dalianis T: A global epidemic increase of an HPV-induced tonsil and tongue base cancer - potential benefit from a pan-gender use of HPV vaccine. *J Intern Med* 2020, 287(2):134-152.

31. Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. *Cell* 2011, 144(5):646-674.

32. Sami A, Elimairi I, Stanton C, Ross RP, Ryan CA: The Role of the Microbiome in Oral Squamous Cell Carcinoma with Insight into the Microbiome-Treatment Axis. *Int J Mol Sci* 2020, 21(21).

33. Tuominen H, Rautava J: Oral Microbiota and Cancer Development. *Pathobiology* 2020:1-11.

34. Binder Gallimidi A, Fischman S, Revach B, Bulvik R, Maliutina A, Rubinstein AM, Nussbaum G, Elkin M: Periodontal pathogens Porphyromonas gingivalis and Fusobacterium nucleatum promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget* 2015, 6(26):22613-22623.

35. Al-hebshi NN, Nasher AT, Maryoud MY, Homeida HE, Chen T, Idris AM, Johnson NW: Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. *Scientific Reports* 2017, 7(1):1834.

36. Yang C-Y, Yeh Y-M, Yu H-Y, Chin C-Y, Hsu C-W, Liu H, Huang P-J, Hu S-N, Liao C-T, Chang K-P *et al*: Oral Microbiota Community Dynamics Associated With Oral Squamous Cell Carcinoma Staging. *Front Microbiol* 2018, 9:862-862.

37. Zhao H, Chu M, Huang Z, Yang X, Ran S, Hu B, Zhang C, Liang J: Variations in oral microbiota associated with oral cancer. *Scientific Reports* 2017, 7(1):11773.

38. Nave H, Gebert A, Pabst R: Morphology and immunology of the human palatine tonsil. *Anat Embryol (Berl)* 2001, 204(5):367-373.

39. Pavlova SI, Wilkening RV, Federle MJ, Lu Y, Schwartz J, Tao L: Streptococcus endopeptidases promote HPV infection in vitro. *Microbiologyopen* 2019, 8(1):e00628.

40. Mukherjee PK, Wang H, Retuerto M, Zhang H, Burkey B, Ghannoum MA, Eng C: Bacteriome and mycobiome associations in oral tongue cancer. *Oncotarget* 2017, 8(57):97273-97289.

41. Guerrero-Preston R, Godoy-Vitorino F, Jedlicka A, Rodríguez-Hilario A, González H, Bondy J, Lawson F, Folawiyo O, Michailidi C, Dziedzic A *et al*: 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. *Oncotarget* 2016, 7(32):51320-51334.

42. Banerjee S, Tian T, Wei Z, Peck KN, Shih N, Chalian AA, O'Malley BW, Weinstein GS, Feldman MD, Alwine J *et al*: Microbial Signatures Associated with Oropharyngeal and Oral Squamous Cell Carcinomas. *Scientific Reports* 2017, 7(1):4036.

43. Börnigen D, Ren B, Pickard R, Li J, Ozer E, Hartmann EM, Xiao W, Tickle T, Rider J, Gevers D *et al*: Alterations in oral bacterial communities are associated with risk factors for oral and oropharyngeal cancer. *Sci Rep* 2017, 7(1):17686.

44. Lim Y, Fukuma N, Totsika M, Kenny L, Morrison M, Punyadeera C: The Performance of an Oral Microbiome Biomarker Panel in Predicting Oral Cavity and Oropharyngeal Cancers. *Front Cell Infect Microbiol* 2018, 8:267.

45. Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM: 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* 2005, 3:27.

46. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013, 79(17):5112-5120.

47. Martin M: Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* 2011, 17(1):3.

48. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP: DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 2016, 13(7):581-583.

49. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM: Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014, 42(Database issue):D633-642.

50. Wang Q, Garrity GM, Tiedje JM, Cole JR: Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007, 73(16):5261-5267.

51. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S: msa: an R package for multiple sequence alignment. *Bioinformatics* 2015, 31(24):3997-3999.

52. Patino LH, Muñoz M, Cruz-Saavedra L, Muskus C, Ramírez JD: Genomic Diversification, Structural Plasticity, and Hybridization in Leishmania (Viannia) braziliensis. *Frontiers in Cellular and Infection Microbiology* 2020, 10(635).

53. McMurdie PJ, Holmes S: phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013, 8(4):e61217.

54. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P *et al*: vegan: Community Ecology Package. R package version 2.5-7. 2018.

55. Foster ZSL, Sharpton TJ, Grünwald NJ: Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLOS Computational Biology* 2017, 13(2):e1005404.

56. Liaw A, Wiener M: Classification and Regression by randomForest. *R News* 2002, 2(3):18–22.

57. Evans JS, Murphy MA: rfUtilities. R package version 2.1-5, https://cran.r-project.org/package=rfUtilities. 2018.

58. Kuhn M: caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret. 2020.

59. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA: Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 1990, 56(6):1919-1925.
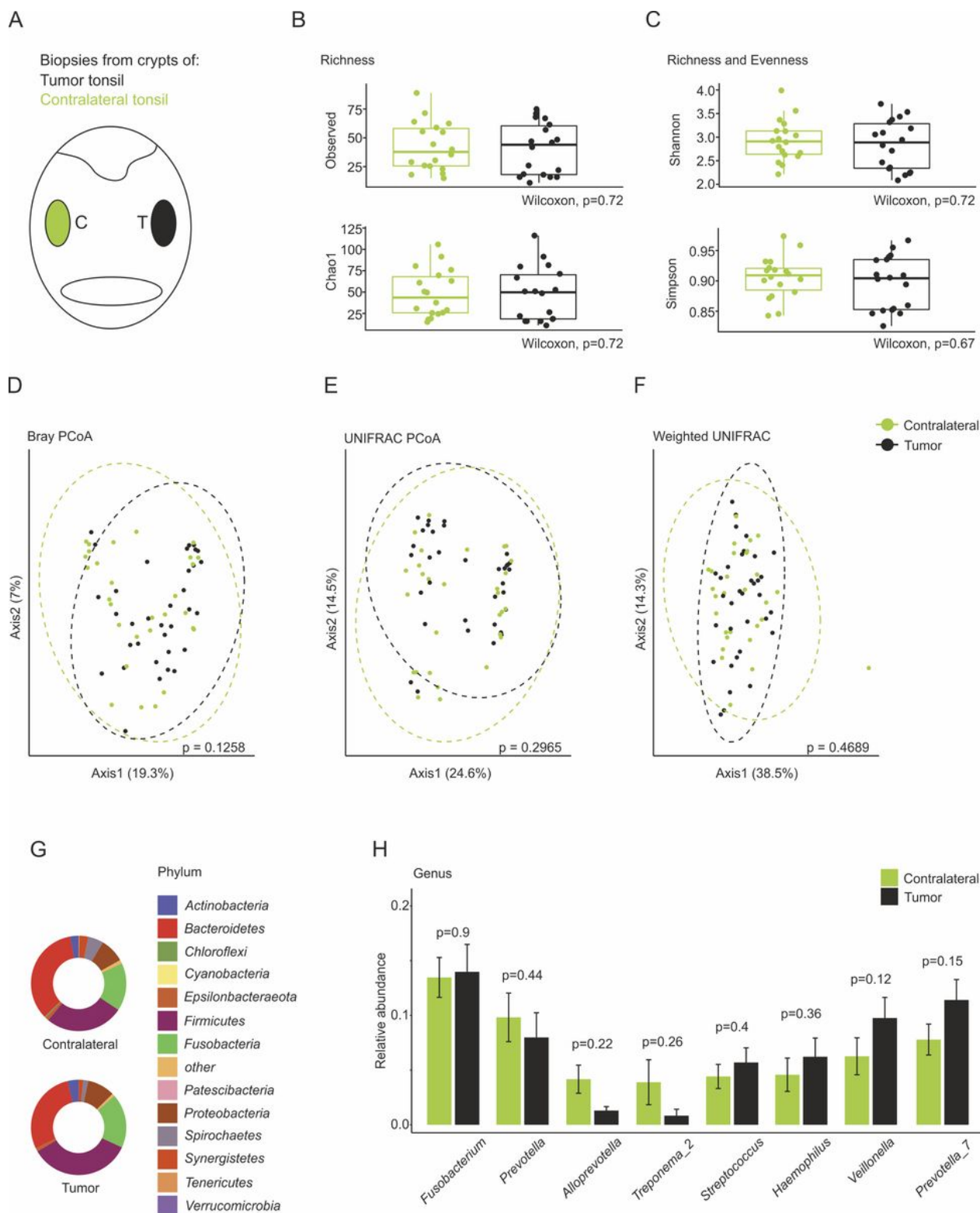
Figures
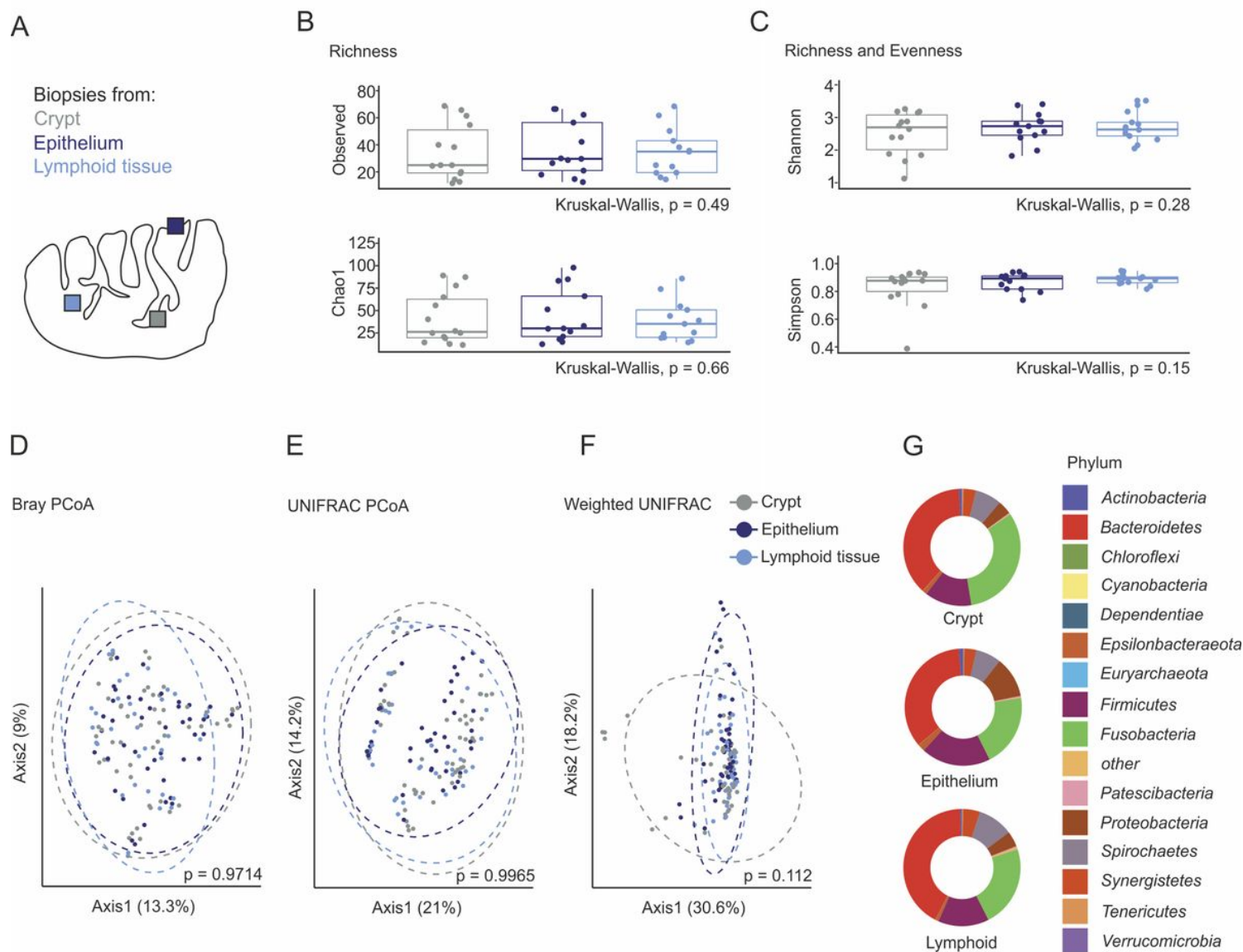
Figure 1

Microbiome diversity in different tonsillar compartments of obstructive sleep apnea (OSA) patients. (A) Schematic representation of the human palatine tonsil. The different sampling sites within the tonsil are indicated in grey (crypt), dark blue (epithelium) and bright blue (lymphoid tissue). (B-C) α-diversity in different sampling sites of OSA patients (n=14 patients, only crypt samples have been available for patient 1). (B) Species richness measured in terms of observed number of species and Chao1 index. (C) Shannon and Simpson indices estimating species richness and evenness. (D-F) Principle coordinate analysis of the tonsillar microbiome in distinct sampling sites calculated using BrayCurtis (D), UNIFRAC (E) and weighted UNIFRAC (F) distance matrices. Dots represent crypt, epithelial and lymphoid tissue biopsies of individual patients (n=14 patients). Dotted circles respresent superimposed normal probability over datapoints. (G) Bacterial community composition in different sampling locations of OSA at the phylum level. Statistical analysis was performed using Kruskal-Wallis (B-C) or paired permanova (D-F) tests.
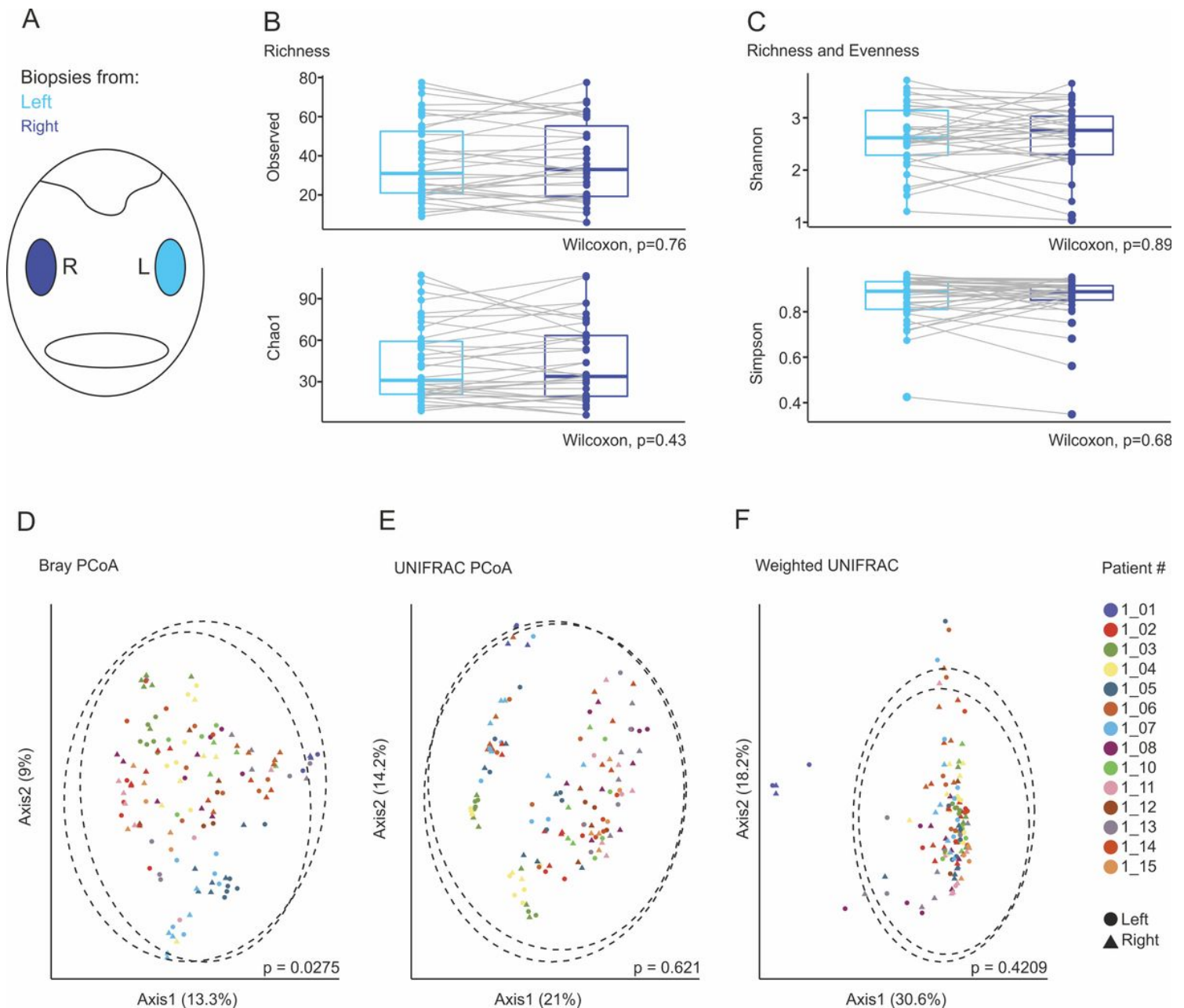
Figure 2

Microbial composition in ipsi- and contralateral tonsils of patients with obstructive sleep apnea. (A) Schematic representation of the oropharynx representing samples collected from right (dark blue) and left (bright blue) tonsils. (B-C) α-diversity displayed by observed number of species and Chao1 indices (B) and Shannon and Simpson indices (C). Dots represent crypt, epithelial and lymphoid tissue biopsies of individual patients (n=14 patients). (C-E) Principle coordinate analysis calculated using BrayCurtis (C), UNIFRAC (D) and weighted UNIFRAC (E) and distance matrices. Samples of each patient are displayed by a distinct color. Dotted circles respresent superimposed normal probability over datapoints. Statistical analysis was performed using Wilcoxon signed-rank test (B-C) or paired permanova (D-F) test.

Figure 3

The tonsillar microbiome composition in tonsil cancer patients. (A) Schematic representation of the oropharynx illustrating crypt biopsies collected from tumor-affected and non-affected contralateral tonsils. (B-C) Comparison of α-diversity between tumor-affected and contralateral tonsils. Observed number of species, Chao1, Simpson and Shannon indices are shown. (D-F) Principle coordinate analysis using BrayCurtis (D), UNIFRAC (E) and weighted UNIFRAC (F) distance matrices. Dotted circles respresent

superimposed normal probability over datapoints. (G) Taxonomic characterisation of the microbiome in tonsil cancer patients on the phylum level. (H) Relative abundance of the top eight highest abundant genera detected in tumor and contralateral tonsils. Statistical analysis was performed using Wilcoxon-Mann-Whitney test (B-C, H) or paired permanova test (D-F).
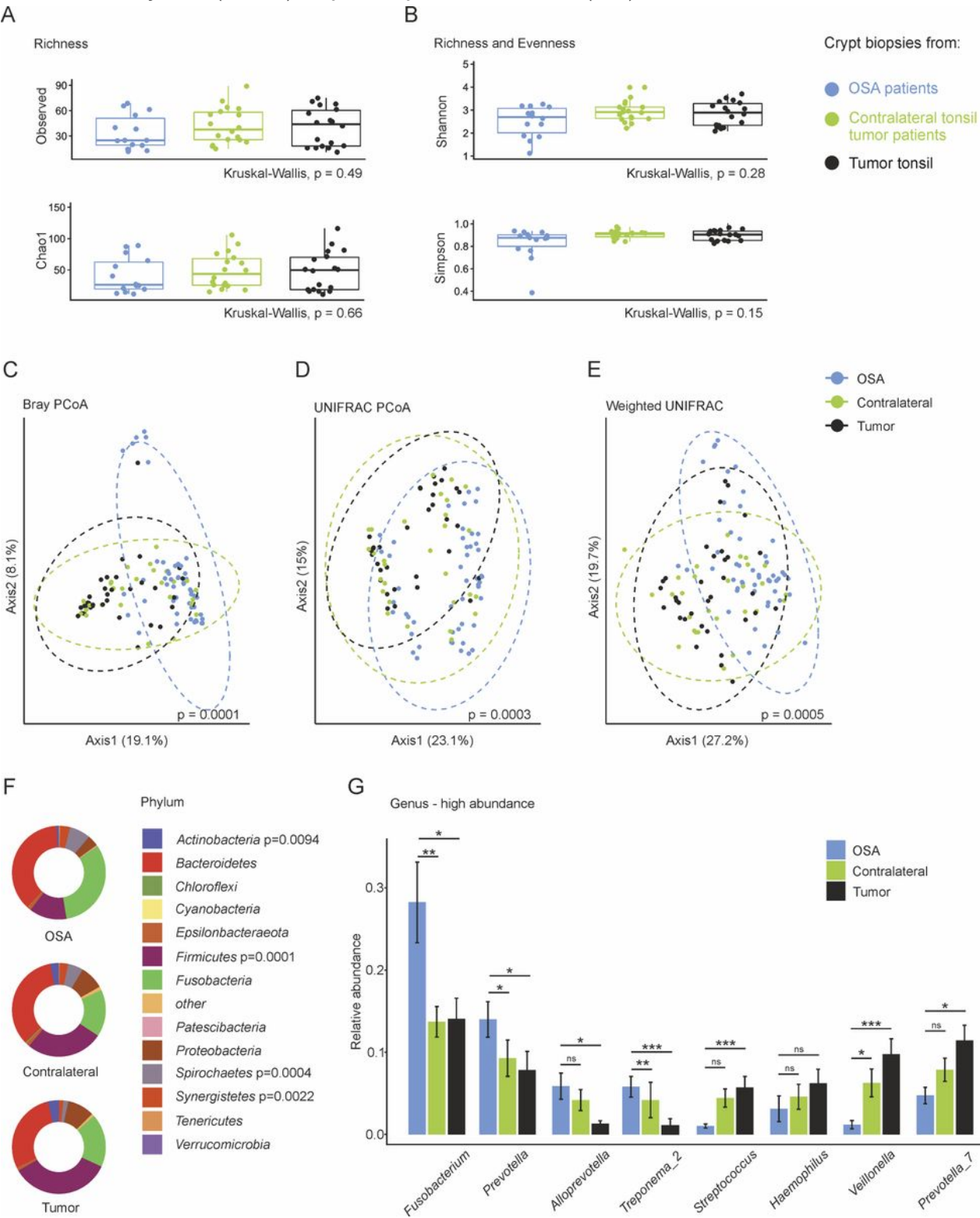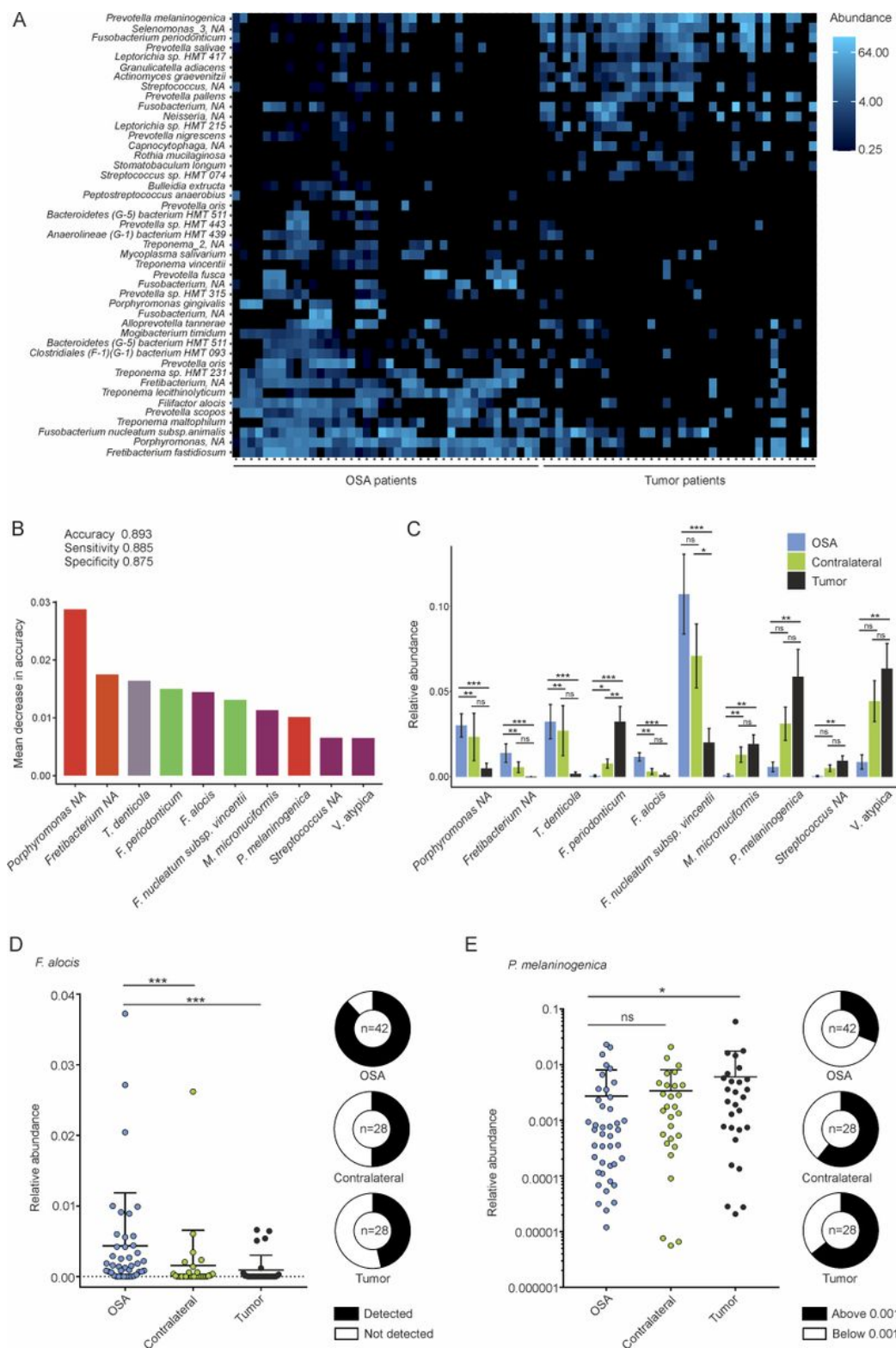

Figure 4

Comparison of the crypt microbiome in tonsil cancer and obstructive sleep apnea (OSA) patients. (A-B) α-diversity in crypt biopsies of OSA and tonsil cancer patients with (A) species richness assessed by the observed number of species and Chao1 index and (B) Shannon and Simpson indices indicating species richness and evenness. (C-E) Principle coordinate analysis using BrayCurtis (C), UNIFRAC (D) and weighted UNIFRAC (E) distance matrices. Dotted circles respresent superimposed normal probability over datapoints. (F) Bacterial community composition in tumor and OSA patients samples on the phylum level. Significant p-values from the statistical analysis are indicated. (G) Comparison of the top eight genera with the highest overall relative abundance. Statistical analysis was performed using Kruskall-Wallis (A-B, F), paired permanova test (C-E), or Wilcoxon-Mann-Whitney test (G) with *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant.

Figure 5

Characterization of tonsil cancer and obstructive sleep apnea (OSA) patient microbiomes at species level. (A) Heatmap showing top 45 differentially abundant species in crypts of tumor patients compared to OSA patients. (B) Mean decrease in accuracy of top ten tumor-predictive species calculated by means of random forest analysis. Bar color indicates phylum affiliation as used in Figure 4F of respective species. (C) Relative abundance of top ten predictive species plotted for tumor-bearing and contralateral tonsils of

tumor and OSA patients. Statistical analysis was performed using Wilcoxon-Mann-Whitney test (C-E) with *, p < 0.05; **, p < 0.01; ***, p < 0.001; ns, not significant.

Supplementary Files

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

- SupplementaryFigures.pdf
- SupplementaryTables1and2.pdf