8-25-2021

Association of oral microbiota with lung cancer risk in a low-income population in the Southeastern USA

Jiajun Shi
Vanderbilt University

Yaohua Yang
Vanderbilt University

Hua Xie
Meharry Medical College

Xiaofei Wang
Tennessee State University

Jie Wu
Vanderbilt University

See next page for additional authors

Recommended Citation

Shi, J., Yang, Y., Xie, H. et al. Association of oral microbiota with lung cancer risk in a low-income population in the Southeastern USA. Cancer Causes Control 32, 1423–1432 (2021). https://doi.org/10.1007/s10552-021-01490-6

Cancer Causes & Control (2021) 32:1423–1432
https://doi.org/10.1007/s10552-021-01490-6

ORIGINAL PAPER

Association of oral microbiota with lung cancer risk in a low-income population in the Southeastern USA

Jiajun Shi[1] · Yaohua Yang[1] · Hua Xie[2] · Xiaofei Wang[3] · Jie Wu[1] · Jirong Long[1] · Regina Courtney[1] · Xiao-Ou Shu[1] · Wei Zheng[1] · William J. Blot[1] · Qiuyin Cai[1]

Received: 11 March 2021 / Accepted: 17 August 2021 / Published online: 25 August 2021

Abstract

Purpose Oral microbiome plays an important role in oral health and systemic diseases, including cancer. We aimed to prospectively investigate the association of oral microbiome with lung cancer risk.

Methods We analyzed 156 incident lung cancer cases (73 European Americans and 83 African Americans) and 156 individually matched controls nested within the Southern Community Cohort Study. Oral microbiota were assessed using 16S *rRNA* gene sequencing in pre-diagnostic mouth rinse samples. Paired t test and the permutational multivariate analysis of variance test were used to evaluate lung cancer risk association with alpha diversity or beta diversity, respectively. Conditional logistic regression models were used to evaluate the association of individual bacterial abundance or prevalence with lung cancer risk.

Results No significant differences were observed for alpha or beta diversity between lung cancer cases and controls. Abundance of families *Lachnospiraceae_[XIV]*, *Peptostreptococcaceae_[XI]*, and *Erysipelotrichaceae* and species *Parvimonas micra* was associated with decreased lung cancer risk, with odds ratios (ORs) and 95% confidence intervals (CIs) of 0.76 (0.59–0.98), 0.80 (0.66–0.97), 0.81 (0.67–0.99), and 0.83 (0.71–0.98), respectively (all $p < 0.05$). Prevalence of five predefined oral pathogens were not significantly associated with overall lung cancer risk. Prevalence of genus *Bacteroidetes_[G-5]* and species *Alloprevotella sp._oral_taxon_912*, *Capnocytophaga sputigena*, *Lactococcus lactis*, *Peptoniphilaceae_[G-1] sp._oral_taxon_113*, *Leptotrichia sp._oral_taxon_225*, and *Fretibacterium fastidiosum* was associated with decreased lung cancer risk, with ORs and 95% CIs of 0.55 (0.30–1.00), 0.36 (0.17–0.73), 0.53 (0.31–0.92), 0.43 (0.21–0.88), 0.43 (0.19–0.94), 0.57 (0.34–0.99), and 0.54 (0.31–0.94), respectively (all $p < 0.05$). Species *L. sp._oral_taxon_225* was significantly associated with decreased lung cancer risk in African Americans (OR [95% CIs] 0.28 [0.12–0.66]; $p = 0.00012$).

Conclusion Results from this study suggest that oral microbiota may play a role in the development of lung cancer.

Keywords Oral microbiome · Lung cancer · 16S *rRNA* sequencing · Low-income population

Introduction

Lung cancer is the most common cause of cancer mortality and one of the most common cancers in both men and women worldwide [1, 2]. Cigarette smoking is attributed to approximately 80% of the lung cancer burden in men and more than 50% in women [3, 4]. Other lung cancer risk factors include alcohol drinking, unhealthy diet, physical inactivity, history of lung diseases, and environmental and occupational exposures. However, there is a need to better understand alternative risk factors, particularly with regard to how they interact with known risk factors as well as the underlying mechanisms of carcinogenesis.

Oral microbiome plays an important role in oral health and systemic diseases, including cancer [5–7]. Dysbiosis of

✉ Qiuyin Cai
qiuyin.cai@vanderbilt.edu

[1] Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, 1161 21st Avenue South, Nashville, TN, USA

[2] School of Dentistry, Meharry Medical College, Nashville, TN, USA

[3] Department of Biological Sciences, Tennessee State University, Nashville, TN 37232, USA

oral microbiota has been proposed to contribute to cancer development via various mechanisms, including inducing chronic inflammation, inhibiting apoptosis, activating cell proliferation, promoting cellular invasion, and/or producing carcinogens [8, 9]. Regarding lung cancer, several lines of observational findings have suggested a possible association between oral microbiota and risk of the disease. First, an increased risk for lung cancer has been observed in patients with periodontal diseases [10–12], which is likely caused by oral microbial pathogens and dysbiotic oral microbiota [13]. Second, individuals carrying specific oral bacteria, such as *Chlamydia pneumoniae*, have an increased risk for lung cancer [14–16]. Third, differential abundance of certain oral bacteria were reported to be associated with lung cancer risk in case–control studies [17–20]. However, no study has prospectively investigated the association of oral microbiome with lung cancer risk in low-income populations, especially among African Americans (AAs).

In the present investigation, we conducted a case–control study nested within the Southern Community Cohort Study (SCCS), a low-income population living in the Southeastern USA, to systematically evaluate the associations between oral microbiota and lung cancer risk among AAs and European Americans (EAs).

Materials and methods

Study participants and data collection

The SCCS is an ongoing prospective cohort study, with the aims to investigate risk factors of cancer and chronic diseases among a low-income population, described in detail elsewhere [21]. Briefly, approximately 86,000 adults, two-thirds of whom were AAs, were recruited between March 2002 and September 2009 from 12 Southeastern states of the USA. Approximately 86% were recruited from community health centers (CHCs), institutions providing basic health care and preventative services in underserved areas, so that the cohort included a large number of individuals of low income and educational status. The remaining 14% of cohort members were recruited through mail-based general population sampling. During enrollment, mouth rinse samples were collected from ~ 34,100 participants [22]. The SCCS was reviewed and approved by the institutional review boards at Vanderbilt University and Meharry Medical College. Written informed consent was obtained from all study participants.

In the baseline survey, participants completed a comprehensive questionnaire, used to collect information on anthropometric characteristics, lifestyle factors, disease history, medication use, and other characteristics. Passive cohort follow-up by record linkage to state cancer registries operating in the 12-state study area and the national death index registry started immediately upon completion of the baseline survey. Active follow-up surveys started in 2008. In mailed or telephone follow-up surveys, participants were asked about their personal medical histories and medication use. Cigarette smokers were defined as those who reported smoking at least 100 cigarettes in their lifetime. Former smokers were defined as those who had quit smoking for at least one year, whereas those who had quit for less than one year were considered current smokers. Lung cancer cases were defined according to the International Classification of Diseases (ICD-10) for Oncology, Second Edition (ICD-O-2) and included all invasive cancers coded as C34.0–C34.9.

For this study, we selected 166 incident lung cancer cases and 166 individually matched controls who donated mouth rinse samples at enrollment. For each incident lung cancer case, one control was randomly selected and individually matched to cases by age of enrollment (±2 years), race, sex, smoking status at baseline (never/former/current smoker), date of mouth rinse sample collection (±90 days), CHC recruitment site, and recruitment source. We excluded six participants who used antibiotics ten or more times during the prior year or took any dose of antibiotics during the past week before mouth rinse sample collection. Participants with a history of HIV infection, other cancers, diabetes, stroke, or myocardial infarction were also excluded.

During enrollment, mouth rinse samples were collected from participants, who were asked to swish vigorously for 45 s with 10 mL Scope mouthwash containing a 15 wt% alcohol content (Procter & Gamble, Cincinnati, OH) and then expectorate into a specimen container. Mouth rinse samples were shipped to the Molecular Epidemiology Laboratory at Vanderbilt University Medical Center for processing and stored at −80 °C until DNA extraction. The mean time between mouth rinse sample collection and lung cancer diagnosis is 3.5 years.

DNA extraction and 16S *rRNA* gene sequencing

Total DNA, including bacterial DNA, was isolated from mouth rinse samples using the QIAmp DNA kit (Qiagen, Germantown, MD, USA). Sequencing libraries were prepared using the NEXTflex 16S V4 Amplicon-Seq Kit (Bioo Scientific 4201-05), following the manufacturer's protocol. This kit was designed to sequence approximately 253 bp of the fourth hypervariable (V4) domain of the 16S *rRNA* gene [23, 24]. Pair-end reads of 250 bp were obtained using the Illumina HiSeq System. Each 96-well plate was sequenced with one negative control (i.e., distilled water) and two duplicate quality control (QC) samples. In this study, each of the two discrete QC samples comprising of mixed mouth rinse DNA samples of de-identified volunteers was sequenced four times. Comparable microbial profiles were

observed for the same QC samples. For example, for alpha diversity measurements within each sample, using Shannon and phylogenetic diversity (PD) whole tree indexes, the average coefficients of variability among repeated QC samples were 2.0% and 4.8%, respectively. For individual species-level taxa, Pearson correlation coefficients among QC samples ranged from 97.9% to 99.9%, with a median of 99.6%.

Sequence data analysis and quality control

Raw sequencing reads, ranging from 36,323 to 351,766 (with a mean of 114,420 and standard deviation of 37,394) among the participants, were trimmed and filtered to remove bases and reads of low quality using the Sickle tool [25]. An average of 9,827 reads (standard deviation = 10,831) were filtered. Then, BayesHammer [26] was utilized for correcting sequencing errors and PANDAseq [27] for stitching paired-end reads [28]. Clean reads were then clustered into Operational Taxonomic Units (OTUs) at 97% sequence identity, using the closed reference OTU picking strategy with the Human Oral Microbiome Database (HOMD) [29] as reference via the Quantitative Insights Into Microbial Ecology (QIIME), v1.9.1 [23]. Four participants with clean sequencing reads < 20,000 were removed. We excluded ten matched case–control pairs from which any paired participant either used antibiotics ($n = 6$) or had low sequencing reads ($n = 4$). A total of 156 lung cancer case–control pairs were included in the final analysis.

Statistical analyses

The QIIME version 1.9.1 [23] was used to rarefy the OTU/species table at a sequencing depth of 20,000 and estimate observed bacterial OTUs and alpha diversity indices including Chao1, Shannon, and PD whole tree. Paired *t* test was used to compare bacterial richness and alpha diversity between lung cancer cases and matched controls. Beta diversity (the total variance of an oral microbial composition) between lung cancer cases and controls was compared by the Permutational Multivariate Analysis of Variance (PERMANOVA-S) test implemented in the miProfile software [30]. This software can construct and produce *p*-values for each of the six beta diversity matrices (Jaccard, Bray–Curtis, presence-weighted UniFrac with parameter 1 or 0, unweighted UniFrac, and weighted UniFrac) and a unified *p*-value by combining all abundance and presence-absence distances [30], adjusting for unmatched covariates including smoking pack-years, alcohol intake (none, light, moderate, or heavy; by sex), total energy intake, body mass index (BMI), last time of dentistry visit, and sequencing batch.

Associations between individual bacterial taxa at phylum, family, genus, and OTU levels and lung cancer risk were evaluated by two approaches based on taxonomic relative abundance and prevalence (carriage frequency). First, taxa with a median relative abundance > 0.01% among control participants were selected for abundance association with lung cancer risk. There were 119 such taxa, including six phyla, 23 families, 33 genera, and 57 species. At each taxonomic level, the sequencing read counts for each taxon were normalized using centered log-ratio transformation after adding 1 as a pseudo-count [31]. Conditional logistic regression was conducted in the stratum for matched cases and controls, adjusting for the above-mentioned unmatched covariates for abundance associations. Second, taxa with a median relative abundance ≤ 0.01% and a prevalence of > 20% among control participants were selected for prevalence association with lung cancer risk. For 219 such taxa, including three phyla, 19 families, 50 genera, and 147 species (total $n = 219$), conditional logistic regression models were used to compare the prevalence of carriers of each taxon among lung cancer cases and controls, adjusted for the above-mentioned covariates and sequencing depth. In addition, taxonomic prevalence in associations with lung cancer risk were also tested for five pre-defined oral pathogens, including *P. gingivalis*, *Treponema denticola*, *Tannerella forsythia*, *A. actinomycetemcomitans*, and *Prevotella intermedia* [32, 33].

Stratified analyses by race, sex, smoking status (ever/never smoking), and time between mouth rinse sample collection and lung cancer diagnosis (<2 years: $n = 94$; >2 years, $n = 62$) were conducted for both abundance and prevalence associations. Adjusting for diet quality score (based on the healthy eating index 2010 edition [34]) rather than total energy intake in above logistic regression models did not materially change the results. Thus, results without adjustment for healthy eating index are reported. We did not collect detailed oral health information at baseline enrollment. We used time since last dentistry visit as a proxy variable for oral health, which was adjusted in the regression models for association analyses between oral microbiota and lung cancer risk.

Considering the fact that taxa of different taxonomic ranks are correlated, we used Galwey's method [35], implemented in the function "meff" of the R package "poolR" [36], to estimate the effective number of independent tests, used for Bonferroni correction. All of the analyses were carried out using SAS statistical software (SAS Enterprise Guide 7.1, SAS Institute Inc.) or R version 3.4.3 unless otherwise indicated.

Results

Characteristics of the study subjects

Table 1 presents the distribution of selected demographic characteristics of the participants. Lung cancer cases and

controls were well matched with respect to age, race, sex, and smoking status. Compared with controls, ever-smoking participants with lung cancer had significantly higher pack-years ($p = 7.1 \times 10^{-4}$). No significant differences were observed in the distributions of alcohol drinking status, family history of lung cancer, dental visit, healthy eating index, or BMI between case and control groups.

Association of overall oral microbiome composition with lung cancer risk

Observed OTUs among lung cancer cases and controls are summarized in Fig. 1A, separately for EA cases, EA controls, AA cases, and AA controls. Among 708 OTUs, 534 were observed in all four groups, while 32, 29, 30, and 27 were unique among EA cases, EA controls, AA cases, and AA controls, respectively. Compared with controls, lung cancer cases had lower observed OTUs (a mean ± standard deviation of 172.0 ± 54.9 vs. 181.3 ± 56.5; $p = 0.09$), Chao1 (195.2 ± 57.7 vs. 204.8 ± 58.2; $p = 0.08$), PD whole tree (8.5 ± 2.4 vs. 8.9 ± 2.4; $p = 0.06$), and Shannon (3.8 ± 0.8 vs. 3.9 ± 0.6; $p = 0.31$), respectively (Fig. 1B). No significant differences were observed in any of the six distance matrices of beta diversity between lung cancer cases and controls (all $p > 0.1$). Results of principal coordinate analysis (PCoA) of Bray–Curtis dissimilarity are presented in Fig. 1C.

Associations of oral bacterial taxa with lung cancer risk

We evaluated the differences of abundance for 119 oral taxa between lung cancer cases and controls. Six taxa, all of them are within phylum *Firmicute*, were associated with reduced lung cancer risk at a $p < 0.05$ (Table 2). Family *Peptoniphilaceae* and its highly correlated descendants, genus *Parvimonas*, and species *P. micra* (Pearson correlation coefficient > 0.99) were associated with reduced lung cancer risk, with similar ORs (0.82–0.83). Three other families, including *Lachnospiraceae_[XIV]*, *Peptostreptococcaceae_[XI]*, and *Erysipelotrichaceae* were also associated with decreased lung cancer risk, with ORs (95% CIs) of 0.76 (0.59–0.98), 0.80 (0.66–0.97), and 0.81 (0.67–0.99), respectively. No significant associations were shown after Bonferroni correction for 41 independent tests for the 119 taxa. When tests were stratified by race, sex, or cancer diagnosed within two years after mouth rinse sample collection, most of the associations largely remained (Supplementary Table 1). No significantly different associations between lung cancer subgroups were detected by formal tests of multiplicative interactions (data not shown).

Among 219 taxa for prevalence comparison between cases and controls, three genera and six species were associated with lung cancer risk at a $p < 0.05$. These taxa showed around 10% to 15% lower prevalence in lung cancer cases compared with controls; however, none of the associations were significant after Bonferroni correction for 70 independent tests (Table 3). As shown in Supplementary Table 2, a stronger association was observed between the presence of species *Leptotrichia sp._oral_taxon_225* and lung cancer risk in AAs (OR = 0.28, 95% CI 0.12–0.66; $p = 1.2 \times 10^{-4}$; Bonferroni-corrected $p = 0.006$ for 48 independent effective tests), showing a significant interaction with race ($p = 0.019$). The association of species *Capnocytophaga sputigena* was possibly caused by cancer development, as a more evident association was found for cases diagnosed within two years after mouth rinse sample collection (OR = 0.26, 95% CI 0.11–0.61; $p = 7.2 \times 10^{-4}$; Bonferroni-corrected $p = 0.034$) compared with cases who were diagnosed more than two years after mouth rinse sample collection (OR = 1.06, 95% CI 0.43–2.57; Supplementary Table 2).

Associations of pre-defined periodontal pathogens with lung cancer risk

Among the five periodontal pathogens, prevalence of *P. intermedia* and *P. gingivalis* was correlated with a Spearman correlation coefficient of 0.72. The prevalence of species *A. actinomycetemcomitans* was 10.3% in lung cancer cases and 5.1% in controls, whereas a lower prevalence of the other four pathogens was found in lung cancer cases compared with controls. However, none of the differences were statistically significant (Table 4). Compared to controls, a nominally higher prevalence of *A. actinomycetemcomitans* was observed in lung cancer cases among males (13.4% vs. 4.9%; $p = 0.041$) and ever-smokers (9.8% vs. 3.8%; $p = 0.021$) and a lower prevalence of *T. forsythia* was observed in AA lung cancer cases (56.6% vs. 74.7%; $p = 0.020$). However, none of these differences were significant after Bonferroni correction (Supplementary Table 3).

Discussion

In this nested case–control study among a low-income population, we found that abundance of six taxa and prevalence of nine taxa at OTU, genus, or family phylogenetic levels were nominally associated with decreased lung cancer risk. The prevalence of species *L. sp._oral_taxon_225* was associated with decreased risk of lung cancer among AAs, after Bonferroni correction. Our findings warrant validation in independent large-scale studies to further understand the role of oral microbiota in lung cancer development.

We observed that abundance of six taxa were lower in lung cancer cases compared to controls. Similar patterns were observed in subgroup analyses stratified by race, sex, or whether lung cancer cases were diagnosed within

Table 1 Selected characteristics of lung cancer cases and controls in the Southern Community Cohort Study

Characteristics	Cases ($n=156$)	Controls ($n=156$)	p[a]
Race (n [%])			
European American	73 (46.8%)	73 (46.8%)	1.000
African American	83 (53.2%)	83 (53.2%)	
Sex (n [%])			
Male	82 (52.6%)	82 (52.6%)	1.000
Female	74 (47.4%)	74 (47.4%)	
Age at sample collection (n [%])			
40–49 years	35 (22.4%)	38 (24.4%)	0.964
50–59 years	55 (35.3%)	56 (35.9%)	
60–69 years	45 (28.8%)	43 (27.6%)	
≥70 years	21 (13.5%)	19 (12.2%)	
Body mass index (n [%])			
<18.5	9 (5.8%)	3 (1.9%)	0.260
18.5–25	54 (34.6%)	51 (32.7%)	
25–30	54 (34.6%)	54 (34.6%)	
≥30	39 (25.0%)	48 (30.8%)	
BMI (kg/m^2)[b]	27.3 ± 6.8	28.3 ± 6.5	0.176
Smoking status (n [%])			
Current smoker	90 (57.7%)	90 (57.7%)	1.000
Former smoker	43 (27.6%)	43 (27.6%)	
Never smoker	23 (14.7%)	23 (14.7%)	
Smoking pack-years[b]	39.7 ± 34.4	29.0 ± 22.8	7.1×10^{-4}
Alcohol drinking (n [%])			
None	82 (52.6%)	82 (52.6%)	0.477
Light, <1 drink per day	35 (22.4%)	46 (29.5%)	
Moderate, 1–2 drink/day	16 (10.3%)	13 (8.3%)	
Heavy, >2 drinks per day	20 (12.8%)	15 (9.6%)	
Missing	3 (1.9%)	0 (0%)	
Total energy intake (kcal/day)[b]	2,517 ± 1,463	2,368 ± 1,243	0.296
Healthy eating index[b]	55.2 ± 12.3	56.8 ± 12.8	0.436
Family history of lung cancer (n [%])			
Yes	22 (14.1%)	16 (10.3%)	0.299
No	134 (85.9%)	140 (89.7%)	
Last visit to a dentistry (n [%])			
<6 months	34 (21.8%)	42 (26.9%)	0.418
6–12 months	25 (16.0%)	25 (16.0%)	
12–36 months	24 (15.4%)	29 (18.6%)	
≥36 months	62 (39.7%)	49 (31.4%)	
Missing	11 (7.1%)	11 (7.1%)	
Teeth lost due to decay or gum disease (n [%])			
None	6 (3.8%)	6 (3.8%)	0.045
1 to 10	14 (9.0%)	44 (28.2%)	
>10, not all	8 (5.1%)	30 (19.2%)	
All	14 (9.0%)	17 (10.9%)	
Missing	114 (73.1%)	59 (37.8%)	

[a]Chi-squared test for categorical variables and paired t test for continuous variables

[b]Mean ± standard deviation (SD); smoking pack-years were calculated among ever-smokers

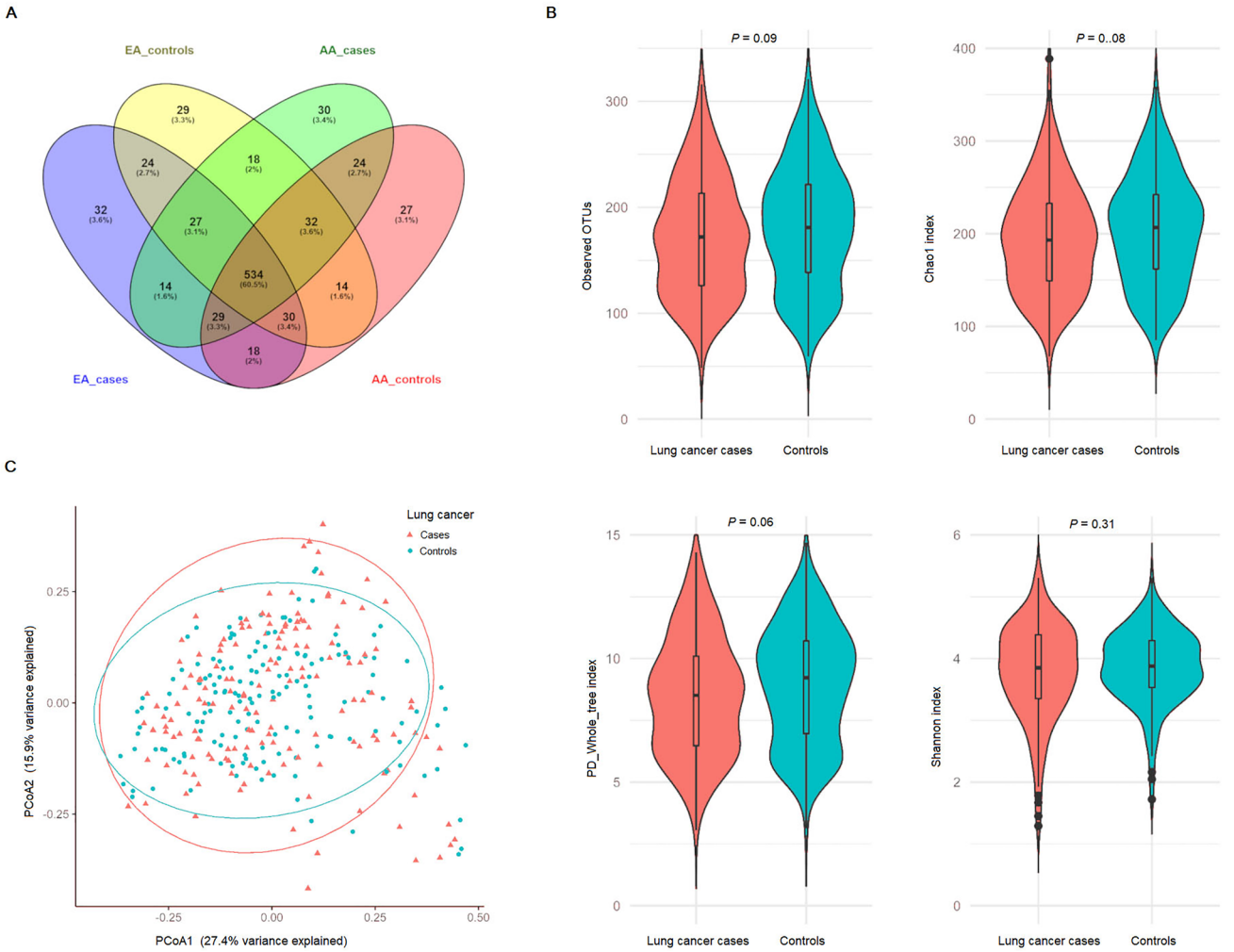

Fig. 1 Observed operational taxonomic units (OTUs) and oral OTU-level bacterial diversity in lung cancer cases and controls. **A** Shared and unique OTUs observed among lung cancer cases and controls of European Americans and African Americans. **B** Comparing mean observed OTUs and three alpha diversity indexes (Chao1, PD whole tree, and Shannon) between lung cancer cases and controls (paired t test, all $p > 0.05$). **C** Principal coordinate analysis (PCoA) of Bray–Curtis beta diversity ($p = 0.93$ from PERMANOVA-S, permutation number = 1,000,000)

two years after enrollment. The lower abundance of families *Peptoniphilaceae* (including its genus *Parvimonas* and species *P. micra*), *Peptostreptococcaceae_[XI]*, and *Erysipelotrichaceae* were more evident among ever-smoking cases, whereas enriched in never-smoking cases compared with corresponding controls (Supplementary Table 1), which suggest a strong impact of cigarette smoking [37–39]. Studies have shown that the genus *Parvimonas* and its species *P. micra* was more abundant in smokers' oral cavity than in non-smokers [38, 40, 41]. Species *P. micra* within the genus *Parvimonas* and family *Peptoniphilaceae*, a Gram-positive anaerobic cocci species, is commensal in the oral cavity and can cause periodontitis [42]. It is also related to infections of other organs including lung abscesses [43]. Oral *Parvimonas* and *P. micra* levels have been reported to be increased in oral squamous cell carcinoma [44–46] and decreased in colorectal cancer [47]. Although it is possible that oral bacteria such as *P. micra* can translocate to lung and cause cancer, additional studies are warranted to investigate the underlying mechanisms of lung carcinogenesis.

Decreased prevalence of several taxa were also observed in lung cancer cases. Such association pattern remained for most of these taxa when stratifying the participants by race, sex, time to diagnosis, or smoking status. Interestingly, the prevalence of the species *L. sp._oral_taxon_225* was more significantly decreased in AA lung cancer cases compared to controls (44.6% vs. 67.5%, $p = 1.2 \times 10^{-4}$), whereas it was slightly decreased in EA cases (41.1% vs. 43.8%, $p = 0.57$). These results may suggest a possible race-specific role of this species in lung cancer development. In addition, *L.*

Table 2 Abundance association of oral bacterial taxa with lung cancer risk[a]

Taxon	Median relative abundance (%)			
	Cases ($n=156$)	Controls ($n=156$)	OR (95% CI)[b]	p[b]
Phylum *Firmicute*				
Family *Lachnospiraceae_[XIV]*	0.205	0.226	0.76 (0.59–0.98)	0.037
Family *Peptoniphilaceae*	0.012	0.030	0.82 (0.69–0.97)	0.023
Genus *Parvimonas*	0.012	0.028	0.83 (0.70–0.98)	0.030
Species *P. micra*	0.012	0.028	0.83 (0.71–0.98)	0.031
Family *Peptostreptococcaceae_[XI]*	0.067	0.108	0.80 (0.66–0.97)	0.022
Family *Erysipelotrichaceae*	0.014	0.018	0.81 (0.67–0.99)	0.038

[a]The taxa had median relative abundance > 0.01% among control subjects

[b]Odds ratios (ORs), 95% confidence intervals (CIs), and p values were calculated from conditional logistic regression on normalized taxa counts using centered log-ratio transformation, adjusted for unmatched covariates including smoking pack-years, alcohol drinking status, total energy intake, BMI, last time of dentistry visit, and sequencing batch

Table 3 Prevalence association of oral bacterial taxa with lung cancer risk[a]

Taxon	N (%) carriage			
	Cases ($n=156$)	Controls ($n=156$)	OR (95% CI)[b]	p[b]
Phylum *Bacteroidetes*				
Genus *Bacteroidetes_[G-5]*	39 (25.0)	55 (35.3)	0.55 (0.30–1.00)	0.044
Species *Alloprevotella sp._oral_taxon_912*	41 (26.3)	56 (35.9)	0.36 (0.17–0.73)	0.017
Species *Capnocytophaga sputigena*	66 (42.3)	89 (57.1)	0.53 (0.31–0.92)	0.021
Phylum *Firmicute*				
Genus *Lactococcus*	23 (14.7)	39 (25.0)	0.43 (0.21–0.88)	0.017
Species *L. lactis*	23 (14.7)	39 (25.0)	0.43 (0.21–0.88)	0.017
Genus *Peptoniphilaceae_[G-1]*	17 (10.9)	34 (21.8)	0.43 (0.21–0.88)	0.028
Species *P_[G-1]. sp._oral_taxon_113*	17 (10.9)	34 (21.8)	0.43 (0.19–0.94)	0.028
Phylum *Fusobacteria*				
Species *Leptotrichia sp._oral_taxon_225*	67 (42.9)	88 (56.4)	0.57 (0.34–0.99)	0.041
Phylum *Synergistetes*				
Species *Fretibacterium fastidiosum*	46 (29.5)	66 (42.3)	0.54 (0.31–0.94)	0.026

[a]The taxa had median relative abundance ≤ 0.01% and carriage > 20% among control subjects

[b]Odds ratios (ORs), 95% confidence intervals (CIs), and p values were calculated from conditional logistic regression with non-carriers as the reference, adjusted for unmatched covariates including smoking pack-years, alcohol drinking status, total energy intake, BMI, last time of dentistry visit, sequencing batch, and sequencing depth

sp._oral_taxon_225 was reportedly significantly abundant in dental caries-free children [48] and its affiliated genus *Leptotrichia* was abundant in pancreatic cancer- and liver cancer-free controls [49, 50]. Further studies are needed to confirm and disentangle the possible protective role of *L. sp._oral_taxon_225* in lung cancer.

Increasing evidence has shown that oral pathogens can cause chronic inflammatory periodontal diseases, which are associated with an increased risk for cancers of oral cavity and other body sites [8, 9]. Among five pre-defined oral pathogens, lung cancer cases had a higher prevalence of *A. actinomycetemcomitans* compared to controls (10.3% vs 5.1%). A higher prevalence of *A. actinomycetemcomitans* in oral mouthwash samples has been found to be associated with increased risk of pancreatic cancer [49]. As a keystone pathogen for aggressive periodontitis, *A. actinomycetemcomitans* has been hypothesized to initiate Toll-like receptor signaling pathways [51], which are involved in inflammatory tumorigenesis [52–54]. We observed lower prevalence of other four periodontal pathogens in lung cancer cases. The lower prevalence of these pathogens in cases was unlikely to be caused by the cancer diagnosis, as similar results were also observed in those diagnosed within two years after sample collection.

Table 4 Prevalence association of pre-defined periodontal pathogens with lung cancer risk

Periodontal pathogens	*N* (%) carriage[a]			
	Cases	Controls	OR (95% CI)[b]	p[b]
	$n=156$	$n=156$		
Aggregatibacter actinomycetemcomitans	16 (10.3)	8 (5.1)	2.35 (0.92–6.00)	0.066
Porphyromonas gingivalis	101 (64.7)	115 (73.7)	0.60 (0.34–1.05)	0.069
Prevotella intermedia	102 (65.4)	108 (69.2)	0.80 (0.47–1.38)	0.424
Tannerella forsythia	83 (53.2)	97 (62.2)	0.72 (0.40–1.31)	0.280
Treponema denticola	77 (49.4)	92 (59.0)	0.67 (0.37–1.21)	0.177

[a]Number and percentage of cases and controls carrying the periodontal pathogen

[b]Odds ratios (ORs), 95% confidence intervals (CIs), and p values were calculated from conditional logistic regression with non-carriers as the reference, adjusted for unmatched covariates including smoking pack-years, alcohol drinking status, total energy intake, BMI, last time of dentistry visit, sequencing batch, and sequencing depth

However, the associations of these four pre-defined oral bacterial pathogens with lung cancer risk were not statistically significant.

This prospective study, to date, is the largest in sample size to investigate the role of oral microbiome in the development of lung cancer. Other strengths of this study include the nested case–control study design, participants from a low-income population of both African and European ancestries, as well as the ability to match multiple sociodemographic factors and smoking behavior. Several limitations in this study should be pointed out. First, although lung cancer risk association patterns were the same for most microbial bacteria in both EAs and AAs, our study with 312 samples may be underpowered. Second, the present study lacked a systematic assessment of oral health at the baseline examination, when samples were collected. In this study, we used time since last dentistry visit as a surrogate measurement to control for confounding oral hygiene and/or oral disease. Studies among participants with more detailed oral health, including tooth loss, may further clarify the role of oral microbiota in lung cancer. Third, the 16S *rRNA* gene sequencing method and using QIIME V1 instead of QIIME V2 may have a limited accuracy to profile taxa at the OTU/species level [55, 56]. Further studies using metagenomics approaches and state-of-the-art analysis pipeline are warranted to confirm our findings.

In summary, we found that multiple oral bacterial families, genera, and species may be associated with risk of lung cancer in a low-income population, although most of the associations were not statistically significant after controlling multiple testing. Further studies with larger sample sizes and using metagenomics approaches, as well as in vitro and in vivo studies are warranted to investigate the role of oral microbiota in the development of lung cancer.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10552-021-01490-6.

Acknowledgments The authors wish to thank all individuals who took part in the study and all researchers, clinicians, technicians, and administrative staff who enabled this work to be carried out. We thank Dr. Mary Shannon Byers for assistance with manuscript preparation.

Author contributions WZ, WJB, and QC conceived the study. JS, YY, and JL analyzed the data and drafted the manuscript. HX, XW, JW, RC, XOS, WZ, WJB, and QC contributed resources, collected, and processed samples. All the authors edited and approved the final draft.

Funding This study was supported by the National Institutes of Health (R01CA207466, R01CA92447, U01CA202979, and U54CA163072). Sample preparation was conducted at the Survey and Biospecimen Shared Resources, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485). The 16S *rRNA* gene sequencing was performed at the VANderbilt Technologies for Advanced Genomics (VANTAGE) Core, which is partly supported by NIH/NCRR grant G20 RR030956. The data analyses were conducted using the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, which is supported in part by the National Institutes of Health S10 Shared Instrumentation (1S10OD023680). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability Data and statistical codes used in the present study can be requested through the SCCS Online request System (https://ors.southerncommunitystudy.org).

Declarations

Conflict of interest The authors have declared no conflicts of interest.

Ethical approval The Southern Community Cohort Study was reviewed and approved by the institutional review boards at Vanderbilt University and Meharry Medical College.

Consent to participate Written informed consent was obtained from all individuals, and the research was performed by the principles of the Helsinki Declaration.

Patient consent for publication All study participants provided written informed consent for the publication of any data and associated images.

References

1. Torre LA, Bray F, Siegel RL et al (2015) Global cancer statistics, 2012. CA Cancer J Clin 65:87–108. https://doi.org/10.3322/caac.21262
2. Bray F, Ferlay J, Soerjomataram I et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394–424. https://doi.org/10.3322/caac.21492
3. Ezzati M, Lopez AD (2003) Estimates of global mortality attributable to smoking in 2000. Lancet Lond Engl 362:847–852. https://doi.org/10.1016/S0140-6736(03)14338-3
4. Ezzati M, Henley SJ, Lopez AD, Thun MJ (2005) Role of smoking in global and regional cancer epidemiology: current patterns and data needs. Int J Cancer 116:963–971. https://doi.org/10.1002/ijc.21100
5. Sharma N, Bhatia S, Sodhi AS, Batra N (2018) Oral microbiome and health. AIMS Microbiol 4:42–66. https://doi.org/10.3934/microbiol.2018.1.42
6. Verma D, Garg PK, Dubey AK (2018) Insights into the human oral microbiome. Arch Microbiol 200:525–540. https://doi.org/10.1007/s00203-018-1505-3
7. Willis JR, Gabaldón T (2020) The human oral microbiome in health and disease: from sequences to ecosystems. Microorganisms. https://doi.org/10.3390/microorganisms8020308
8. Irfan M, Delgado RZR, Frias-Lopez J (2020) The oral microbiome and cancer. Front Immunol 11:591088. https://doi.org/10.3389/fimmu.2020.591088
9. Tuominen H, Rautava J (2021) Oral microbiota and cancer development. Pathobiol J Immunopathol Mol Cell Biol 88:116–126. https://doi.org/10.1159/000510979
10. Hiraki A, Matsuo K, Suzuki T et al (2008) Teeth loss and risk of cancer at 14 common sites in Japanese. Cancer Epidemiol Biomarkers Prev 17:1222–1227. https://doi.org/10.1158/1055-9965.EPI-07-2761
11. Chung M, York BR, Michaud DS (2019) Oral health and cancer. Curr Oral Health Rep 6:130–137. https://doi.org/10.1007/s40496-019-0213-7
12. Yoon H-S, Wen W, Long J et al (2019) Association of oral health with lung cancer risk in a low-income population of African Americans and European Americans in the Southeastern United States. Lung Cancer Amst Neth 127:90–95. https://doi.org/10.1016/j.lungcan.2018.11.028
13. Sampaio-Maia B, Caldas IM, Pereira ML et al (2016) The oral microbiome in health and its implication in oral and systemic diseases. Adv Appl Microbiol 97:171–210. https://doi.org/10.1016/bs.aambs.2016.08.002
14. Brenner DR, McLaughlin JR, Hung RJ (2011) Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. PLoS One 6:e17479. https://doi.org/10.1371/journal.pone.0017479
15. Zhan P, Suo L, Qian Q et al (2011) Chlamydia pneumoniae infection and lung cancer risk: a meta-analysis. Eur J Cancer 47:742–747. https://doi.org/10.1016/j.ejca.2010.11.003
16. Hua-Feng X, Yue-Ming W, Hong L, Junyi D (2015) A meta-analysis of the association between Chlamydia pneumoniae infection and lung cancer risk. Indian J Cancer 52(Suppl 2):e112-115. https://doi.org/10.4103/0019-509X.172506
17. Hosgood HD, Sapkota AR, Rothman N et al (2014) The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. Environ Mol Mutagen 55:643–651. https://doi.org/10.1002/em.21878
18. Yan X, Yang M, Liu J et al (2015) Discovery and validation of potential bacterial biomarkers for lung cancer. Am J Cancer Res 5:3111–3122
19. Yang J, Mu X, Wang Y et al (2018) Dysbiosis of the salivary microbiome is associated with non-smoking female lung cancer and correlated with immunocytochemistry markers. Front Oncol 8:520. https://doi.org/10.3389/fonc.2018.00520
20. Hosgood HD, Cai Q, Hua X et al (2021) Variation in oral microbiome is associated with future risk of lung cancer among never-smokers. Thorax 76:256–263. https://doi.org/10.1136/thoraxjnl-2020-215542
21. Signorello LB, Hargreaves MK, Steinwandel MD et al (2005) Southern community cohort study: establishing a cohort to investigate health disparities. J Natl Med Assoc 97:972–979
22. Long J, Cai Q, Steinwandel M et al (2017) Association of oral microbiome with type 2 diabetes risk. J Periodontal Res 52:636–643. https://doi.org/10.1111/jre.12432
23. Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303
24. Caporaso JG, Lauber CL, Walters WA et al (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624. https://doi.org/10.1038/ismej.2012.8
25. Joshi N, Fass J (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle
26. Nikolenko SI, Korobeynikov AI, Alekseyev MA (2013) BayesHammer: bayesian clustering for error correction in single-cell sequencing. BMC Genomics 14(Suppl 1):S7. https://doi.org/10.1186/1471-2164-14-S1-S7
27. Masella AP, Bartram AK, Truszkowski JM et al (2012) PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13:31. https://doi.org/10.1186/1471-2105-13-31
28. Schirmer M, Ijaz UZ, D'Amore R et al (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res 43:e37. https://doi.org/10.1093/nar/gku1341
29. Dewhirst FE, Chen T, Izard J et al (2010) The human oral microbiome. J Bacteriol 192:5002–5017. https://doi.org/10.1128/JB.00542-10
30. Tang Z-Z, Chen G, Alekseyenko AV (2016) PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. Bioinformatics 32:2618–2625. https://doi.org/10.1093/bioinformatics/btw311
31. Fernandes AD, Reid JN, Macklaim JM et al (2014) Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome 2:15. https://doi.org/10.1186/2049-2618-2-15
32. Holt SC (2000) Ebersole JL (2005) Porphyromonas gingivalis, Treponema denticola, and Tannerella forsythia: the "red complex", a prototype polybacterial pathogenic consortium in periodontitis. Periodontol 38:72–122. https://doi.org/10.1111/j.1600-0757.2005.00113.x
33. Cullinan MP, Hamlet SM, Westerman B et al (2003) Acquisition and loss of porphyromonas gingivalis, actinobacillus actinomycetemcomitans and prevotella intermedia over a 5-year period: effect of a triclosan/copolymer dentifrice. J Clin Periodontol 30:532–541
34. Yu D, Sonderman J, Buchowski MS et al (2015) Healthy eating and risks of total and cause-specific death among low-income populations of african-americans and other adults in the

southeastern United States: a prospective cohort study. PLoS Med 12:e1001830. https://doi.org/10.1371/journal.pmed.1001830
35. Galwey NW (2009) A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. Genet Epidemiol 33:559–568. https://doi.org/10.1002/gepi.20408
36. Cinar O, Viechtbauer W (2016) poolr: package for pooling the results from (dependent) tests. https://ozancinar.github.io/poolr/
37. Yang Y, Zheng W, Cai Q-Y et al (2019) Cigarette smoking and oral microbiota in low-income and African-American populations. J Epidemiol Community Health 73:1108–1115. https://doi.org/10.1136/jech-2019-212474
38. Wu J, Peters BA, Dominianni C et al (2016) Cigarette smoking and the oral microbiome in a large study of American adults. ISME J 10:2435–2446. https://doi.org/10.1038/ismej.2016.37
39. Huang C, Shi G (2019) Smoking and microbiome in oral, airway, gut and some systemic diseases. J Transl Med 17:225. https://doi.org/10.1186/s12967-019-1971-7
40. Delima SL, McBride RK, Preshaw PM et al (2010) Response of subgingival bacteria to smoking cessation. J Clin Microbiol 48:2344–2349. https://doi.org/10.1128/JCM.01821-09
41. Shchipkova AY, Nagaraja HN, Kumar PS (2010) Subgingival microbial profiles of smokers with periodontitis. J Dent Res 89:1247–1253. https://doi.org/10.1177/0022034510377203
42. Belstrøm D, Fiehn N-E, Nielsen CH et al (2015) Differentiation of salivary bacterial profiles of subjects with periodontitis and dental caries. J Oral Microbiol 7:27429. https://doi.org/10.3402/jom.v7.27429
43. Zhang Y, Song P, Zhang R et al (2021) Clinical characteristics of chronic lung abscess associated with parvimonas micra diagnosed using metagenomic next-generation sequencing. Infect Drug Resist 14:1191–1198. https://doi.org/10.2147/IDR.S304569
44. Lee W-H, Chen H-M, Yang S-F et al (2017) Bacterial alterations in salivary microbiota and their association in oral cancer. Sci Rep 7:16540. https://doi.org/10.1038/s41598-017-16418-x
45. Yang C-Y, Yeh Y-M, Yu H-Y et al (2018) Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. Front Microbiol 9:862. https://doi.org/10.3389/fmicb.2018.00862
46. Zhao H, Chu M, Huang Z et al (2017) Variations in oral microbiota associated with oral cancer. Sci Rep 7:11773. https://doi.org/10.1038/s41598-017-11779-9
47. Flemer B, Warren RD, Barrett MP et al (2018) The oral microbiota in colorectal cancer is distinctive and predictive. Gut 67:1454–1463. https://doi.org/10.1136/gutjnl-2017-314814
48. Al-Hebshi NN, Baraniya D, Chen T et al (2019) Metagenome sequencing-based strain-level and functional characterization of supragingival microbiome associated with dental caries in children. J Oral Microbiol 11:1557986. https://doi.org/10.1080/20002297.2018.1557986
49. Fan X, Alekseyenko AV, Wu J et al (2018) Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. Gut 67:120–127. https://doi.org/10.1136/gutjnl-2016-312580
50. Li D, Xi W, Zhang Z et al (2020) Oral microbial community analysis of the patients in the progression of liver cancer. Microb Pathog 149:104479. https://doi.org/10.1016/j.micpath.2020.104479
51. Sun Y, Shu R, Li C-L, Zhang M-Z (2010) Gram-negative periodontal bacteria induce the activation of Toll-like receptors 2 and 4, and cytokine production in human periodontal ligament cells. J Periodontol 81:1488–1496. https://doi.org/10.1902/jop.2010.100004
52. Elinav E, Nowarski R, Thaiss CA et al (2013) Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. Nat Rev Cancer 13:759–771. https://doi.org/10.1038/nrc3611
53. Sheflin AM, Whitney AK, Weir TL (2014) Cancer-promoting effects of microbial dysbiosis. Curr Oncol Rep 16:406. https://doi.org/10.1007/s11912-014-0406-0
54. Dajon M, Iribarren K, Cremer I (2017) Toll-like receptor stimulation in cancer: a pro- and anti-tumor double-edged sword. Immunobiology 222:89–100. https://doi.org/10.1016/j.imbio.2016.06.009
55. Ranjan R, Rani A, Metwally A et al (2016) Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun 469:967–977. https://doi.org/10.1016/j.bbrc.2015.12.083
56. Bolyen E, Rideout JR, Dillon MR et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9