Prevention and Epidemiology

Cancer Research

Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers

Brandilyn A. Peters[1], Jing Wu[1,2], Zhiheng Pei[2,3,4], Liying Yang[5], Mark P. Purdue[6], Neal D. Freedman[6], Eric J. Jacobs[7], Susan M. Gapstur[7], Richard B. Hayes[1,2], and Jiyoung Ahn[1,2]

Abstract

Bacteria may play a role in esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC), although evidence is limited to cross-sectional studies. In this study, we examined the relationship of oral microbiota with EAC and ESCC risk in a prospective study nested in two cohorts. Oral bacteria were assessed using 16S rRNA gene sequencing in prediagnostic mouthwash samples from $n = 81/160$ EAC and $n = 25/50$ ESCC cases/matched controls. Findings were largely consistent across both cohorts. Metagenome content was predicted using PiCRUST. We examined associations between centered log-ratio transformed taxon or functional pathway abundances and risk using conditional logistic regression adjusting for BMI, smoking, and alcohol. We found the periodontal pathogen *Tannerella forsythia* to be associated with higher risk of EAC. Furthermore, we found that depletion of the commensal genus *Neisseria* and the species *Streptococcus pneumoniae* was associated with lower EAC risk. Bacterial biosynthesis of carotenoids was also associated with protection against EAC. Finally, the abundance of the periodontal pathogen *Porphyromonas gingivalis* trended with higher risk of ESCC. Overall, our findings have potential implications for the early detection and prevention of EAC and ESCC. *Cancer Res; 77(23); 6777–87. ©2017 AACR.*

Introduction

Esophageal cancer is the eighth most common cancer and sixth most common cause of cancer-related death worldwide (1). Because late-stage presentation is common in most cases, esophageal cancers are highly fatal; 5-year survival rates range from 15% to 25% in most countries (2). Consequently, there is a critical need for new avenues of prevention, risk stratification, and early detection.

The two main types, esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC), differ greatly in incidence, geography, and etiology. ESCC, the most common type worldwide, predominates in developing countries, while EAC has become the predominant type in developed countries as incidence rates continue to rise (2, 3). Known risk factors include gastroesophageal reflux disease (GERD), obesity, low fruit/vegetable intake, and smoking for EAC, and alcohol drinking, low fruit/vegetable intake, and smoking for ESCC (4), but the etiology of these diseases cannot be fully explained by these factors.

Recently, upper digestive tract microbiota have been suggested to play a role in esophageal cancer etiology, and in particular in the rising incidence of EAC in developed countries (5). The complex microbial community of the upper digestive tract, consisting of mutualists, commensals, and pathogens, could facilitate carcinogenesis via activation of Toll-like receptors (6), or protect against carcinogenesis via synthesis of vitamins or providing barriers to pathogen invasion (5). Cross-sectional studies report distinct differences in upper digestive tract microbiota between GERD (7–9), Barrett's esophagus (an EAC precursor; refs. 7–10), EAC (7, 11), esophageal squamous dysplasia (ESD, an ESCC precursor; ref. 12), or ESCC (13) cases and controls. In addition, periodontitis (a disease of oral dysbiosis) may be associated with increased esophageal cancer risk (14). However, no studies have prospectively examined whether upper digestive tract microbiota influence risk for subsequent esophageal cancer.

We hypothesized that oral microbiota influence development of esophageal cancer. The oral microbiota shape the esophageal microbiome (15), due to migration of oral bacteria to the esophagus (16) and, therefore, may contribute to esophageal carcinogenesis. We conducted a prospective study nested in two large U.S. cohorts, to determine whether oral microbiota are associated with subsequent EAC or ESCC risk.

Patients and Methods

Parent cohorts

Participants were drawn from two U.S. cohorts: the NCI Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial cohort and American Cancer Society (ACS) Cancer Prevention Study II (CPS-II) Nutrition cohort. Characteristics of these

[1]Division of Epidemiology, Department of Population Health, NYU School of Medicine, New York, New York. [2]NYU Perlmutter Cancer Center, New York, New York. [3]Department of Pathology, NYU School of Medicine, New York, New York. [4]Department of Veterans Affairs New York Harbor Healthcare System, New York, New York. [5]Division of Translational Medicine, Department of Medicine, NYU School of Medicine, New York, New York. [6]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland. [7]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Current address for J. Wu: Microbiology Section, New York City Public Health Laboratory, New York, NY.

Corresponding Author: Jiyoung Ahn, NYU School of Medicine, 650 1st Avenue, New York, NY 10016. Phone: 212-263-3390; Fax: 212-263-8570; E-mail: Jiyoung.Ahn@nyumc.org

doi: 10.1158/0008-5472.CAN-17-1296

cohorts are comparable, with both cohorts collecting oral wash samples and comprehensive demographic information, and following prospectively for cancer incidence.

PLCO (17) is a large population-based randomized trial designed to determine effects of screening on cancer-related mortality in men and women aged 55 to 74, recruited in 1993 to 2001, and followed for cancer incidence. Participants were randomized to a screening or control arm. Oral wash samples were collected in the control arm only ($n = 52{,}000$). Incident cancers were ascertained by annual mailed questionnaire and verified through medical records or death certificates.

CPS-II (18) includes >184,000 participants, aged 50 to 74 who completed a mailed baseline questionnaire in 1992. Follow-up questionnaires have been sent to cohort members every other year to update information and ascertain incident cancers, which are also verified through medical records, state registries, or death certificates. During 2000 to 2002, oral wash samples were collected from 70,004 participants.

Nested case–control study

Incident cases were cohort participants diagnosed with esophageal cancer any time after oral wash collection (collection to diagnosis time ranged from <1 year to 9 years; first quartile, median, third quartile = 1, 3, 5 years) and had no prior cancer history (except nonmelanoma skin cancer). Matched controls were selected at a case:control ratio of 1:2 by incidence density sampling without replacement among participants who provided an oral wash sample in the same year as the index case, had no cancer at or prior to index case diagnosis, and were of the same cohort, age, sex, and race as the index case.

A total of 368 samples were provided and successfully sequenced, including 117 complete sets (1 case: 2 controls), 2 reduced sets (1 case: 1 control), and 13 unmatched controls (due to missing case, case failing sequencing, or nonesophageal case). On the basis of ICD morphology codes (EAC: 8140, 8144, 8480, 8481, 8560; ESCC: 8070, 8071, 8072, 8074, 8052), we included 81 EAC cases (with 160 matched controls) and 25 ESCC cases (with 50 matched controls) in the current analysis ($N = 316$). Cases of other or missing morphology ($n = 13$), their matched controls ($n = 26$), and unmatched controls ($n = 13$) were excluded.

This study was conducted in accordance with the U.S. Common Rule and approved by the IRB of New York University School of Medicine (New York, NY), NCI (Bethesda, MD), and ACS, and participants provided informed consent.

Covariate assessment

Covariate information was extracted from questionnaires preceding oral sample collection for each participant. Body mass index (BMI) was categorized as normal or underweight (BMI < 25 kg/m^2), overweight (25 $\leq$ BMI < 30 kg/m^2), or obese (BMI $\geq$ 30 kg/m^2). Smoking status was classified as never, former, or current. Drinking level was classified as never, moderate, or heavy (19). Servings of fruits and vegetables per day, derived from food frequency questionnaire responses, were categorized as low or high based on cohort-specific medians.

Oral wash sample collection

Participants were asked to swish with 10 mL Scope mouthwash (P&G) and expectorate into a tube (17, 18). Samples were shipped to each cohort's biorepository and stored at −80°C. The oral microbiome is highly stable over time (20–22) and shows much greater interindividual than intraindividual variation, indicating that a one-time oral sample collection is appropriate for assessing oral microbial risk factors in a cohort study.

Microbiome assay

We extracted DNA from oral wash samples using the PowerSoil DNA Isolation Kit (Mo Bio). Barcoded amplicons were generated covering the 16S rRNA gene V4 region using F515/R806 primers. The PCR reaction used FastStart High Fidelity PCR System, dNTP pack (Roche) as follows: initial denaturing at 94°C for 3 minutes, followed by 25 cycles of 94°C for 15 seconds, 52°C for 45 seconds and 72°C for 1 minute, and a final extension at 72°C for 8 minutes. PCR products were purified using Agencourt AMPure XP (Beckman Coulter Life Sciences), quantified using Agilent 4200 TapeStation (Agilent Technologies), pooled at equimolar concentrations and sequenced on Illumina MiSeq with a 300-cycle (2 × 151 bp) reagent kit.

Sequence data processing

Paired-end reads were joined and demultiplexed, and poor-quality reads excluded, using default parameters in QIIME (23). The 11,422,831 quality-filtered reads (from $N = 368$ samples) were clustered into operational taxonomic units (OTU) against the Human Oral Microbiome Database (HOMD) reference sequence collection (version 14.5; ref. 24), and assigned HOMD taxonomy, using QIIME script *pick_closed_reference_otus.py* (23). This method discards reads not matching the database, leaving 11,074,719 reads [mean ± SD = 30,094 ± 21,059; range = (4,965–203,242)] and 569 OTUs. We generated a phylogenetic tree from aligned HOMD reference sequences using FastTree (25).

Quality control

All samples underwent DNA extraction and sequencing in the same laboratory, with personnel blinded to case/control status. DNA from volunteer oral wash samples was included in the sequencing batches: six replicates from each of 4 volunteers in the CPS-II batch, and eight replicates from each of the same 4 volunteers in the PLCO batch. Intraclass correlation coefficients for the Shannon diversity index and relative abundance of major oral phyla were high (Supplementary Table S1), and principal coordinate analysis of UniFrac distances (26) showed clustering of repeat samples for each volunteer, indicating excellent reproducibility (Supplementary Fig. S1).

Statistical analysis

We used multiple imputation ("mice" package, R; ref. 27) to impute missing data for three important predictors of esophageal cancer, BMI, alcohol drinking, and fruit and vegetable intake. A total of 23 participants (7.3%) were missing BMI, 36 (11.4%) were missing alcohol drinking, and 39 (12.3%) were missing fruit and vegetable intake (% missing by case/control group shown in Table 1). Predictors of BMI category (<25, 25–30, ≥30) used in imputation were sex, race, age, cohort, smoking status, education level, and ethanol intake. Predictors of alcohol drinking (none, moderate, heavy) and fruit/vegetable intake (low or high) used in imputation were sex, race, age, cohort, smoking status, education level, and continuous BMI. Ten imputed datasets were used in analysis, and we present pooled estimates and P values.

α-Diversity (within-subject diversity) was assessed by richness and the Shannon diversity index, calculated in 100 iterations of

rarefied OTU tables of 4,500 sequence reads per sample. This depth was chosen to sufficiently reflect sample diversity (Supplementary Fig. S2) while retaining all participants. We examined whether α-diversity differed between cases and controls in conditional logistic regression using matched sets as strata and adjusting for smoking status, BMI category, and alcohol drinking level.

β-Diversity (between-subject diversity) was assessed at OTU level using unweighted and weighted UniFrac distances (26). Permutational multivariate analysis of variance (PERMANOVA; "adonis" function, "vegan" package, R; ref. 28) was used to examine statistically whether overall bacterial community composition differed by case/control status, using matched sets as strata and adjusting for smoking status, BMI category, and alcohol drinking level.

The 569 OTUs were agglomerated to 12 phyla, 26 classes, 42 orders, 70 families, 149 genera, and 513 species. We applied the centered log-ratio (clr) transformation (29) to the taxa counts at each level (e.g., phylum, class, etc.) after adding a pseudocount of 1. We used conditional logistic regression, using matched sets as strata and adjusting for smoking status, BMI category, and alcohol drinking level, to determine whether abundance of bacterial taxa predicts esophageal cancer risk. This analysis included only taxa present in ≥15% of the 316 participants (10 phyla, 20 classes, 28 orders, 46 families, 85 genera, 266 species), to exclude rare taxa and thereby minimize the number of statistical tests conducted. *A priori* species of interest were "red complex" periodontal pathogens: *Tannerella forsythia* (*T. forsythia*), *Porphyromonas gingivalis* (*P. gingivalis*), and *Treponema denticola* (*T. denticola*; ref. 30). For other taxa, *P* values were adjusted for the FDR.

Metagenome content was predicted using PiCRUSt (31). Because PiCRUST gene content is precomputed for the GreenGenes database of 16S rRNA genes, for this analysis, we performed closed-reference OTU picking against the GreenGenes database prior to PiCRUST. The 5507 KEGG (32) gene orthologs were grouped into 270 KEGG pathways. We applied the clr transformation (29) to pathway counts after adding a pseudocount of 1, filtered to include pathways present in ≥15% of participants (255 pathways), and used conditional logistic regression, as described above, to determine whether abundance of functional pathways predicts esophageal cancer risk.

Ecological networks among species were inferred using the SPIEC-EASI (SParse InversE Covariance Estimation for Ecological Association Inference) algorithm (33). This statistical method, designed for ecological network inference from amplicon sequencing datasets, accounts for compositional data structure using the clr transformation and assumes a sparse underlying ecological association network. We applied SPIEC-EASI separately to EAC cases and matched controls, and ESCC cases and matched controls. The "igraph" package in R was used for network visualization.

All statistical tests were two-sided. A $P < 0.05$ was considered of nominal significance, and an FDR-adjusted *P* value (q-value) <0.10 was considered significant after multiple comparisons adjustment. Analyses were conducted using R 3.2.1.

Results

Participant characteristics

Cases and their matched controls did not differ on matching factors (Table 1). Although obesity, low fruit/vegetable intake, and smoking are recognized risk factors for EAC, and alcohol drinking, low fruit/vegetable intake, and smoking are recognized risk factors for ESCC, only alcohol drinking was associated with ESCC ($P = 0.004$).

Overall microbiota diversity in relation to EAC and ESCC

EAC and ESCC cases did not differ significantly from matched controls in oral α-diversity, as measured by species richness and the Shannon diversity index, or overall oral microbiome composition (β-diversity), as measured by unweighted and weighted UniFrac distances (Supplementary Table S2).

Taxa associated with EAC

For the *a priori* "red complex" periodontal pathogens (30), a doubling of *T. forsythia* abundance relative to the geometric mean of all taxa was associated with 1.21 [95% confidence interval (CI), 1.01–1.46] times higher odds of EAC ($P = 0.04$), while abundance of *P. gingivalis* and *T. denticola* was not associated with EAC risk (Table 2; Fig. 1). We identified several other oral taxa nominally associated with EAC risk (Table 3; Fig. 1), although none reached the significance threshold after FDR adjustment (all q-value > 0.32). Increased abundance of species *Actinomyces cardiffensis*, *Selenomonas oral taxon 134*, and *Veillonella oral taxon 917* was associated with higher EAC risk (all $P < 0.05$). Conversely, increased abundance of *Corynebacterium durum*, *Prevotella nanceiensis*, *Streptococcus pneumoniae*, *Lachnoanaerobaculum umeaense*, *Oribacterium parvum*, *Solobacterium moorei*, *Neisseria sicca*, *Neisseria flavescens*, and *Haemophilus oral taxon 908* was associated with lower EAC risk (all $P <$ 0.05). Additional adjustment for fruit/vegetable intake did not impact effect estimates (percent change in β-coefficient for all nominally significant taxa <12%).

We observed that the majority of these species were associated with each other in an ecological network analysis (Fig. 2A). The protective species in phylum Proteobacteria (*Neisseria sicca*, *Neisseria flavescens*, and *Haemophilus oral taxon 908*) were closely connected, as were some of the protective species in phylum Firmicutes (*Solobacterium moorei*, *Oribacterium parvum*, and *Lachnoanaerobaculum umeaense*). Some of the species formed their own networks (i.e., unrelated to other EAC-associated species), including *Streptococcus pneumoniae* and *Selenomonas oral taxon 134*.

We additionally explored heterogeneity of taxon abundance–EAC associations by years from oral wash collection to diagnosis (≤ or > median of 3 years), cohort (CPS-II or PLCO), smoking status (ever or never), obesity (nonobese or obese), and fruit/vegetable intake (low or high). Taxon findings were consistent across years to diagnosis subgroups (all $P_{\text{interaction}} > 0.12$; Supplementary Table S3). Similarly, taxon findings were largely consistent across cohorts (Supplementary Table S4; Supplementary Fig. S3); in particular, *Streptococcus pneumoniae*, *Solobacterium moorei*, *Veillonella oral taxon 917*, *Neisseria sicca*, *Neisseria flavescens*, and *Haemophilus oral taxon 908* showed homogenous associations with EAC in both cohorts. *Selenomonas oral taxon 134* was associated with higher EAC risk in the PLCO cohort only ($P_{\text{interaction}} = 0.02$). Similarly, the periodontal pathogens tended to be associated with higher EAC risk only in PLCO ($P_{\text{interaction}} = 0.11$, 0.35, and 0.04 for *P. gingivalis*, *T. forsythia*, and *T. denticola*, respectively). When stratifying by smoking status, we observed that *Lachnoanaerobaculum umeaense* was associated with lower EAC risk only in smokers

Table 1. Prediagnosis demographic characteristics of esophageal adenocarcinoma and squamous cell carcinoma cases and matched controls

	Adenocarcinoma			Squamous cell carcinoma		
Characteristics	**Cases ($n = 81$)**	**Matched controls ($n = 160$[a])**	***P***	**Cases ($n = 25$)**	**Matched controls ($n = 50$)**	***P***
Sex[b] (%)			1.00[c]			1.00[c]
Women	7.4	7.5		60.0	60.0	
Men	92.6	92.5		40.0	40.0	
Age[b] (mean ± SD)	68.0 ± 6.7	68.0 ± 6.6	0.95[d]	66.6 ± 6.5	66.8 ± 6.4	0.83[d]
Race[b] (%)			1.00[c]			1.00[c]
White	97.5	97.5		84.0	84.0	
Other	2.5	2.5		16.0	16.0	
BMI[e] (%)			0.38[c,f]			0.07[c,f]
Normal weight	22.2	21.9		52.0	38.0	
Overweight	50.6	55.6		36.0	32.0	
Obese	21.0	13.8		4.0	26.0	
Missing	6.2	8.8		8.0	4.0	
Smoking (%)			0.12[c]			0.36[c]
Never	25.9	36.9		36.0	44.0	
Current	9.9	5.0		16.0	6.0	
Former	64.2	58.1		48.0	50.0	
Alcohol drinking[g] (%)			0.20[c,f]			0.004[c,f]
Never	21.0	25.6		24.0	26.0	
Moderate	51.9	53.1		32.0	54.0	
Heavy	17.3	9.4		36.0	6.0	
Missing	9.9	11.9		8.0	14.0	
Fruit and vegetable intake[h] (%)			0.85[c,f]			1.00[c,f]
Low	44.4	42.5		48.0	48.0	
High	43.2	45.6		40.0	38.0	
Missing	12.3	11.9		12.0	14.0	

[a]There were two incomplete case sets (1 case: 1 control).
[b]Sex, age, and race were matching factors.
[c]Differences between cases and controls were detected using the χ^2 test.
[d]Differences between cases and controls were detected using the Wilcoxon rank-sum test.
[e]Normal-weight: BMI < 25 kg/m^2; overweight: 25 ≤ BMI < 30 kg/m^2; obese: BMI ≥ 30 kg/m^2.
[f]*P* value determined after exclusion of those missing the variable.
[g]Moderate drinker: >0 but ≤1 drinks/day for women and >0 but ≤2 drinks/day for men; heavy drinker: >1 drinks/day for women and >2 drinks/day for men.
[h]Low and high intake groups reflect participants below or above cohort-specific median of servings of fruit and vegetables/day. CPS-II median = 4.62 servings/day; PLCO median = 6.10 servings/day.

($P_{interaction} = 0.02$; Supplementary Table S5), while other taxon–EAC associations did not differ significantly between ever and never smokers ($P_{interaction} > 0.19$). When stratifying by obesity, we observed that *Actinomyces cardiffensis* was associated with higher EAC risk only in nonobese ($P_{interaction} = 0.02$), while other taxon–EAC associations did not differ significantly between nonobese and obese ($P_{interaction} > 0.11$; Supplementary Table S6). Finally, when we stratified by fruit and vegetable intake, order *Actinomycetales* was associated with higher EAC risk only in those with higher fruit and vegetable intake ($P_{interaction} = 0.05$), while other interactions were nonsignificant ($P_{interaction} > 0.18$; Supplementary Table S7).

Taxa associated with ESCC

The periodontal pathogen *P. gingivalis* was marginally associated with higher ESCC risk [OR (95% CI) = 1.30 (0.96–1.77); $P = 0.09$; Table 2; Fig. 1]. Several other species were nominally associated with ESCC risk (Table 3; Fig. 1), although none reached the significance threshold after FDR adjustment (all q-value > 0.80). Increased abundance of *Prevotella*

Table 2. Periodontal pathogens[a] and risk for incident esophageal adenocarcinoma or squamous cell carcinoma

	Adenocarcinoma				Squamous cell carcinoma			
	Median relative abundance (% carriage[b])				Median relative abundance (% carriage[b])			
Periodontal pathogen	**EAC cases ($n = 81$)**	**Matched controls ($n = 160$)**	**OR (95% CI)[c]**	***P*[c]**	**ESCC cases ($n = 25$)**	**Matched controls ($n = 50$)**	**OR (95% CI)[c]**	***P*[c]**
Porphyromonas gingivalis	0.00[d] (23.5)	0.00 (25.0)	1.06 (0.93–1.20)	0.40	0.00 (32.0)	0.00 (20.0)	1.30 (0.96–1.77)	0.09
Tannerella forsythia	0.005 (56.8)	0.00 (47.5)	1.21 (1.01–1.46)	0.04	0.004 (52.0)	0.01 (58.0)	0.95 (0.58–1.55)	0.84
Treponema denticola	0.00 (39.5)	0.00 (37.5)	0.99 (0.83–1.17)	0.87	0.00 (20.0)	0.00 (44.0)	1.09 (0.72–1.66)	0.67

[a]Taxon raw counts were normalized with the clr transformation and used as predictors in conditional logistic regression models; models used matched sets as strata and adjusted for smoking status, BMI category, and alcohol drinking level.
[b]Percent of participants with presence of particular taxon in their oral cavity.
[c]Model parameters and *P* values were pooled over 10 models from 10 imputed datasets (missing values in BMI category and alcohol drinking level were imputed) using "mice" package, R.
[d]Zeros in table are true zeros, as when >50% of participants do not carry a taxon, the median relative abundance will be zero.

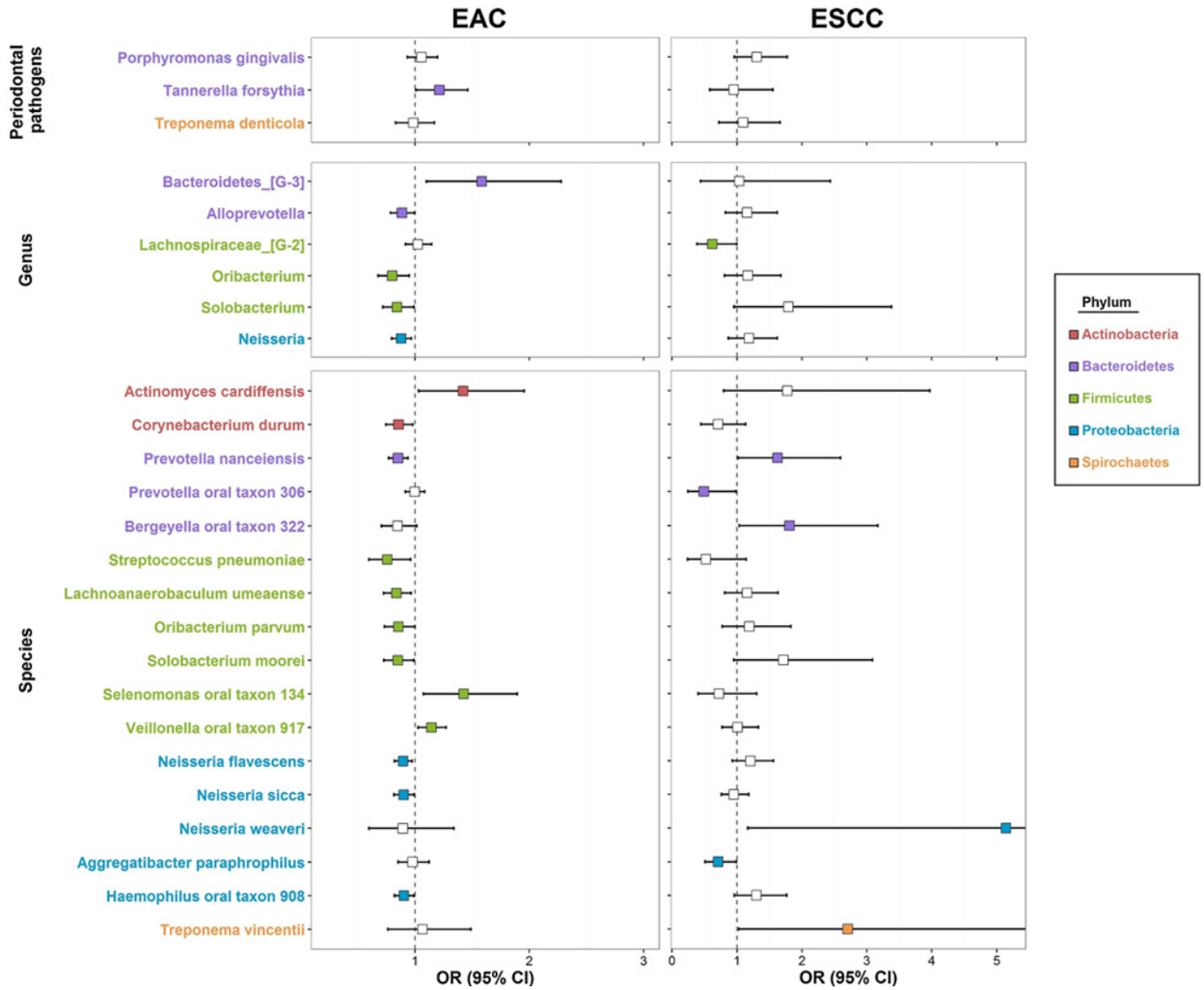

Figure 1.

Forest plot of ORs and 95% CI for associations of clr-transformed periodontal pathogen (*a priori*), genus, and species abundance with EAC and ESCC risk in conditional logistic regression models. See Tables 2 and 3 for numeric display of the OR (95% CI) estimates. Taxa names are colored by phylum; OR estimates are colored only if nominally statistically significant ($P < 0.05$).

nanceiensis, *Bergeyella oral taxon 322*, *Neisseria weaveri*, and *Treponema vincentii* was associated with higher ESCC risk, while increased abundance of *Prevotella oral taxon 306* and *Aggregatibacter paraphrophilus* was associated with lower ESCC risk (all $P < 0.05$). Additional adjustment for fruit/vegetable intake did not impact effect estimates (percent change in β-coefficient for all nominally significant taxa <11%). We did not perform stratified analysis of taxonomic findings for ESCC due to small sample size.

All of the species nominally associated with ESCC were associated with each other in an ecological network analysis (Fig. 2B). Interestingly, *Treponema vincentii*, which was associated with increased ESCC risk and has been previously associated with periodontal disease (34, 35), was linked to other periodontal pathogens (*P. gingivalis*, *T. forsythia*) in the ecological network.

Inferred metagenomic analysis

Analysis of inferred metagenomes revealed a number of metabolic pathways nominally associated with EAC risk (Table 4), although none reached the significance threshold after FDR adjustment. Increased abundance of endocytosis, sulfur relay system, biosynthesis of siderophore groups, and bisphenol degradation pathways was associated with higher EAC risk, and α-linolenic acid (ALA) metabolism and carotenoid biosynthesis pathways with lower risk (all $P < 0.05$). We did not identify any pathways associated with ESCC risk. Species *Neisseria sicca* and *Neisseria flavescens*, associated with reduced EAC risk, were positively correlated with the protective carotenoid biosynthesis and ALA metabolism pathways (Fig. 3).

Discussion

In this first prospective study of oral microbiota and esophageal cancer risk, we did not observe significant associations between overall microbiota diversity or composition and subsequent EAC or ESCC risk. However, several species were

Table 3. Oral taxa[a] associated with incident esophageal adenocarcinoma or squamous cell carcinoma, by phylum

	Adenocarcinoma				Squamous cell carcinoma			
	Median relative abundance (% carriage[b])				Median relative abundance (% carriage[b])			
Taxon (class; order; family; genus; species)	EAC cases ($n = 81$)	Matched controls ($n = 160$)	OR (95% CI)[c]	P^c	ESCC cases ($n = 25$)	Matched controls ($n = 50$)	OR (95% CI)[c]	P^c
Actinobacteria								
Actinobacteria; Actinomycetales (order)	8.04 (100)	6.99 (100)	1.34 (1.01–1.78)	0.05	7.89 (100)	5.82 (100)	0.94 (0.45–1.94)	0.86
Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces; cardiffensis (species)	0.00[d] (46.9)	0.00 (38.8)	1.42 (1.03–1.96)	0.03	0.00 (44.0)	0.00 (30.0)	1.77 (0.79–3.97)	0.17
Actinobacteria; Corynebacteriales; *Corynebacteriaceae; Corynebacterium; durum (species)*	0.02 (65.4)	0.07 (76.2)	0.86 (0.75–0.98)	0.03	0.09 (76.0)	0.11 (84.0)	0.71 (0.44–1.13)	0.15
Bacteroidetes								
Bacteroidetes C-1; Bacteroidetes O-1; Bacteroidetes F-1; Bacteroidetes G-3 (genus)	0.00 (19.8)	0.00 (15.6)	1.58 (1.10–2.28)	0.01	0.00 (20.0)	0.00 (20.0)	1.04 (0.44–2.44)	0.94
Bacteroidia; Bacteroidales; Prevotellaceae; Alloprevotella (genus)	0.47 (88.9)	0.56 (95.0)	0.89 (0.79–1.00)	0.05	0.64 (100)	0.73 (88.0)	1.15 (0.82–1.62)	0.41
Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella; nanceiensis (species)	0.07 (74.1)	0.21 (86.9)	0.85 (0.77–0.94)	0.001	0.55 (92.0)	0.18 (78.0)	1.63 (1.02–2.6)	0.04
Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella; oral taxon 306 (species)	0.05 (77.8)	0.08 (75.6)	1.00 (0.92–1.08)	0.93	0.02 (64.0)	0.17 (82.0)	0.49 (0.24–0.99)	0.05
Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bergeyella; oral taxon 322 (species)	0.07 (87.7)	0.08 (90.0)	0.85 (0.71–1.01)	0.07	0.10 (96.0)	0.06 (96.0)	1.81 (1.03–3.17)	0.03
Firmicutes								
Bacilli; Lactobacillales; Streptococcaceae; Streptococcus; pneumoniae (species)	0.09 (98.8)	0.10 (99.4)	0.76 (0.60–0.96)	0.02	0.08 (100)	0.14 (98.0)	0.52 (0.24–1.14)	0.10
Clostridia; Clostridiales; Lachnospiraceae XIV; Lachnoanaerobaculum; umeaense (species)	0.02 (72.8)	0.06 (81.2)	0.84 (0.73–0.97)	0.02	0.05 (84.0)	0.05 (78.0)	1.15 (0.81–1.63)	0.42
Clostridia; Clostridiales; Lachnospiraceae XIV; Lachnospiraceae G-2 (genus)	0.01 (63)	0.02 (56.9)	1.02 (0.92–1.14)	0.67	0.00 (44.0)	0.04 (76.0)	0.62 (0.38–0.99)	0.05
Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium (genus)	0.21 (90.1)	0.21 (96.2)	0.80 (0.68–0.95)	0.01	0.13 (92.0)	0.13 (88.0)	1.16 (0.81–1.68)	0.41
Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium; parvum (species)	0.00 (30.9)	0.00 (40.6)	0.85 (0.73–1.00)	0.05	0.01 (52.0)	0.004 (52.0)	1.19 (0.77–1.83)	0.43
Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium (genus)	0.04 (82.7)	0.08 (91.9)	0.84 (0.72–0.99)	0.04	0.07 (96.0)	0.07 (84.0)	1.79 (0.95–3.38)	0.07
Erysipelotrichia; Erysipelotrichales; *Erysipelotrichaceae; Solobacterium; moorei (species)*	0.04 (82.7)	0.08 (91.9)	0.85 (0.73–0.99)	0.04	0.07 (96.0)	0.07 (84.0)	1.71 (0.95–3.09)	0.08
Negativicutes; Selenomonadales; Veillonellaceae; Selenomonas; oral taxon 134 (species)	0.00 (45.7)	0.00 (31.9)	1.43 (1.07–1.89)	0.02	0.00 (24.0)	0.00 (40.0)	0.72 (0.40–1.30)	0.28
Negativicutes; Selenomonadales; Veillonellaceae; Veillonella; oral taxon 917 (species)	0.00 (35.8)	0.00 (18.8)	1.14 (1.03–1.27)	0.01	0.00 (28.0)	0.00 (20.0)	1.01 (0.77–1.33)	0.94
Proteobacteria								
Betaproteobacteria (class)	1.50 (96.3)	2.59 (96.9)	0.87 (0.78–0.97)	0.02	3.58 (96.0)	2.40 (98.0)	1.15 (0.80–1.64)	0.45
Betaproteobacteria; Neisseriales (order)	1.32 (96.3)	2.47 (96.9)	0.88 (0.79–0.98)	0.02	3.37 (96.0)	2.29 (98.0)	1.18 (0.84–1.67)	0.34
Betaproteobacteria; Neisseriales; Neisseriaceae (family)	1.32 (96.3)	2.47 (96.9)	0.88 (0.79–0.98)	0.02	3.37 (96.0)	2.29 (98.0)	1.19 (0.85–1.66)	0.32
Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria (genus)	1.20 (93.8)	2.42 (95.6)	0.88 (0.80–0.97)	0.01	3.23 (96.0)	2.13 (98.0)	1.19 (0.87–1.62)	0.29
Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; flavescens (species)	0.60 (85.2)	1.24 (92.5)	0.89 (0.82–0.98)	0.01	1.76 (96.0)	1.13 (96.0)	1.20 (0.93–1.56)	0.16
Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; sicca (species)	0.10 (75.3)	0.19 (85.0)	0.90 (0.81–0.99)	0.04	0.05 (88.0)	0.18 (88.0)	0.95 (0.76–1.18)	0.64
Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; weaveri (species)	0.00 (17.3)	0.00 (21.9)	0.89 (0.60–1.34)	0.59	0.00 (36.0)	0.00 (16.0)	5.14 (1.17–22.64)	0.03
Gammaproteobacteria; Pasteurellales; *Pasteurellaceae; Aggregatibacter; paraphrophilus (species)*	0.00 (42.0)	0.00 (41.2)	0.98 (0.85–1.12)	0.75	0.00 (28.0)	0.01 (56.0)	0.71 (0.51–0.99)	0.04
Gammaproteobacteria; Pasteurellales; *Pasteurellaceae; Haemophilus; oral taxon 908 (species)*	0.07 (69.1)	0.28 (81.9)	0.90 (0.82–0.99)	0.04	0.51 (84.0)	0.31 (74.0)	1.3 (0.96–1.76)	0.09
Spirochaetes								
Spirochaetia; Spirochaetales; Spirochaetaceae; Treponema; vincentii (species)	0.00 (16.0)	0.00 (18.1)	1.06 (0.76–1.49)	0.71	0.00 (20.0)	0.00 (26.0)	2.71 (1.03–7.14)	0.04

[a]Taxon raw counts were normalized with the clr transformation and used as predictors in conditional logistic regression models; models used matched sets as strata and adjusted for smoking status, BMI category, and alcohol drinking level. All taxa (classes, orders, families, genera, species) with $P < 0.05$ are included in the table. We did not observe phylum-level associations with EAC or ESCC risk.

[b]Percent of participants with presence of particular taxon in their oral cavity.

[c]Model parameters and *P* values were pooled over 10 models from 10 imputed datasets (missing values in BMI category and alcohol drinking level were imputed) using "mice" package, R.

[d]Zeros in table are true zeros, as when >50% of participants do not carry a taxon, the median relative abundance will be zero.

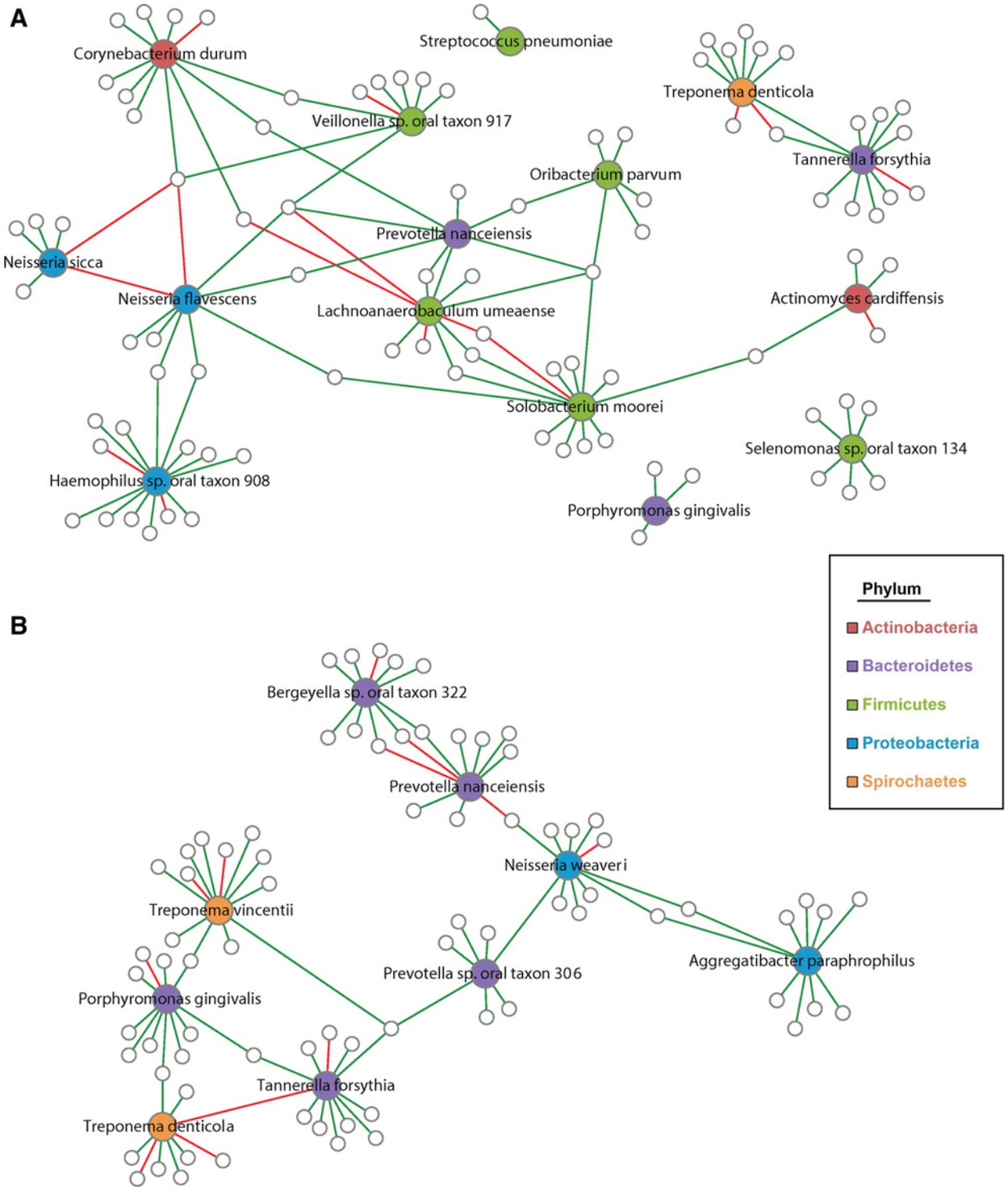

Figure 2.

Ecological networks among bacterial species associated with EAC or ESCC risk. The SPIEC-EASI algorithm (33) was used to infer microbial ecological networks. **A,** Algorithm was applied to EAC cases and matched controls ($n = 241$), and only networks related to EAC-associated species or *a priori* periodontal pathogens are shown. **B,** Algorithm was applied to ESCC cases and matched controls ($n = 75$), and only networks related to ESCC-associated species or *a priori* periodontal pathogens are shown. Species associated with EAC or ESCC are colored by phylum; other species in networks are indicated by small gray-outlined circles. Lines connecting species are colored by sign (positive, green; negative, red).

Table 4. KEGG pathways[a] associated with incident esophageal adenocarcinoma or squamous cell carcinoma

	Adenocarcinoma[b]		Squamous cell carcinoma[b]	
KEGG pathway	**OR (95% CI)[c]**	***P*[c]**	**OR (95% CI)[c]**	***P*[c]**
Cellular processes				
Meiosis – yeast	1.73 (1.10–2.70)	0.02	1.39 (0.50–3.86)	0.52
Endocytosis	1.46 (1.09–1.96)	0.01	0.80 (0.27–2.34)	0.68
Genetic information processing				
Sulfur relay system	5.21 (1.19–22.7)	0.03	0.49 (0.01–48.48)	0.76
Metabolism				
Glycosphingolipid biosynthesis - globo series	0.43 (0.18–0.99)	0.05	0.69 (0.29–1.68)	0.42
ALA metabolism	0.78 (0.64–0.95)	0.01	1.18 (0.73–1.90)	0.50
Porphyrin and chlorophyll metabolism	4.02 (1.23–13.15)	0.02	1.47 (0.17–12.99)	0.73
Biosynthesis of siderophore group nonribosomal peptides	2.03 (1.10–3.75)	0.02	0.79 (0.20–3.14)	0.73
Carotenoid biosynthesis	0.84 (0.70–1.00)	0.05	1.12 (0.70–1.81)	0.63
Bisphenol degradation	3.07 (1.30–7.24)	0.01	2.37 (0.35–16.16)	0.38

[a]KEGG pathway raw counts were normalized with the clr transformation and used as predictors in conditional logistic regression models; models used matching set as strata and adjusted for smoking status, BMI category, and alcohol drinking level. All pathways with $P < 0.05$ are included in the table.
[b]Adenocarcinoma includes 81 EAC cases and 160 matched controls, and squamous cell carcinoma includes 25 ESCC cases and 50 matched controls.
[c]Model parameters and *P* values were pooled over 10 models from 10 imputed datasets (missing values in BMI category and alcohol drinking level were imputed) using "mice" package, R.

nominally associated with risk, among them potential pathogens and also commensal species. Interestingly, bacterial taxon associations observed were unique to either EAC or ESCC, in line with the fundamentally different origins of these cancer types. We also show replication of several taxonomic findings in both the CPS-II and PLCO cohorts. Our biologically plausible findings warrant further investigation in larger studies, to fully explore prospects of modulating the oral microbiota for

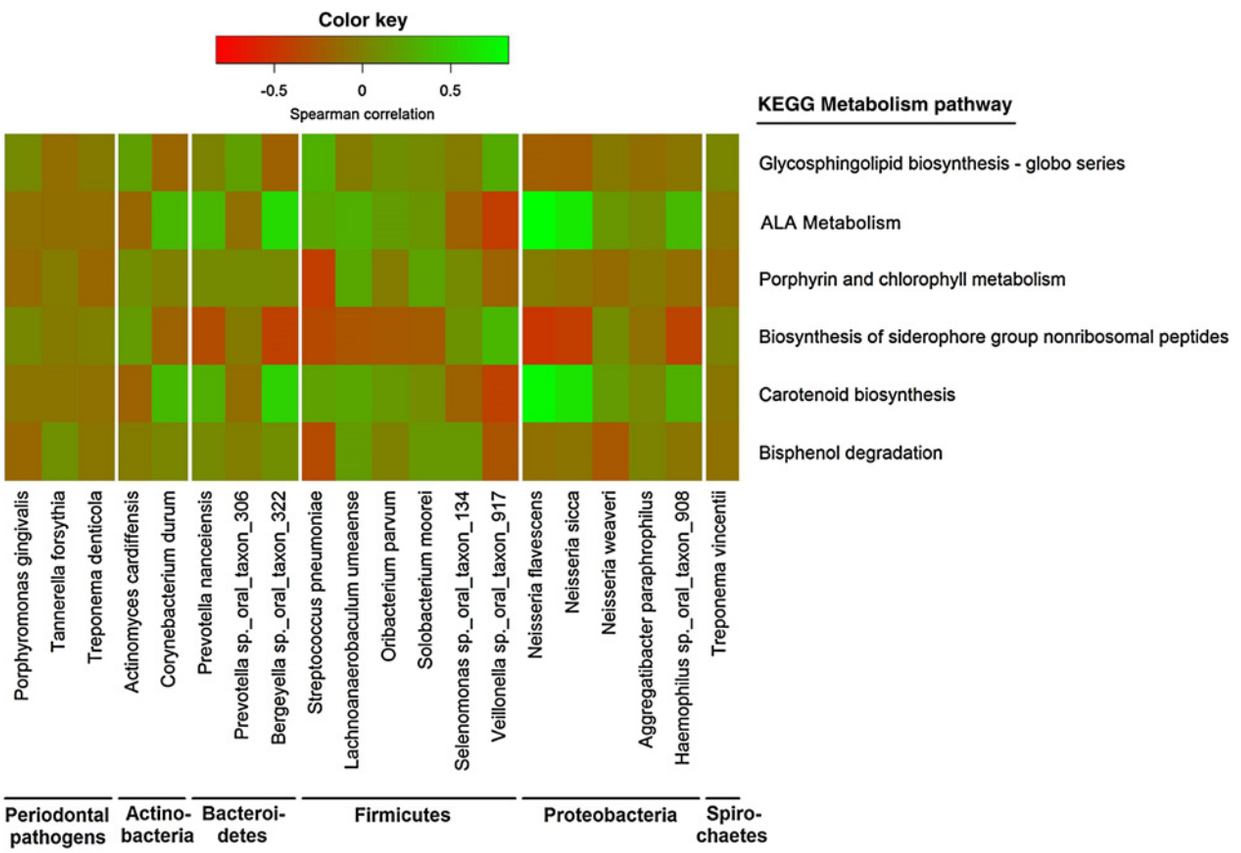

Figure 3.
Correlations of bacterial species and inferred metagenomic functions. Species and KEGG pathway counts were clr-transformed. Partial Spearman correlation coefficients were estimated for each pairwise comparison of species and KEGG pathway abundance, adjusting for age, sex, cohort, race, and smoking. Only KEGG pathways relating to metabolism, and periodontal pathogens or species associated with EAC or ESCC ($P < 0.05$) are included in the heatmap.

esophageal cancer prevention or utilizing it for risk stratification and early detection.

Studies of oral disease and cancer provide evidence that oral health (tooth loss, poor oral hygiene, and possibly periodontal disease) is linked to esophageal cancer risk (14, 36–38). We observed that *T. forsythia* was associated with higher EAC risk, and *P. gingivalis* with ESCC risk. These two species are members of the "red complex" of periodontal pathogens, that is, the species most strongly associated with severe periodontitis (30). A recent report revealed that *P. gingivalis* was detected at a higher rate in ESCC tumor tissue, compared with adjacent normal and healthy control mucosa; moreover, *P. gingivalis* presence was associated with ESCC lymph node metastasis and decreased survival time (39). More research is needed to determine whether periodontal disease and/or periodontal pathogens play a role in EAC/ESCC carcinogenesis, particularly as periodontal pathogen–EAC risk associations were inconsistent between the CPS-II and PLCO cohorts.

Several small studies have characterized the esophageal microbiota in relation to EAC (7, 11) or its precursors, GERD (7–9) and Barrett's esophagus (7–10). *Campylobacter* species were shown to dominate GERD and Barrett's esophagus biopsies compared with controls in two culture-based studies of subjects from the United Kingdom (7, 10). Yang and colleagues surveyed 16S rRNA genes from distal esophageal biopsies of 12 controls, 12 GERD patients, and 12 Barrett's esophagus patients in the United States (8); they observed a distinctly different microbial composition in GERD and Barrett's esophagus patients compared with controls, characterized by greater diversity, decreased *Streptococcus*, and increased abundance of Gram-negative anaerobes, including *Veillonella*, *Neisseria*, *Prevotella*, *Campylobacter*, *Porphyromonas*, *Fusobacterium*, and *Actinomyces*. Similarly, Japanese patients with Barrett's esophagus had decreased *Streptococcus* and increased *Veillonella*, *Neisseria*, and *Fusobacterium* in distal esophageal biopsies compared with controls (9). Finally, Zaidi and colleagues observed decreased abundance of *Streptococcus pneumoniae* in dysplastic, tumor-adjacent normal, and EAC biopsy samples compared with normal and Barrett's esophagus samples from U.S. patients (11). We observed an inverse association between *Streptococcus pneumoniae* and incident EAC, consistent with above-mentioned studies. In contrast to above-mentioned studies, we observed an inverse association of genus *Neisseria* with EAC risk. *Neisseria* species are oral cavity commensals (40), and we and others previously showed that oral *Neisseria* are depleted by cigarette smoking (41–43), a cause of EAC. Interestingly, we found that *Neisseria* were only associated with lower EAC risk in smokers (although interaction was not significant), possibly suggesting a joint effect of smoking and *Neisseria* depletion. Differing findings from previous literature may relate to differences in study design (cross-sectional vs. prospective) and sample origin (biopsy vs. oral).

Other studies have characterized the microbiota related to ESCC (13) and its precursor, ESD (12). Yu and colleagues observed that lower microbial richness and altered composition of upper digestive tract microbiota were associated with ESD in Chinese subjects (12). Likewise, Chen and colleagues reported differences in carriage and/or relative abundance of oral genera between 87 ESCC cases and 85 controls, including increased relative abundance of *Prevotella*, *Streptococcus*, and *Porphyromonas* in ESCC cases. These authors did not report findings at species level, making comparison with our mostly species-level findings for ESCC difficult.

Analysis of inferred metagenomes revealed several pathways associated with EAC, albeit not after FDR adjustment; some appeared biologically plausible. Bacterial carotenoid biosynthesis was associated with lower EAC risk, with *Neisseria* species potentially contributing to this protective pathway. Carotenoids are phytochemicals in fruits and vegetables, many acting as antioxidants (44). Higher fruit and vegetable intake and higher β-carotene intake have been associated with reduced EAC risk (4, 45). In addition, β-carotene therapy was shown to ameliorate GERD symptoms (46). Bacterial biosynthesis of siderophores (iron-chelating compounds) was associated with higher EAC risk. Although excessive iron may promote carcinogenesis (47) and iron chelation has been considered as a potential EAC therapy (48), iron is an essential trace element with deficiency leading to inflammation (49). Bacterial siderophore synthesis may upset iron homeostasis and thus might increase EAC risk. These inferred metagenomic functions provide insight into bacterial actions that may potentially impact EAC risk and warrant further investigation with full metagenomic sequencing.

Strengths of our study included the prospective design, comprehensive 16S rRNA gene sequencing, inclusion of two cohorts, and adjustment for EAC/ESCC risk factors throughout analysis. Our study also had several limitations. Lack of periodontal status of participants did not allow us to determine whether periodontal pathogens are implicated independently of periodontal disease. We also lacked data on presence of esophageal cancer precursor conditions (i.e., GERD and Barrett's esophagus) in the participants, which could mediate or confound oral microbiome-esophageal cancer associations and data on medications (e.g., proton-pump inhibitors, antibiotics), which could confound these associations (50, 51). In addition, although our study is the largest of its kind, case sample sizes ($n = 81$ EAC and $n = 25$ ESCC) remained small, limiting statistical power to detect FDR-adjusted significant associations, and our study population was mostly white, limiting generalizability.

In summary, we found evidence that specific bacterial pathogens may play a role in esophageal cancer risk, whereas other bacterial types may be associated with reduced risk. Larger studies are needed to confirm our findings, particularly among smokers and nonsmokers to clarify joint effects, followed by experimental animal models to clarify causal relationships. Identification of oral bacteria causal or protective in esophageal cancer could lead to interventions for their eradication or colonization in at-risk individuals. Continued study of oral microbiota in esophageal cancer may lead to actionable means for prevention of this highly fatal disease.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

Z. Pei is a staff physician at the Department of Veteran Affairs New York Harbor Healthcare System. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the U.S. Department of Veterans Affairs, or the United States Government.

Authors' Contributions

Conception and design: Z. Pei, R.B. Hayes, J. Ahn

Development of methodology: J. Wu, Z. Pei, L. Yang, M.P. Purdue, R.B. Hayes, J. Ahn

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.P. Purdue, E.J. Jacobs, S.M. Gapstur, R.B. Hayes, J. Ahn
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): B.A. Peters, N.D. Freedman, E.J. Jacobs, S.M. Gapstur, R.B. Hayes, J. Ahn
Writing, review, and/or revision of the manuscript: B.A. Peters, Z. Pei, M.P. Purdue, N.D. Freedman, E.J. Jacobs, S.M. Gapstur, R.B. Hayes, J. Ahn
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Z. Pei, N.D. Freedman, J. Ahn
Study supervision: Z. Pei, J. Ahn

Acknowledgments

This work was supported in part by the NCI at the NIH (R01CA159036, U01CA182370, R01CA164964, R03CA159414, P30CA016087, and R21CA183887). The NYU School of Medicine Genome Technology Center where samples were sequenced is partially supported by the Perlmutter Cancer Center support grant P30CA016087. The American Cancer Society (ACS) funds the creation, maintenance, and updating of the Cancer Prevention Study II cohort.

Samples were sequenced at the NYU School of Medicine Genome Technology Center.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 2, 2017; revised July 10, 2017; accepted August 28, 2017; published online December 1, 2017.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359–86.
2. Gupta B, Kumar N. Worldwide incidence, mortality and time trends for cancer of the oesophagus. Eur J Cancer Prev 2017;26:107–18.
3. Rubenstein JH, Shaheen NJ. Epidemiology, diagnosis, and management of esophageal adenocarcinoma. Gastroenterology 2015;149:302–17.
4. Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL, et al. Population attributable risks of esophageal and gastric cancers. J Natl Cancer Inst 2003;95:1404–13.
5. Yang L, Chaudhary N, Baghdadi J, Pei Z. Microbiome in reflux disorders and esophageal adenocarcinoma. Cancer J 2014;20:207–10.
6. Kauppila JH, Selander KS. Toll-like receptors in esophageal cancer. Front Immunol 2014;5:200.
7. Blackett KL, Siddhi SS, Cleary S, Steed H, Miller MH, Macfarlane S, et al. Oesophageal bacterial biofilm changes in gastro-oesophageal reflux disease, Barrett's and oesophageal carcinoma: association or causality? Aliment Pharmacol Ther 2013;37:1084–92.
8. Yang L, Lu X, Nossa CW, Francois F, Peek RM, Pei Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. Gastroenterology 2009;137:588–97.
9. Liu N, Ando T, Ishiguro K, Maeda O, Watanabe O, Funasaka K, et al. Characterization of bacterial biota in the distal esophagus of Japanese patients with reflux esophagitis and Barrett's esophagus. BMC Infect Dis 2013;13:130.
10. Macfarlane S, Furrie E, Macfarlane GT, Dillon JF. Microbial colonization of the upper gastrointestinal tract in patients with Barrett's esophagus. Clin Infect Dis 2007;45:29–38.
11. Zaidi AH, Kelly LA, Kreft RE, Barlek M, Omstead AN, Matsui D, et al. Associations of microbiota and toll-like receptor signaling pathway in esophageal adenocarcinoma. BMC Cancer 2015;16:52.
12. Yu G, Gail MH, Shi J, Klepac-Ceraj V, Paster BJ, Dye BA, et al. Association between upper digestive tract microbiota and cancer-predisposing states in the esophagus and stomach. Cancer Epidemiol Biomarkers Prev 2014;23:735–41.
13. Chen X, Winckler B, Lu M, Cheng H, Yuan Z, Yang Y, et al. Oral microbiota and risk for esophageal squamous cell carcinoma in a high-risk area of China. PLoS One 2015;10:e0143603.
14. Michaud DS, Kelsey KT, Papathanasiou E, Genco CA, Giovannucci E. Periodontal disease and risk of all cancers among male never smokers: an updated analysis of the Health Professionals Follow-up Study. Ann Oncol 2016;27:941–7.
15. Pei Z, Bini EJ, Yang L, Zhou M, Francois F, Blaser MJ. Bacterial biota in the human distal esophagus. Proc Natl Acad Sci U S A 2004;101:4250–5.
16. Lawson RD, Coyle WJ. The noncolonic microbiome: does it really matter? Curr Gastroenterol Rep 2010;12:259–62.
17. Hayes RB, Reding D, Kopp W, Subar AF, Bhat N, Rothman N, et al. Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. Control Clin Trials 2000;21(6 Suppl):349S–55S.
18. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer 2002;94:2490–501.
19. U.S. Department of Agriculture, U.S. Department of Health and Human Services. Dietary Guidelines for Americans, 2010. Washington, DC: U.S. Government Printing Office; 2010.
20. Belstrom D, Holmstrup P, Bardow A, Kokaras A, Fiehn NE, Paster BJ. Temporal stability of the salivary microbiota in oral health. PLoS One 2016;11:e0147472.
21. Zhou Y, Gao H, Mihindukulasuriya KA, La Rosa PS, Wylie KM, Vishnivetskaya T, et al. Biogeography of the ecosystems of the healthy human body. Genome Biol 2013;14:R1.
22. 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;326:1694–7.
23. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–6.
24. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database 2010;2010:baq013.
25. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;26:1641–50.
26. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environmental Microbiol 2007;73:1576–85.
27. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations. R. J Stat Softw 2011;45:1–67.
28. McArdle BH, Anderson MJ. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 2001;82:290–7.
29. Fernandes AD, Reid JNS, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome 2014;2:15.
30. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL Jr. Microbial complexes in subgingival plaque. J Clin Periodontol 1998;25:134–44.
31. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 2013;31:814–21.
32. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012;40:D109–14.
33. Kurtz ZD, Muller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol 2015;11:e1004226.

34. Dabu B, Mironiuc-Cureu M, Jardan D, Szmal C, Dumitriu S. Identification of four *Treponema* species in subgingival samples by nested-PCR and their correlation with clinical diagnosis. Roum Arch Microbiol Immunol 2012;71:43–7.
35. Willis SG, Smith KS, Dunn VL, Gapter LA, Riviere KH, Riviere GR. Identification of seven Treponema species in health- and disease-associated dental plaque by nested PCR. J Clin Microbiol 1999;37:867–9.
36. Hiraki A, Matsuo K, Suzuki T, Kawase T, Tajima K. Teeth loss and risk of cancer at 14 common sites in Japanese. Cancer Epidemiol Biomarkers Prev 2008;17:1222–7.
37. Abnet CC, Qiao YL, Mark SD, Dong ZW, Taylor PR, Dawsey SM. Prospective study of tooth loss and incident esophageal and gastric cancers in China. Cancer Causes Control 2001;12:847–54.
38. Guha N, Boffetta P, Wunsch Filho V, Eluf Neto J, Shangina O, Zaridze D, et al. Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies. Am J Epidemiol 2007;166:1159–73.
39. Gao S, Li S, Ma Z, Liang S, Shan T, Zhang M, et al. Presence of *Porphyromonas gingivalis* in esophagus and its association with the clinicopathological characteristics and survival in patients with esophageal cancer. Infect Agent Cancer 2016;11:3.
40. Liu G, Tang CM, Exley RM. Non-pathogenic Neisseria: members of an abundant, multi-habitat, diverse genus. Microbiology 2015;161:1297–312.
41. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, et al. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J 2016;10:2435–46.
42. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med 2013;187:1067–75.
43. Colman G, Beighton D, Chalk AJ, Wake S. Cigarette smoking and the microbial flora of the mouth. Aust Dent J 1976;21:111–8.
44. Liu RH. Health-promoting components of fruits and vegetables in the diet. Adv Nutr 2013;4:384s–92s.
45. Kubo A, Corley DA. Meta-analysis of antioxidant intake and the risk of esophageal and gastric cardia adenocarcinoma. Am J Gastroenterol 2007;102:2323–30.
46. Dutta SK, Agrawal K, Girotra M, Fleisher AS, Motevalli M, Mah'moud MA, et al. Barrett's esophagus and beta-carotene therapy: symptomatic improvement in GERD and enhanced HSP70 expression in esophageal mucosa. Asian Pac J Cancer Prev 2012;13:6011–6.
47. Huang X. Iron overload and its association with cancer risk in humans: evidence for iron as a carcinogenic metal. Mutat Res 2003;533:153–71.
48. Keeler BD, Brookes MJ. Iron chelation: a potential therapeutic strategy in oesophageal cancer. Br J Pharmacol 2013;168:1313–5.
49. Ganz T, Nemeth E. Iron homeostasis in host defence and inflammation. Nat Rev Immunol 2015;15:500–10.
50. Freedberg DE, Lebwohl B, Abrams JA. The impact of proton pump inhibitors on the human gastrointestinal microbiome. Clin Lab Med 2014;34:771–85.
51. Abeles SR, Jones MB, Santiago-Rodriguez TM, Ly M, Klitgord N, Yooseph S, et al. Microbial diversity in individuals and their household contacts following typical antibiotic courses. Microbiome 2016;4:39.