JAMA Oncology | Original Investigation

Oral Microbiome and Subsequent Risk of Head and Neck Squamous Cell Cancer

Soyoung Kwak, PhD; Chan Wang, PhD; Mykhaylo Usyk, PhD; Feng Wu, PhD; Neal D. Freedman, PhD; Wen-Yi Huang, PhD; Marjorie L. McCullough, ScD; Caroline Y. Um, PhD; Martha J. Shrubsole, PhD; Qiuyin Cai, PhD; Huilin Li, PhD; Jiyoung Ahn, PhD; Richard B. Hayes, DDS, PhD

IMPORTANCE The oral microbiota may be involved in development of head and neck squamous cell cancer (HNSCC), yet current evidence is largely limited to bacterial 16S amplicon sequencing or small retrospective case-control studies.

OBJECTIVE To test whether oral bacterial and fungal microbiomes are associated with subsequent risk of HNSCC development.

DESIGN, SETTING, AND PARTICIPANTS Prospective nested case-control study among participants providing oral samples in 3 epidemiological cohorts, the American Cancer Society Cancer Prevention Study II Nutrition Cohort, the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial, and the Southern Community Cohort Study. Two hundred thirty-six patients who prospectively developed HNSCC were identified during a mean (SD) of 5.1 (3.6) years of follow-up. Control participants who remained HNSCC free were selected by 2:1 frequency matching on cohort, age, sex, race and ethnicity, and time since oral sample collection. Data analysis was conducted in 2023.

EXPOSURES Characterization of the oral bacterial microbiome using whole-genome shotgun sequencing and the oral fungal microbiome using internal transcribed spacer sequencing. Association of bacterial and fungal taxa with HNSCC was assessed by analysis of compositions of microbiomes with bias correction. Association with red and orange oral pathogen complexes was tested by logistic regression. A microbial risk score for HNSCC risk was calculated from risk-associated microbiota.

MAIN OUTCOMES AND MEASURES The primary outcome was HNSCC incidence.

RESULTS The study included 236 HNSCC case participants with a mean (SD) age of 60.9 (9.5) years and 24.6% women during a mean of 5.1 (3.6) years of follow-up, and 485 matched control participants. Overall microbiome diversity at baseline was not related to subsequent HNSCC risk; however 13 oral bacterial species were found to be differentially associated with development of HNSCC. The species included the newly identified *Prevotella salivae*, *Streptococcus sanguinis*, and *Leptotrichia* species, as well as several species belonging to beta and gamma Proteobacteria. The red/orange periodontal pathogen complex was moderately associated with HNSCC risk (odds ratio, 1.06 per 1 SD; 95% CI, 1.00-1.12). A 1-SD increase in microbial risk score (created based on 22 bacteria) was associated with a 50% increase in HNSCC risk (multivariate odds ratio, 1.50; 95% CI, 1.21-1.85). No fungal taxa associated with HNSCC risk were identified.

CONCLUSIONS AND RELEVANCE This case-control study yielded compelling evidence that oral bacteria are a risk factor for HNSCC development. The identified bacteria and bacterial complexes hold promise, along with other risk factors, to identify high-risk individuals for personalized prevention of HNSCC.

JAMA Oncol. doi:10.1001/jamaoncol.2024.4006
Published online September 26, 2024.

Supplemental content

Author Affiliations: Department of Population Health, NYU Grossman School of Medicine, New York, New York (Kwak, Wang, Usyk, Wu, Li, Ahn, Hayes); NYU Laura and Isaac Perlmutter Cancer Center, New York, New York (Kwak, Wang, Usyk, Wu, Li, Ahn, Hayes); Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland (Freedman, Huang); Department of Population Science, American Cancer Society, Atlanta, Georgia (McCullough, Um); Vanderbilt-Ingram Cancer Center, Nashville, Tennessee (Shrubsole, Cai); Department of Medicine, Division of Epidemiology, Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, Tennessee (Shrubsole, Cai).

Corresponding Authors: Jiyoung Ahn, PhD (jiyoung.ahn@nyulangone.org), and Richard B. Hayes, DDS, PhD (richard.b.hayes@nyulangone.org), NYU Grossman School of Medicine, 180 Madison Ave, New York, NY 10016.

Head and neck squamous cell cancer (HNSCC; cancers of the oral cavity, pharynx, and larynx) confers significant physical disfigurement and a poor prognosis with about 60% 5-year survival.[1] In addition to established risk factors—smoking,[2] alcohol use,[3] and human papillomavirus (HPV)[4]—recent attention has focused on the possible role of other microbiota in HNSCC development,[5-7] potentially related to microbial metabolism of HNSCC carcinogens[8] or to microbial factors associated with poor oral hygiene and periodontal disease.[9-11] Recognizing this, there is a critical need to more comprehensively identify novel microbial risk factors that potentially could be used to help prevent HNSCC.

The human oral cavity hosts a diverse bacterial community or microbiome.[12] Studies point to the differential prevalence of certain oral bacteria in oral or tumor samples of patients with HNSCC[13]; however, there is no consensus from these small retrospective case-control studies regarding the specific oral taxa involved etiologically in HNSCC. Addressing this limitation, we recently carried out a prospective study using prediagnostic oral samples and provided evidence that several genera in Proteobacteria are associated with reduced HNSCC risk.[14] However, this evidence was based on low-resolution 16S ribosomal RNA-based microbiome profiling. We now require high-resolution bacterial species-level identification in large prospective cohorts for the purposes of microbially targeted early detection and intervention.

Oral fungi have been less studied in HNSCC.[15] Fungi can activate carcinogens and promote inflammation and neoplasia.[16-20] Fungi are abundant in oral samples[21] and in HNSCC tumor tissue[22,23] yet have largely been unexamined in prospective studies. Furthermore, few studies have collectively examined both bacterial and fungal microbiome,[7,24] although bacteria and fungi may interact physically and metabolically in the oral cavity[25] and often occur symbiotically.[26]

We conducted a prospective case-control study nested in 3 large, well-established, prospective US cohorts to determine whether the oral bacterial and fungal microbiome is associated with subsequent HNSCC risk. We assessed risk of HNSCC associated with oral taxa and with periodontal disease-related bacterial complexes,[27,28] individually and as summarized in a microbial risk score.[29]

Key Points

Question Is the prediagnostic oral bacterial and fungal microbiome associated with subsequent development of head and neck squamous cell cancer (HNSCC)?

Findings In 3 large prospective cohorts including 159 840 individuals, we identified novel oral bacterial species were identified, including both commensals and periodontal pathogenic bacterial complexes that together conferred a 50% increased risk of HNSCC. No significant associations between oral fungi and risk of HNSCC were found.

Meaning The identified bacteria and bacterial complexes hold promise, along with other risk factors, for identifying high-risk individuals for personalized prevention of HNSCC.

Methods

All participants provided written informed consent and all protocols were approved by the NYU Grossman School of Medicine institutional review board. Data analysis was conducted in 2023.

Study Population

Parent Cohorts

The 3 US cohorts on which this study is based represent a diversity of US populations resident across the US. The American Cancer Society Cancer Prevention Study II Nutrition Cohort (ACS-CPS-II)[30] included more than 184 000 participants aged 50 to 74 years from 21 US states. Oral wash samples were collected during 2001-2002 from 70 000 cohort members. Follow-up for incident cancer was conducted through June 2009. The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)[31] was a large randomized trial at 10 US screening centers that examined the effect of screening on cancer mortality in participants aged 55 to 74 years. Oral wash samples were collected in the control group only (n = 52 000) between 1993 and 2001. Cancer follow-up was conducted through December 2010. The Southern Community Cohort Study (SCCS)[32] was a prospective cohort of 85 000 adults, two-thirds African American, aged 40 to 70 years, from 12 southeastern states, conducted between 2002 and 2009. Oral wash samples were collected from 37 836 participants at enrollment. Follow-up for cancer was conducted through 2021.

Nested Case-Control Study

Case participants had histologically confirmed incident HNSCC, including cancers of the oral cavity (*International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* [*ICD-10*] codes C02.0-C06.9; excluding salivary glands), pharynx (*ICD-10* codes C09.0-C10.9 and C12-C13.9; excluding nasopharynx), and larynx (*ICD-10* codes C32.0-32.9). Cohort nested control participants were selected by incidence-density sampling (2:1 frequency matched) by cohort, age, sex, race and ethnicity (African American, White, or other/unknown), and time since oral wash sample collection (in years). All case and control participants provided written informed consent, provided oral wash samples (prediagnostically for case patients), and had no history of cancer (except nonmelanoma skin cancer). Fourteen controls were excluded owing to unavailability of matching controls and sequencing failure, resulting in 14 case patients having only 1 matched control.

Measurements

Demographic Information

At enrollment and during follow-up, cohort participants completed structured questionnaires. We used covariates from the questionnaires most closely preceding oral wash sample collection. Oral HPV-16 status was available in the ACS-CPS-II and PLCO cohorts.[4] HPV-16 typing was conducted using the MY09/11 L1-targeted degenerate primer polymerase chain reaction system, a multiplexed next-generation sequencing

method, and a real-time polymerase chain reaction assay, as previously described.[4]

Oral Wash Sample Collection

Study participants from the 3 cohorts provided oral wash samples following similar collection protocols.[30-32] The oral microbiome tends to be stable over time[33-35] and shows much greater interindividual than intraindividual variation, indicating that a 1-time oral sample collection is useful for assessing the oral microbiome in a cohort study.

Bacterial Microbiome Whole-Genome Sequencing Assay and Data Processing

Using DNA from oral samples, whole-genome sequencing was carried out at NYU Langone's Genome Technology Center. Library preparation was performed using the Nextera DNA Flex Library Preparation Kit (Illumina), and further sequenced using the Illumina NovaSeq SP 300-cycle system. FASTQ sequence reads were demultiplexed, and human reads and low-quality reads were removed. Using the SHOGUN pipeline,[36] reads were mapped and labeled with the National Center for Biotechnology Information (NCBI) taxonomic annotation at the species level. Reads mapping to multiple reference genomes are labeled as the last common ancestor according to the NCBI taxonomy.[37] The whole-genome sequencing assay generated sufficient reads, represented by major bacterial phyla, distributed similarly across the cohorts (eTables 1 and 2 in Supplement 1).

Fungal Microbiome Internal Transcribed Spacer Sequencing Assay and Data Processing

We conducted an internal transcribed spacer fungal-specific gene sequencing assay.[38] Briefly, the DNA underwent polymerase chain reaction using forward (5′-ACACACCGCCCGTCGCTACT-3′) and reverse primers (5′-TTTCGCTGCGTTCTTCATCG-3′). Polymerase chain reaction amplicons were pooled and purified using Agencourt AMPure XP beads (Beckman Coulter), and the library was prepared using the KAPA LTP library preparation kit (KAPA Biosystems). Sequencing used Illumina MiSeq. Reads were analyzed using DADA2[39] and taxonomy was assigned using the UNITE database.[40] Open-reference operational taxonomic unit (OTU) picking was done using the QIIME version 1.9 open-reference OTU picking protocol and the UNITE database for the reference-based clustering component. VSEARCH version 1.4.0 was substituted for USEARCH to facilitate higher throughput.[41] The OTU clustering threshold was set at 99% sequence identity to account for fungal heterogeneity. Sequence dereplication and chimera removal were performed using the QIIME quality control protocol. Representative sequences for each OTU cluster were chosen based on sequence abundance. BLAST[42] was used to assign taxonomy using the UNITE database. The internal transcribed spacer assay generated sufficient reads for analysis. The vast majority of taxa across the 3 cohorts were of 2 phyla: Ascomycota and Basidiomycota (eTables 1 and 2 in Supplement 1).

Quality Control Assessment

Laboratory personnel were blinded to case and control status. Blinded random repeated quality control samples, positive microbial controls (8 bacterial species and 2 fungal species, ZymoBIOMICS Microbial Community DNA Standard II), and negative controls (distilled water) were inserted across all sequencing batches. The intraclass correlation coefficient[43] for the blinded random repeated controls showed good agreement (a coefficient between 0.90 and 1 for whole-genome sequencing and a coefficient between 0.62 and 1 for internal transcribed spacer sequencing) (eFigure 1 in Supplement 1), with distinct clusters identified for random quality control, positive control, and negative control samples. All species in the positive controls were detected, and the negative controls showed no microbial DNA.

Statistical Analysis

Microbial composition β diversity (between-participant diversity), assessed by Jensen-Shannon divergence[44] or Jaccard distance for case and control participants, was compared using permutational multivariate analysis of variance,[45] adjusting for covariates. The α diversity (within-participant diversity), assessed by richness, Shannon Diversity Index, and Pielou evenness, was compared using linear regressions. The α diversity and β diversity were calculated for bacteria after rarefaction at 100 000 read depth and for fungi after rarefaction at 500 read depth, for which the thresholds were determined based on the rarefaction curves (eFigure 2 in Supplement 1).

We compared bacterial and fungal taxa between HNSCC case and control participants up to the species level using analysis of compositions of microbiomes with bias correction (ANCOM-BC),[46] including adjustments for covariates (age, sex, race and ethnicity, smoking status, number of cigarettes per day for those reporting ever smoking, alcohol consumption status, grams of ethanol consumed per day for those reporting consuming alcohol, and oral HPV-16 status). The default settings for the ANCOM-BC function at α= .05 were used, except for zero_cut = 1, lib_cut = 10 000 for bacteria, and lib_cut = 500 for fungi. We limited bacterial and fungal taxa for analysis to those with mean relative abundance of 0.001% or greater and filtered to include only taxa with at least 2 sequences in at least 10% (for bacteria) or 5% (for fungi) of the participants. First, cohort-specific estimates of differential abundance by species were obtained. Then, cohort-specific estimates and their corresponding standard errors at α < .05 were used to perform random-effects meta-analysis[47] to generate the pooled estimates for the 3 cohorts. The results were visualized as a cladogram using GraPhlAn.[48]

We further confirmed the robustness of ANCOM-BC–selected bacterial species using a permutation test with 500 iterations. The *P* value for *j*th species was calculated as

$$\frac{\sum_{i}^{500} I(P_{ij} \geq O_j) + 1}{501}$$

where $I(\cdot)$ was the indicator function, O_j was the frequency of *j*th species identified (ie, with a meta $P < .05$ in differential abundance test) using the original cohorts, and P_{ij} was the frequency of *j*th species identified using the permuted data in *i*th

Table 1. Selected Characteristics of 236 Incident Cases of Head and Neck Squamous Cell Cancer and 458 Control Participants in 3 US Cohorts

Characteristics	Cohort					
	ACS-CPS-II		PLCO		SCCS	
	Cases (n = 55)	Controls (n = 106)	Cases (n = 71)	Controls (n = 138)	Cases (n = 110)	Controls (n = 214)
Age, mean (SD), y	71.2 (6.3)	71.4 (6.3)	62.7 (4.8)	63.3 (5.0)	54.5 (7.9)	54.4 (7.8)
Sex, No. (%)						
Female	15 (27.3)	29 (27.4)	13 (18.3)	25 (18.1)	30 (27.3)	58 (21.7)
Male	40 (72.7)	77 (72.6)	58 (81.7)	113 (81.9)	80 (72.7)	156 (72.9)
Race and ethnicity, No. (%)						
African American or other[a]	0	0	5 (7.0)	9 (6.5)	81 (73.6)	156 (72.9)
White	55 (100)	106 (100)	66 (93.0)	129 (93.5)	28 (25.5)	56 (26.2)
Smoking status, No. (%)						
Never	10 (18.2)	57 (53.8)	8 (11.3)	69 (50.0)	13 (11.8)	65 (30.4)
Former	30 (54.5)	46 (43.4)	38 (53.5)	62 (44.9)	25 (22.7)	68 (31.8)
Current[b]	15 (27.3)	3 (2.8)	25 (35.2)	7 (5.1)	69 (62.7)	79 (36.9)
No. of cigarettes per day for ever smoking, mean (SD)[c]	24.9 (12.1)	19.7 (11.6)	26.1 (12.6)	21.5 (13.7)	18.0 (13.0)	16.5 (13.0)
Alcohol consumption, No. (%)						
No[d]	12 (21.8)	24 (22.6)	8 (11.3)	43 (31.2)	34 (30.9)	85 (39.7)
Yes	34 (61.8)	62 (58.5)	50 (70.4)	88 (63.8)	74 (67.3)	125 (58.4)
Missing data	9 (16.4)	20 (18.9)	13 (18.3)	7 (5.1)	2 (1.8)	4 (1.9)
Ethanol consumption among those reporting alcohol consumption, mean (SD), g/d[e]	17.0 (14.6)	12.5 (17.5)	38.3 (80.6)	11.4 (16.6)	53.4 (80.8)	34.5 (60.4)
Human papillomavirus 16 status, No. (%)[d]	4 (7.3)	1 (0.9)	7 (9.9)	1 (0.7)	NA	NA
Cancer site						
Oral cavity	18 (32.7)		22 (31.0)		22 (20)	
Pharynx, base of tongue	12 (21.8)		18 (25.4)		33 (30)	
Larynx	25 (45.5)		31 (43.7)		55 (50)	
Time to diagnosis, mean (SD), y[f]	2.9 (2.0)		4.3 (2.5)		6.7 (4.1)	

Abbreviations: ACS-CPS-II, American Cancer Society Cancer Prevention Study II Nutrition Cohort; NA, data not available; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SCCS, Southern Community Cohort Study.

[a] Due to the small sample size of the "other" group, it was combined with the "African American" group for the analysis.

[b] $P < .001$ by χ^2 test for comparison between case and control participants in each cohort.

[c] $P < .05$ by t test for comparison between case and control participants in the ACS-CPS-II and PLCO cohorts.

[d] $P < .05$ by χ^2 test for comparison between case and control participants in the PLCO cohort.

[e] $P < .05$ by t test for comparison between case and control participants in the PLCO cohort.

[f] Time from oral sample collection to head and neck squamous cell cancer diagnosis (in years).

iteration. The Q value was then calculated adjusting for the false discovery rate across all species.

We also examined periodontal bacterial pathogen complexes, selected a priori,[27,28] for their association with HNSCC, including the red complex (*Treponema denticola*, *Porphyromonas gingivalis*, *and Tannerella forsythia*), the orange complex (*Fusobacterium nucleatum*, *Prevotella intermedia*, *Prevotella nigrescens*, *Eubacterium nodatum*, *Campylobacter showae*, and *Campylobacter gracilis*), and their combination (ie, red/orange complex). Unconditional logistic regression was used for this testing, with adjustments made for the aforementioned covariates.

Finally, building on the concept of a polygenic risk score used productively in genomic research,[49] we constructed a microbial risk score[29] to summarize the microbial profiles in relation to HNSCC risk. The microbial risk score was calculated as a weighted sum of the relative abundance of selected bacterial species, with weights assigned according to the estimated effect sizes for individual species, determined via microbiome-wide ANCOM-BC. To evaluate the reproducibility of a microbial risk score, we conducted Monte Carlo cross-validation[50] by randomly splitting data into training and test sets at a 70% to 30% ratio, with this process repeated 50 times.

Results

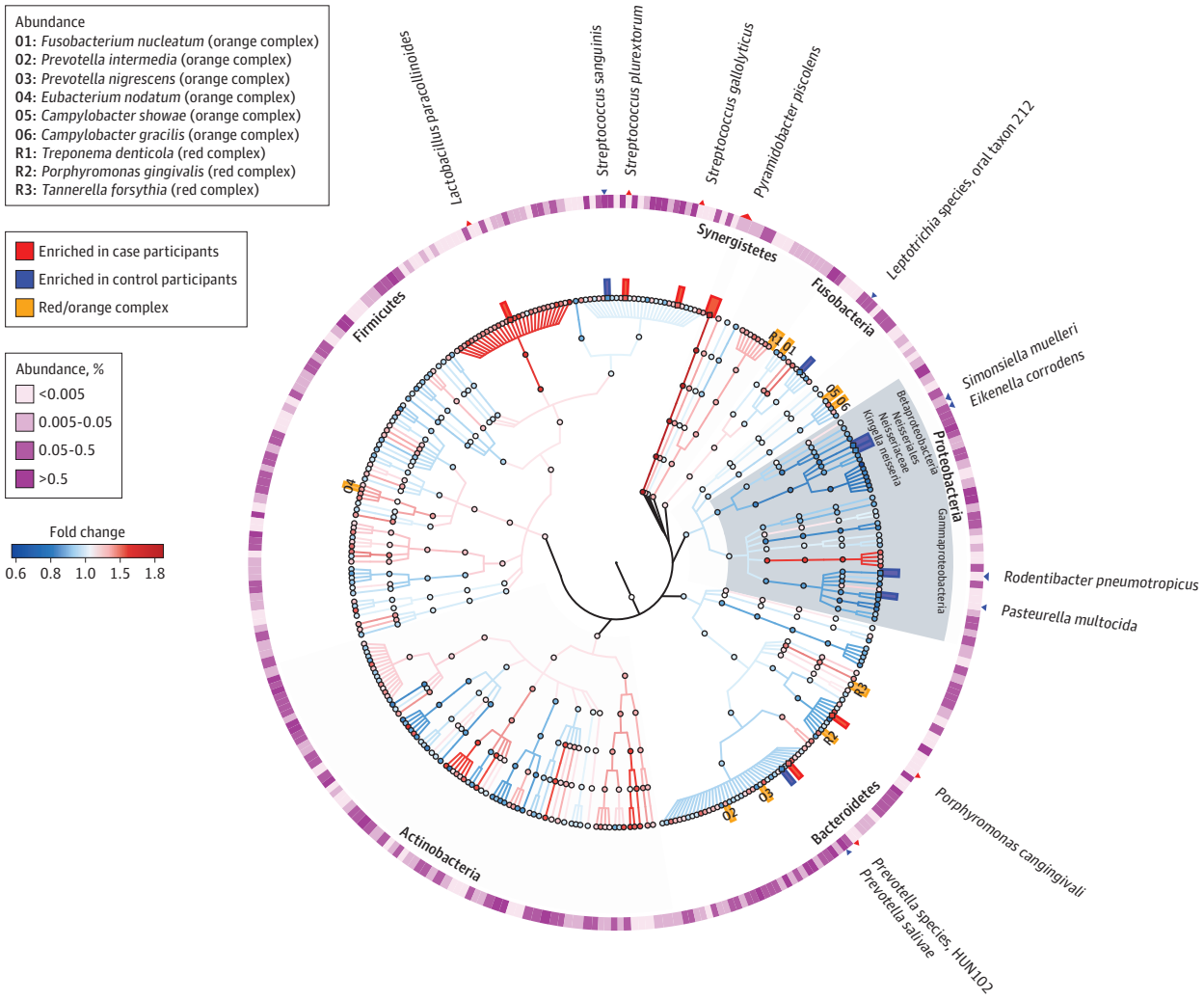
Participant Characteristics

This study included 236 men and women diagnosed with HNSCC during a mean of 5.1 years of follow-up and 458 matched control participants. The characteristics of the study participants are presented in (**Table 1**). Matching factors including age, sex, and race and ethnicity did not differ by case and control status for the 3 cohorts. As expected, case participants tended to use tobacco and alcohol more frequently and had a greater prevalence of oral HPV-16. Study participants in the ACS-

Figure 1. Oral Bacteria and Risk of HNSCC

A, Fold changes and 95% CIs for the association between selected taxa and risk of head and neck squamous cell cancer (HNSCC) from random-effects meta-analysis of analysis of compositions of microbiomes with bias correction abundance estimates of the 3 cohorts (American Cancer Society Cancer Prevention Study II Nutrition Cohort [ACS-CPS-II]; Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial [PLCO]; and Southern Community Cohort Study [SCCS]). Complete results for all bacteria are shown in eTable 4 in Supplement 1. B, Fold changes and 95% CIs for the association between selected taxa and risk of HNSCC for the 3 individual cohorts. C, Odds ratios and 95% CIs for the association between a priori periodontal pathogen complexes and risk of HNSCC from random-effects meta-analysis of logistic regression estimates for the 3 cohorts. The red complex includes *Treponema denticola*, *Porphyromonas gingivalis*, and *Tannerella forsythia* and the orange complex includes *Fusobacterium nucleatum*, *Prevotella intermedia*, *Prevotella nigrescens*, *Eubacterium nodatum*, *Campylobacter showae*, and *Campylobacter gracilis*. D, Odds ratios and 95% CIs for the association between red and orange periodontal pathogen complexes and risk of HNSCC for the 3 cohorts. All models were adjusted for age, sex, race and ethnicity, smoking status, number of cigarettes per day for those reporting ever smoking, alcohol consumption status, grams of ethanol consumed per day for those reporting alcohol consumption, and oral human papillomavirus 16 status as covariates. Fold change values greater than 1 indicate enrichment in HNSCC case participants, while values less than 1 indicate enrichment in control participants. The size of the points indicates relative abundance.

Figure 2. Cladogram Representation of Oral Microbiome Species Associated With HNSCC



The color of nodes and branches represents fold changes in head and neck squamous cell cancer (HNSCC) case participants compared with control participants, ranging from blue to red. A red branch indicates a taxon enriched in case participants with HNSCC and a blue branch indicates a taxon enriched in control participants. Thirteen differentially abundant species are highlighted and labeled on the outer ring. Oral bacterial periodontal pathogenic complexes are highlighted with yellow shading and labeled in the top left corner of the cladogram. The intensity of magenta shading corresponds to the relative abundance of bacterial species. A total of 320 species are included in the cladogram, representing taxa with at least 2 sequences in at least 10% of participants and with mean relative abundance of 0.001% or greater. The cladogram was created using GraPhlAn.

CPS-II and PLCO cohorts were predominantly White (>90%), while the SCCS included 73% African American participants.

Oral Bacterial Microbiome and Risk of HNSCC

HNSCC case participants did not differ significantly from matched control participants in oral bacterial α diversity, as measured by species richness and the Shannon Diversity Index, or in overall oral microbiome composition (β diversity), as measured by Jensen-Shannon divergence distance (eFigure 3 in Supplement 1).

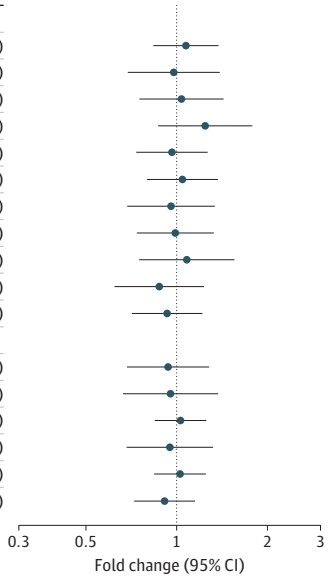
In the oral microbiome-wide meta-analysis of the 3 cohorts, we found 13 oral bacterial species associated with the risk of HNSCC development (**Figure 1**A; results for all bacteria identified are shown in eTable 4 in Supplement 1), verified in the permutation test (eTable 4 in Supplement 1). *Prevotella salivae*, *Streptococcus sanguinis*, and *Leptotrichia* species oral taxon 212 were related to lower HNSCC risk. In addition, 4 bacterial species in Proteobacteria were associated with lower risks: *Eikenella corrodens*, *Simonsiella muelleri*, *Rodentibacter pneumotropicus*, and *Pasteurella multocida*. These species belong to Betaproteobacteria and Gammaproteobacteria (**Figure 2**). Furthermore, a broader spectrum of Betaproteobacteria and Gammaproteobacteria tended to be associated with lower HNSCC risk. In addition, several rare species were associated with greater HNSCC risk (Figure 1A). The associations of these 13 bacterial species with HNSCC tended to be similar across the 3 cohorts (Figure 1B) and disease sites (oral cavity, pharynx, and larynx) (eFigure 4A in Supplement 1).

The oral periodontal pathogen red complex, orange complex, and combined red/orange complexes were moder-

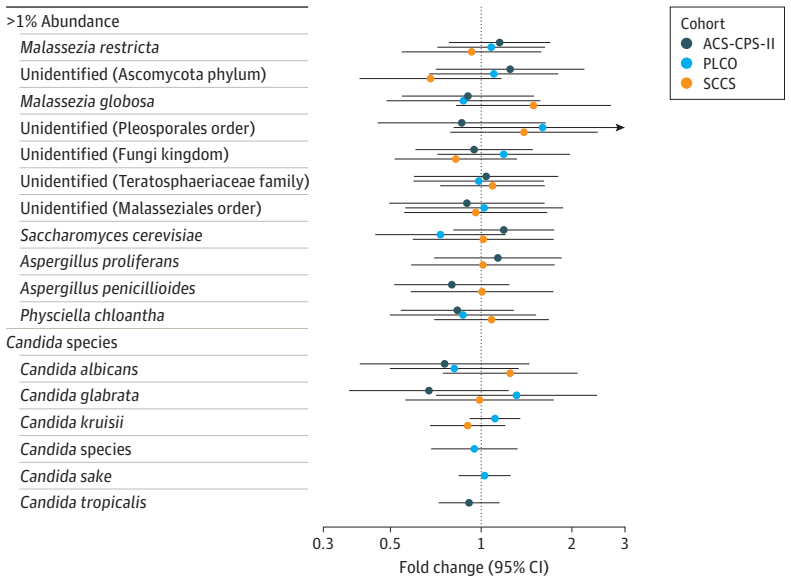
Figure 3. Oral Fungi and Risk of HNSCC

A Oral fungi and risk of HNSCC, meta-analysis

	Abundance, %	Fold change (95% CI)
>1% Abundance		
Malassezia restricta	20	1.07 (0.84-1.38)
Unidentified (Ascomycota phylum)	12	0.98 (0.69-1.39)
Malassezia globosa	4.9	1.04 (0.75-1.43)
Unidentified (Pleosporales order)	3.9	1.24 (0.87-1.78)
Unidentified (Fungi kingdom)	2.7	0.96 (0.73-1.27)
Unidentified (Teratosphaeriaceae family)	1.5	1.04 (0.80-1.37)
Unidentified (Malasseziales order)	1.5	0.96 (0.69-1.34)
Saccharomyces cerevisiae	1.4	0.99 (0.74-1.33)
Aspergillus proliferans	1.4	1.08 (0.75-1.55)
Aspergillus penicillioides	1.1	0.88 (0.62-1.23)
Physciella chloantha	1	0.93 (0.71-1.22)
Candida species		
Candida albicans	42	0.94 (0.68-1.28)
Candida glabrata	1.3	0.95 (0.66-1.37)
Candida kruisii	0.24	1.03 (0.85-1.25)
Candida species	0.22	0.95 (0.68-1.32)
Candida sake	0.018	1.03 (0.84-1.25)
Candida tropicalis	0.014	0.91 (0.72-1.15)



B Oral fungi and risk of HNSCC, by cohort



A, Fold changes and 95% CIs for the association between fungal taxa and risk of head and neck squamous cell cancer (HNSCC) from random-effects meta-analysis of analysis of compositions of microbiomes with bias correction abundance estimates of the 3 cohorts (American Cancer Society Cancer Prevention Study II Nutrition Cohort [ACS-CPS-II]; Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial [PLCO]; and Southern Community Cohort Study [SCCS]). Complete results for all fungi detected are shown in eTable 5 in Supplement 1. B, Fold changes and 95% CIs for the association between fungal taxa and risk of HNSCC for the 3 individual cohorts. All models were adjusted for age, sex, race and ethnicity, smoking status, number of cigarettes per day for those reporting ever smoking, alcohol consumption status, grams of ethanol consumed per day for those reporting alcohol consumption, and oral human papillomavirus 6 status as covariates. Fold change values greater than 1 indicate enrichment in HNSCC case participants, while values less than 1 indicate enrichment in control participants. The figure presents results for all *Candida* species and for other fungi of 1% or greater abundance. Relative abundance and fold changes with 95% CIs for fungal taxa are shown next to the plot.

ately associated with greater HNSCC risk (Figure 1C). The associations of these pathogen complexes with HNSCC tended to be consistent across the 3 cohorts (Figure 1D) and disease sites (oral cavity, pharynx and larynx) (eFigure 4B in Supplement 1).

Oral Fungal Microbiome and HNSCC

Overall fungal α and β diversity were not associated with HNSCC (eFigure 5 in Supplement 1). None of the fungal species were associated with HNSCC risk, including all *Candida* and relatively abundant species (>1%) (**Figure 3**), as well as all other less common species (eTables 5 and 6 in Supplement 1). These null findings were consistent across the 3 cohorts and disease sites (Figure 3; eTables 5 and 6 in Supplement 1). Given these null findings for fungi, we did not further evaluate bacterial-fungal interrelationships.

Microbial Risk Score and HNSCC

Analyzing the microbial profile as a community can characterize more microbial information than analyzing microbes indi-

Table 2. Microbial Risk Score and Risk of HNSCC

Microbial risk score[a]	Odds ratio (95% CI)				
	ACS-CPS-II	PLCO	SCCS	3-Cohort meta-analysis[b]	Monte Carlo cross-validation[c]
ANCOM-BC	1.53 (0.98-2.38)	1.12 (0.76-1.64)	1.31 (0.97-1.77)	1.29 (1.05-1.59)	1.12 (1.06-1.18)
Red complex	2.78 (1.24-6.21)	1.16 (0.79-1.69)	1.33 (1.02-1.73)	1.34 (1.09-1.65)	1.18 (1.08-1.27)
Orange complex	1.47 (0.81-2.67)	1.18 (0.78-1.77)	1.19 (0.91-1.56)	1.22 (0.99-1.51)	1.02 (0.96-1.07)
Red/orange complex[d]	3.01 (1.32-6.85)	1.20 (0.80-1.80)	1.39 (1.05-1.82)	1.40 (1.13-1.75)	1.13 (1.05-1.22)
ANCOM-BC and red/orange complex	2.13 (1.29-3.52)	1.23 (0.82-1.84)	1.47 (1.11-1.96)	1.50 (1.21-1.85)	1.16 (1.08-1.24)

Abbreviations: ACS-CPS-II, American Cancer Society Cancer Prevention Study II Nutrition Cohort; ANCOM-BC, analysis of compositions of microbiomes with bias correction; HNSCC, head and neck squamous cell cancer; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SCCS, Southern Community Cohort Study.

[a] The microbial risk score was calculated as a weighted sum of the relative abundance of the bacterial species listed below, with weights assigned according to their effect sizes as determined by ANCOM-BC.

[b] Meta-analysis estimates were obtained from random-effects meta-analyses of cohort-specific logistic regression estimates. For each cohort and 3-cohort meta-analysis, ANCOM-BC refers to the 13 bacterial species associated with HNSCC, as identified in the meta-analysis of the 3 cohorts' ANCOM-BC estimates at α = .05. These species were *Eikenella corrodens*, *Lactobacillus paracollinoides*, *Leptotrichia* species oral taxon 212, *Porphyromonas cangingivalis*, *Prevotella salivae*, *Prevotella* species HUN102, *Pyramidobacter piscolens*, *Rodentibacter pneumotropicus*, *Simonsiella muelleri*, *Streptococcus gallolyticus*, *Pasteurella multocida*, *Streptococcus plurextorum*, and *Streptococcus sanguinis*.

[c] For the Monte Carlo cross-validation, data were split into training and test sets at a 70% to 30% ratio, with this process repeated 50 times to ensure robustness. ANCOM-BC for Monte Carlo cross-validation refers to the same 13 bacterial species associated with HNSCC, as verified in the permutation test ($Q < 0.05$) (eTable 4 in Supplement 1).

[d] For the red/orange complexes, the red complex includes *Treponema denticola*, *Porphyromonas gingivalis*, and *Tannerella forsythia* and the orange complex includes *Fusobacterium nucleatum*, *Prevotella intermedia*, *Prevotella nigrescens*, *Eubacterium nodatum*, *Campylobacter showae*, and *Campylobacter gracilis*. ANCOM-BC and logistic regression were conducted for HNSCC after adjusting for age, sex, race and ethnicity, smoking status, number of cigarettes per day for those reporting ever smoking, alcohol consumption status, grams of ethanol consumed per day for those reporting alcohol consumption, and oral human papillomavirus 16 status. Odds ratios for the microbial risk score represent changes per 1 SD in the microbial risk score as a continuous predictor.

vidually. We constructed a summary microbial risk score consisting of 13 bacteria associated with HNSCC (Figure 1A) and 9 species in the red and orange periodontal pathogen complexes (**Table 2**). In meta-analyses of the 3 cohorts, increased risk of HNSCC was observed for the 13 bacteria, for the orange and red complex bacteria combined, and for the combination of all 22 risk-associated bacteria. These associations did not differ substantially by study cohort. The microbial risk score validated by Monte Carlo cross-validation showed comparable risks per 1-unit increase in microbial risk score, as was found in the 3-cohort meta-analysis (Table 2; eTable 7 in Supplement 1). Further examination of the association of the combined microbial risk score for 22 bacteria with HNSCC did not reveal substantial variability with respect to disease site (oral cavity, pharynx, or larynx) or in relation to tobacco and alcohol use, HPV-16 status, or follow-up period (eTable 8 in Supplement 1).

Discussion

In a metagenomic survey from 3 large, well-established cohorts, we identified 13 previously unrecognized oral bacterial species that are associated with subsequent HNSCC risk. We also found that the red and orange oral bacterial pathogen complexes were moderately associated with greater HNSCC risk. We summarized these findings in a microbial risk score for the 3 cohorts, identifying a 50% increase in the risk of HNSCC per 1-SD increase in the microbial risk score. A microbial risk score in the Monte Carlo cross-validation showed comparable risks per 1-unit increase in microbial risk score, as was found in the meta-analysis of the 3 cohorts. Oral fungal taxa were not associated with HNSCC risk in our study.

The oral microbiome is an intricate ecosystem that thrives in the dynamic oral environment through a symbiotic relationship with human hosts.[51] Our findings potentially reflect a subtle dysbiosis, years before clinical diagnosis in the oral microbiome, marked by the depletion of commensal organisms and enrichment of oral pathogens, leading to significant increases in HNSCC risk. The identified bacteria and bacterial complexes associated with HNSCC hold promise as potential biomarkers, along with other risk factors, for identifying high-risk individuals for personalized prevention of HNSCC.

The red and orange oral pathogen complexes are associated with periodontal disease,[27] while periodontal disease and other indicators of poor oral health are putative risk factors for head and neck cancer (as reviewed by Gopinath et al[10]). While others identified red complex bacteria in HNSCC tumor tissue,[52-54] the evidence from our prospective study strengthens the hypothesis that oral health status is causally related to HNSCC development and, furthermore, indicates that well-characterized red and orange oral bacterial complexes may be involved early in head and neck carcinogenesis, before HNSCC is overt.

We identified previously unrecognized commensal bacterial species in *Proteobacteria*, including *E corrodens*, *S muelleri*, *R pneumotropicus*, and *P multocida*, as associated with lower HNSCC risk. These findings align with our earlier observations using 16S ribosomal RNA amplicon gene sequencing[14] that Proteobacteria, at the genera and class levels, are associated with lower HNSCC risk, potentially related to oral carcinogen metabolism.[8] However, in this study, we have pinpointed specific species and strains associated with HNSCC. The other newly identified oral species have received less attention in HNSCC research, although *S sanguinis* and *E corrodens* are known to be associated with a healthy oral state.[27,28] Conducting further

functional investigations into the bacteria associated with HNSCC risk, as identified in our metagenomic survey, is imperative for deepening understanding of their role in HNSCC.

We did not observe significant relationships of fungal species with subsequent HNSCC risk. This contrasts with findings from animal models showing that fungi can activate carcinogens including alcohol, promote inflammation, and contribute to neoplasia,[16-20] and contrasts with studies in patients with HNSCC showing high abundance of *Candida albicans* and other pathogenic fungi in their oral samples[21] and tumor tissue.[22,23] It is plausible that oral fungi, as opportunistic pathogens, may exert an influence on HNSCC later in head and neck cancer development, beyond the time frame of our prediagnostic sample collections. Alternatively, technical limitations in assessing fungi[55] could potentially explain our results, although the approach was validated and used previously[56-58] and our quality control measures appeared reasonably robust.

The strengths of this study include (1) the prospective study design; (2) bacterial whole-genome sequencing and fungal internal transcribed spacer sequencing; (3) a substantial sample size, allowing for exploration (albeit with limited power) of HNSCC subsites (oral cavity, pharynx, and larynx) and interplay with alcohol consumption and smoking; and (4) a racially diverse study population, including 36% African American participants. This study is innovative because it is the first comprehensive and prospective study of the oral bacterial and fungal microbiomes in relation to HNSCC risk.

Limitations

There are limitations to our study. First, because it was an observational study, there is potential for confounding by unmeasured factors at study entry and that may develop over the course of follow-up. Second, selection bias due to loss to follow-up may also have impacted risk estimates,[59] although the 3 cohorts limited these losses through active and passive follow-up activities.[30-32] Third, this study had limited mechanistic information on the causal relationship from an observational study, warranting further testing of our observations in experimental systems and human intervention studies.[60,61] Fourth, oral bacteria and fungi commonly organize as complex biofilms,[28] whereas this study used oral wash samples that may incompletely capture these relationships. Fifth, in this study, HPV-16–negative participants predominated, so interrelationships of bacteria and fungi with carriage of HPV-16 will need to be examined in further studies.

Conclusions

This case-control study yielded compelling evidence that oral bacteria are a risk factor for HNSCC development. The identified bacteria and bacterial complexes and the associated microbial risk score hold promise as potential biomarkers, along with other risk factors, to identify high-risk individuals for personalized prevention of HNSCC.

ARTICLE INFORMATION

Accepted for Publication: May 21, 2024.

Published Online: September 26, 2024. doi:10.1001/jamaoncol.2024.4006

Author Contributions: Drs Ahn and Hayes had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Kwak, Ahn, Hayes.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Kwak, Ahn, Hayes.
Critical review of the manuscript for important intellectual content: All authors.
Statistical analysis: Kwak, Wang, Li, Ahn, Hayes.
Obtained funding: Shrubsole, Ahn, Hayes.
Administrative, technical, or material support: Wu, Freedman, Huang, McCullough, Um, Shrubsole, Cai, Ahn.
Supervision: Ahn, Hayes.

Conflict of Interest Disclosures: Dr Hayes reported receipt of grants from the National Institutes of Health outside the submitted work. No other disclosures were reported.

Funding/Support: The research reported herein was supported in part by the National Cancer Institute under awards P20CA252728, R01CA159036, and U01CA250186 and by the National Library of Medicine under grant 1R01LM014085. Samples were sequenced at the NYU School of Medicine Genome Technology Center. The Genome Technology Center is partially supported by Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study II cohort.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

Additional Contributions: We acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries and cancer registries supported by the National Cancer Institute's Surveillance, Epidemiology, and End Results program. We express sincere appreciation to all Cancer Prevention Study II (and/or Cancer Prevention Study 3) participants and to each member of the study and biospecimen management group.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660

2. Wyss A, Hashibe M, Chuang SC, et al. Cigarette, cigar, and pipe smoking and the risk of head and neck cancers: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Am J Epidemiol*. 2013;178(5):679-690. doi:10.1093/aje/kwt029

3. Purdue MP, Hashibe M, Berthiller J, et al. Type of alcoholic beverage and risk of head and neck cancer—a pooled analysis within the INHANCE Consortium. *Am J Epidemiol*. 2009;169(2):132-142. doi:10.1093/aje/kwn306

4. Agalliu I, Gapstur S, Chen Z, et al. Associations of oral α-, β-, and γ-human papillomavirus types with risk of incident head and neck cancer. *JAMA Oncol*. 2016;2(5):599-606. doi:10.1001/jamaoncol.2015.5504

5. Pignatelli P, Romei FM, Bondi D, Giuliani M, Piattelli A, Curia MC. Microbiota and oral cancer as a complex and dynamic microenvironment: a narrative review from etiology to prognosis. *Int J Mol Sci*. 2022;23(15):8323. doi:10.3390/ijms23158323

6. Ganly I, Yang L, Giese RA, et al. Periodontal pathogens are a risk factor of oral cavity squamous cell carcinoma, independent of tobacco and alcohol and human papillomavirus. *Int J Cancer*. 2019;145(3):775-784. doi:10.1002/ijc.32152

7. Wu Z, Han Y, Wan Y, et al. Oral microbiome and risk of incident head and neck cancer: a nested case-control study. *Oral Oncol*. 2023;137:106305. doi:10.1016/j.oraloncology.2022.106305

8. Wu J, Peters BA, Dominianni C, et al. Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J*. 2016;10(10):2435-2446. doi:10.1038/ismej.2016.37

9. Michaud DS, Fu Z, Shi J, Chung M. Periodontal disease, tooth loss, and cancer risk. *Epidemiol Rev*. 2017;39(1):49-58. doi:10.1093/epirev/mxx006

10. Gopinath D, Kunnath Menon R, Veettil SK, Botelho MG, Johnson NW. Periodontal diseases as putative risk factors for head and neck cancer: systematic review and meta-analysis. *Cancers (Basel)*. 2020;12(7):1893. doi:10.3390/cancers12071893

11. Kang EJ, Moon SJ, Lee K, Park IH, Kim JS, Choi YJ. Associations between missing teeth and the risk of cancer in Korea: a nationwide cohort study. *BMC Oral Health*. 2023;23(1):418. doi:10.1186/s12903-023-02997-x

12. Dewhirst FE, Chen T, Izard J, et al. The human oral microbiome. *J Bacteriol*. 2010;192(19):5002-5017. doi:10.1128/JB.00542-10

13. Ting HSL, Chen Z, Chan JYK. Systematic review on oral microbial dysbiosis and its clinical associations with head and neck squamous cell carcinoma. *Head Neck*. 2023;45(8):2120-2135. doi:10.1002/hed.27422

14. Hayes RB, Ahn J, Fan X, et al. Association of oral microbiome with risk for incident head and neck squamous cell cancer. *JAMA Oncol*. 2018;4(3):358-365. doi:10.1001/jamaoncol.2017.4777

15. Ghannoum MA, Jurevic RJ, Mukherjee PK, et al. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog*. 2010;6(1):e1000713. doi:10.1371/journal.ppat.1000713

16. O'Grady JF, Reade PC. *Candida albicans* as a promoter of oral mucosal neoplasia. *Carcinogenesis*. 1992;13(5):783-786. doi:10.1093/carcin/13.5.783

17. Sohrabi N, Hassan ZM, Khosravi AR, et al. Invasive aspergillosis promotes tumor growth and severity in a tumor-bearing mouse model. *Can J Microbiol*. 2010;56(9):771-776. doi:10.1139/W10-064

18. Daley D, Mani VR, Mohan N, et al. Dectin 1 activation on macrophages by galectin 9 promotes pancreatic carcinoma and peritumoral immune tolerance. *Nat Med*. 2017;23(5):556-567. doi:10.1038/nm.4314

19. Ramirez-Garcia A, Rementeria A, Aguirre-Urizar JM, et al. *Candida albicans* and cancer: can this yeast induce cancer development or progression? *Crit Rev Microbiol*. 2016;42(2):181-193. doi:10.3109/1040841X.2014.913004

20. El-Jurdi N, Ghannoum MA. The mycobiome: impact on health and disease states. *Microbiol Spectr*. 2017;5(3). doi:10.1128/microbiolspec.FUNK-0045-2016

21. Alnuaimi AD, Wiesenfeld D, O'Brien-Simpson NM, Reynolds EC, McCullough MJ. Oral *Candida* colonization in oral cancer patients and its relationship with traditional risk factors of oral cancer: a matched case-control study. *Oral Oncol*. 2015;51(2):139-145. doi:10.1016/j.oraloncology.2014.11.008

22. Mukherjee PK, Wang H, Retuerto M, et al. Bacteriome and mycobiome associations in oral tongue cancer. *Oncotarget*. 2017;8(57):97273-97289. doi:10.18632/oncotarget.21921

23. Perera M, Al-Hebshi NN, Perera I, et al. A dysbiotic mycobiome dominated by *Candida albicans* is identified within oral squamous-cell carcinomas. *J Oral Microbiol*. 2017;9(1):1385369. doi:10.1080/20002297.2017.1385369

24. Shay E, Sangwan N, Padmanabhan R, Lundy S, Burkey B, Eng C. Bacteriome and mycobiome and bacteriome-mycobiome interactions in head and neck squamous cell carcinoma. *Oncotarget*. 2020;11(25):2375-2386. doi:10.18632/oncotarget.27629

25. Diaz PI, Strausbaugh LD, Dongari-Bagtzoglou A. Fungal-bacterial interactions and their relevance to oral health: linking the clinic and the bench. *Front Cell Infect Microbiol*. 2014;4:101. doi:10.3389/fcimb.2014.00101

26. Lof M, Janus MM, Krom BP. Metabolic interactions between bacteria and fungi in commensal oral biofilms. *J Fungi (Basel)*. 2017;3(3):40. doi:10.3390/jof3030040

27. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL Jr. Microbial complexes in subgingival plaque. *J Clin Periodontol*. 1998;25(2):134-144. doi:10.1111/j.1600-051X.1998.tb02419.x

28. Abdulkareem AA, Al-Taweel FB, Al-Sharqi AJB, Gul SS, Sha A, Chapple ILC. Current concepts in the pathogenesis of periodontitis: from symbiosis to dysbiosis. *J Oral Microbiol*. 2023;15(1):2197779. doi:10.1080/20002297.2023.2197779

29. Wang C, Segal LN, Hu J, et al. Microbial risk score for capturing microbial characteristics, integrating multi-omics data, and predicting disease risk. *Microbiome*. 2022;10(1):121. doi:10.1186/s40168-022-01310-2

30. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer*. 2002;94(2):500-511. doi:10.1002/cncr.10197

31. Hayes RB, Reding D, Kopp W, et al; Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Project Team. Etiologic and early marker studies in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials*. 2000;21(6)(suppl):349S-355S. doi:10.1016/S0197-2456(00)00101-X

32. Signorello LB, Hargreaves MK, Blot WJ. The Southern Community Cohort Study: investigating health disparities. *J Health Care Poor Underserved*. 2010;21(1)(suppl):26-37. doi:10.1353/hpu.0.0245

33. Belstrøm D, Holmstrup P, Bardow A, Kokaras A, Fiehn NE, Paster BJ. Temporal stability of the salivary microbiota in oral health. *PLoS One*. 2016;11(1):e0147472. doi:10.1371/journal.pone.0147472

34. Zhou Y, Gao H, Mihindukulasuriya KA, et al. Biogeography of the ecosystems of the healthy human body. *Genome Biol*. 2013;14(1):R1. doi:10.1186/gb-2013-14-1-r1

35. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. *Science*. 2009;326(5960):1694-1697. doi:10.1126/science.1177486

36. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Knight R, Knights D. SHOGUN: a modular, accurate and scalable framework for microbiome quantification. *Bioinformatics*. 2020;36(13):4088-4090. doi:10.1093/bioinformatics/btaa277

37. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, et al. Evaluating the information content of shallow shotgun metagenomics. *mSystems*. 2018;3(6):e00069-18. doi:10.1128/mSystems.00069-18

38. Usyk M, Zolnik CP, Patel H, Levi MH, Burk RD. Novel ITS1 fungal primers for characterization of the mycobiome. *mSphere*. 2017;2(6):e00488-17. doi:10.1128/mSphere.00488-17

39. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-583. doi:10.1038/nmeth.3869

40. Abarenkov K, Henrik Nilsson R, Larsson KH, et al. The UNITE database for molecular identification of fungi—recent updates and future perspectives. *New Phytol*. 2010;186(2):281-285. doi:10.1111/j.1469-8137.2009.03160.x

41. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584. doi:10.7717/peerj.2584

42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2

43. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17(4):571-582. doi:10.1080/10543400701329422

44. Fuglede B, Topsoe F. Jensen-Shannon Divergence and Hilbert Space Embedding. In: *Proceedings of the International Symposium on Information Theory*; 2004:31. doi:10.1109/ISIT.2004.1365067

45. Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online*; November 15, 2017. doi:10.1002/9781118445112.stat07841

46. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nat Commun*. 2020;11(1):3514. doi:10.1038/s41467-020-17041-7

47. Hedges LV. A random effects model for effect sizes. *Psychol Bull*. 1983;93(2):388-395. doi:10.1037/0033-2909.93.2.388

48. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ*. 2015;3:e1029. doi:10.7717/peerj.1029

49. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*. 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1

50. Shan G. Monte Carlo cross-validation for a study with binary outcome and limited sample size. *BMC Med Inform Decis Mak*. 2022;22(1):270. doi:10.1186/s12911-022-02016-z

51. Radaic A, Kapila YL. The oralome and its dysbiosis: new insights into oral microbiome-host interactions. *Comput Struct Biotechnol J*. 2021;19:1335-1360. doi:10.1016/j.csbj.2021.02.010

52. Shin JM, Luo T, Kamarajan P, Fenno JC, Rickard AH, Kapila YL. Microbial communities associated with primary and metastatic head and neck squamous cell carcinoma—a high fusobacterial and low streptococcal signature. *Sci Rep*. 2017;7(1):9934. doi:10.1038/s41598-017-09786-x

53. Fan Z, Tang P, Li C, et al. *Fusobacterium nucleatum* and its associated systemic diseases: epidemiologic studies and possible mechanisms. *J Oral Microbiol*. 2022;15(1):2145729. doi:10.1080/20002297.2022.2145729

54. Lamont RJ, Fitzsimonds ZR, Wang H, Gao S. Role of *Porphyromonas gingivalis* in oral and

orodigestive squamous cell carcinoma. *Periodontol 2000*. 2022;89(1):154-165. doi:10.1111/prd.12425

55. Mbareche H, Veillette M, Bilodeau G, Duchaine C. Comparison of the performance of ITS1 and ITS2 as barcodes in amplicon-based sequencing of bioaerosols. *PeerJ*. 2020;8:e8523. doi:10.7717/peerj.8523

56. Rosenbaum J, Usyk M, Chen Z, et al. Evaluation of oral cavity DNA extraction methods on bacterial and fungal microbiota. *Sci Rep*. 2019;9(1):1531. doi:10.1038/s41598-018-38049-6

57. Usyk M, Schlecht NF, Pickering S, et al; Costa Rica HPV Vaccine Trial (CVT) Group. molBV reveals immune landscape of bacterial vaginosis and predicts human papillomavirus infection natural history. *Nat Commun*. 2022;13(1):233. doi:10.1038/s41467-021-27628-3

58. Usyk M, Zolnik CP, Castle PE, et al; Costa Rica HPV Vaccine Trial (CVT) Group. Cervicovaginal microbiome and natural history of HPV in a longitudinal study. *PLoS Pathog*. 2020;16(3):e1008376. doi:10.1371/journal.ppat.1008376

59. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology*. 2016;27(1):91-97. doi:10.1097/EDE.0000000000000409

60. Homayouni Rad A, Pourjafar H, Mirzakhani E. A comprehensive review of the application of probiotics and postbiotics in oral health. *Front Cell Infect Microbiol*. 2023;13:1120995. doi:10.3389/fcimb.2023.1120995

61. Siddiqui R, Badran Z, Boghossian A, Alharbi AM, Alfahemi H, Khan NA. The increasing importance of the oral microbiome in periodontal health and disease. *Future Sci OA*. 2023;9(8):FSO856. doi:10.2144/fsoa-2023-0062