British Journal of Cancer www.nature.com/bjc

ARTICLE

Epidemiology

The oral microbiome in relation to pancreatic cancer risk in African Americans

Jessica L. Petrick[1✉], Jeremy E. Wilkinson[2], Dominique S. Michaud[3], Qiuyin Cai[4], Hanna Gerlovin[1], Lisa B. Signorello[5], Brian M. Wolpin[6], Edward A. Ruiz-Narváez[7], Jirong Long[4], Yaohua Yang[4], W. Evan Johnson[8], Xiao-Ou Shu[4], Curtis Huttenhower[2,9] and Julie R. Palmer[1✉]

BACKGROUND: African Americans have the highest pancreatic cancer incidence of any racial/ethnic group in the United States. The oral microbiome was associated with pancreatic cancer risk in a recent study, but no such studies have been conducted in African Americans. Poor oral health, which can be a cause or effect of microbial populations, was associated with an increased risk of pancreatic cancer in a single study of African Americans.

METHODS: We prospectively investigated the oral microbiome in relation to pancreatic cancer risk among 122 African-American pancreatic cancer cases and 354 controls. DNA was extracted from oral wash samples for metagenomic shotgun sequencing. Alpha and beta diversity of the microbial profiles were calculated. Multivariable conditional logistic regression was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for associations between microbes and pancreatic cancer risk.

RESULTS: No associations were observed with alpha or beta diversity, and no individual microbial taxa were differentially abundant between cases and control, after accounting for multiple comparisons. Among never smokers, there were elevated ORs for known oral pathogens: *Porphyromonas gingivalis* (OR = 1.69, 95% CI: 0.80–3.56), *Prevotella intermedia* (OR = 1.40, 95% CI: 0.69–2.85), and *Tannerella forsythia* (OR = 1.36, 95% CI: 0.66–2.77).

CONCLUSIONS: Previously reported associations between oral taxa and pancreatic cancer were not present in this African-American population overall.

British Journal of Cancer (2022) 126:287–296; https://doi.org/10.1038/s41416-021-01578-5

INTRODUCTION

African Americans have the highest incidence and mortality from pancreatic cancer of any racial/ethnic group in the United States (US). Pancreatic cancer is rare, yet its poor prognosis (5-year survival of 9.0%) underscores the need for identifying modifiable risk factors and early detection strategies [1].

Poor oral health [2–8] has been implicated as a risk factor for pancreatic cancer. Increased pancreatic cancer risk has been consistently associated with periodontal disease or tooth loss, which in turn correspond with changes in the microbiota—the community of microorganisms, including bacteria, fungi, viruses and archaea, that reside within the oral cavity [9]. In our prior investigation of pancreatic cancer in African-American women, adult tooth loss was associated with a 2-fold increased pancreatic cancer risk [2]. In addition, a recent prospective study of the relation between the oral microbiome and pancreatic cancer risk reported associations of a few specific oral pathogens (i.e. *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans*) with an increased risk [10]. Most study participants were White, prohibiting analysis by race/ethnicity. It utilised bacterial 16S ribosomal RNA (rRNA) gene sequencing, which does not allow for characterisation of the functional potential of the microbiota or species-level identification and only identifies bacteria.

African Americans are reported to have worse oral health profiles than other racial/ethnic groups in the US, largely due to socioeconomic disadvantage and racial discrimination in the health care system, which can reduce access to preventive oral health care [11]. An individual's oral microbiome is affected by factors such as periodontal disease and tooth loss, and the reverse association may also occur. Here, we prospectively assessed the association of the oral microbiome with pancreatic risk in African-American women and men.

[1]Slone Epidemiology Center, Boston University, Boston, MA, USA. [2]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, USA. [3]Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, USA. [4]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA. [5]Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA. [6]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. [7]Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, MI, USA. [8]Department of Medicine, Division of Computational Biomedicine, Boston University, Boston, MA, USA. [9]Broad Institute of MIT and Harvard, Cambridge, MA, USA. ✉email: jpetrick@bu.edu; jpalmer@bu.edu

Received: 2 December 2020 Revised: 14 September 2021 Accepted: 1 October 2021
Published online: 30 October 2021

MATERIALS AND METHODS

Study populations

The current study leveraged resources from two prospective cohort studies: Black Women's Health Study (BWHS) and Southern Community Cohort Study (SCCS). All study participants provided written informed consent. The Institutional Review Boards of Boston University (BWHS) and Vanderbilt University Medical Center and Meharry Medical College (SCCS) approved the studies.

The BWHS enrolled 59,000 women aged 21–69 years in 1995 by mailing scannable questionnaires to subscribers of *Essence* magazine [12]. Participants, who reside in all regions of the mainland US, complete follow-up questionnaires every 2 years, either online or by mail.

SCCS participants reside in a 12-state area in the southeastern US [13, 14]. Between 2002 and 2009, 55,362 African-American and 25,305 White men and women aged 40–79 years were enrolled via in-person recruitment at 71 community health centres (86%) or through mailed questionnaires (14%). Follow-up mail questionnaires are sent to participants every 5–6 years. Both the SCCS and BWHS are ongoing.

Cases were BWHS or SCCS participants diagnosed with incident primary pancreatic adenocarcinoma (consistent with the International Classification of Diseases 10 [ICD-10] topography codes C25.0–25.9 and morphology code 8140) during follow-up, who also had a stored, prediagnostic oral wash specimen. The majority of pancreatic cancer cases were identified by cancer registries or National Death Index. Self-report of cancer on follow-up questionnaires was confirmed by a review of hospital and cancer registry data. Participants who reported prevalent pancreatic cancer at study baseline or a diagnosis within the first 2 years after providing an oral wash sample were excluded. Thus, this study included 148 pancreatic cancer cases ($n = 53$ from BWHS and $n = 95$ from SCCS), among whom 122 were African American and 26 White.

Within each cohort, controls were individually matched 3:1 to cases using incidence-density sampling on the following criteria: age (5-year age groups), smoking status (never, former, current), timing of oral wash collection (±12 months) and for SCCS only, sex, race, and community health centre/mail enrolment. For the main analyses, we restricted to African-American cases ($n = 122$) and controls ($n = 354$). In secondary analyses, we used the full dataset of 148 cases and 441 controls and stratified on race.

Oral wash samples

Oral wash sample collection in BWHS and SCCS has previously been described [15, 16]. Briefly, BWHS participants were mailed a mouthwash collection kit and instructed to take a mouthful of Mint Fresh Scope® (at least ½ hour after eating or drinking), swish vigorously for 45 s and spit the sample into a screw-top polypropylene jar. Subjects were asked to record the time and date of oral rinse on the instruction sheet and to mail the instruction sheet to the laboratory with the sample. All BWHS samples were returned via first class mail and processed on the day of receipt. Similar methods and materials were used for SCCS, except that the samples were collected in person and taken directly to the laboratory for processing.

Similar methods were used by the SCCS and BWHS to process and store samples. Buccal cells from oral wash samples were centrifuged and suspended using TE buffer, aliquoted as a pellet into a 2 ml vial, and stored at −80 °C. For SCCS, samples were obtained at study enrolment (2002–2009) from ~40% of participants [17, 18]. For BWHS, samples were obtained from 2004 to 2007 from ~50% of participants [19]. The median interval from the provision of sample to pancreatic cancer diagnosis was 7 years in both studies.

Metagenomic shotgun sequencing

DNA was extracted using the PowerSoil Pro Kit (MoBio Laboratories Inc.), as this has been shown to increase the ratio of bacterial and fungal to human DNA extracted [20]. Using a minimum of 300 ng of DNA, paired-end whole-metagenome shotgun sequencing was performed at BGI Genomics (Shenzhen, China), using the Illumina HiSeq2000 platform with a read length of 100 bp (insert size 350 bp).

Bioinformatic processing of metagenomic data

For BWHS and SCCS, raw sequencing reads were processed by the Harvard TH Chan School of Public Health Microbiome Analysis Core using the bioBakery shotgun metagenomic workflow v0.13.1 [21]. Briefly, raw sequencing reads were run through KneadData, which trims data to remove low-quality reads and human contamination from the samples using a human (hg38) and known sequencing contaminant databases. This retained an average of ~1.6 Gb of high-quality non-human sequences per sample. Next, the marker gene-based approach from MetaPhlAn2 was used to assign taxonomy to each sample [22]. Functional profiling was performed using HUMAnN2 to provide taxon-specific profiles of UniRef gene families, enzymes and MetaCyc pathways [23].

Quality control

Quality control was performed on the taxonomic profiles of the samples in comparison to negative controls (i.e. extraction blanks) to ensure the validity of the sequencing results. In brief, reads were detected in the negative controls [24]. However, the number of reads from the negative controls were orders of magnitude lower than from the study samples. We examined the taxa in the negative controls and determined that they consisted solely of typical contaminant microbes not observed in our study samples. This, coupled with the low read counts in the negative controls, allowed us to conclude that noticeable reagent contamination was not present (and would not need to be removed from the study samples). Further, it would be extremely unusual for reagent contamination to be a substantial problem in a study like this, as none of our samples were low biomass [25]. Laboratory personnel were blinded to case/control status; matched sets were processed within the same batch. Thus, any laboratory variation would be accounted for in the matching factors.

Statistical analysis

Alpha diversities were calculated based on species as Shannon, Simpson, and inverse Simpson indices; richness analysis was calculated based on species as Chao1 index and beta diversities were calculated based on species and DNA pathways as Bray–Curtis dissimilarity to examine within- and between-sample differences in microbial richness and composition [26]. Principal coordinate analysis plots were created using the Bray–Curtis dissimilarities [27]. Between-group differences in microbial composition (Bray–Curtis dissimilarities and Jaccard distance) were also assessed using permutational multivariate analysis of variance. For all analyses, except alpha diversity, microbial features (species and pathways) were filtered requiring a microbial feature to have at least 0.01% relative abundance in at least 10% of all samples.

Seven periodontal bacteria were specified a priori for examination—*A. actinomycetemcomitans*, *Filifactor alocis*, *Fusobacterium nucleatum*, *P. gingivalis*, *Prevotella intermedia*, *Tannerella forsythia* and *Treponema denticola* [10, 28, 29]. For fungi, one phylum (*Ascomycota*) and one genus within that phylum (*Saccharomyces*) was inferred from the gene family relative abundances [30].

Count data were normalised using relative abundance calculations and log-transformed. The taxon variables examined include $\log_{10}$ relative abundances, prevalence (carrier/non-carrier status), and high, low or non-carrier status (categorised using the mean relative abundance of the control group as the threshold for high or low carriers).

To examine bacteria that were specified a priori, conditional logistic regression was used to calculate odds ratios (ORs) and corresponding 95% confidence intervals (CIs) for the association between microbes and pancreatic cancer risk [31]. Data for covariates were taken from the same time period/questionnaire cycle as the oral wash sample. Because smoking is an important risk factor, we additionally controlled for intensity and duration of smoking by including a term for <20 or ≥20 pack-years in the regression models. Other potential covariates considered were body mass index (kg/m^2), type 2 diabetes, alcohol intake, red meat intake [32, 33], and tooth loss. If the log(OR) changed by ≥10% due to variable elimination, the variable was considered a confounder and retained in the model; [31] none of the additional covariates met this criterion. Based on aetiologic evidence, which suggests differences in the oral microbiome by smoking status [34–36], and for comparison with the prior report [10], we conducted analyses stratified by smoking status, as well as by sex and by race.

Featurewise analysis was conducted using MaAsLin2 on $\log_{10}$ relative abundances and pathways to determine if there were differentially abundant microbes or pathways. All p values are two-sided [37]. Multiple comparisons were adjusted for using the q value for species and pathway-level classifications, based on the Benjamini–Hochberg procedure for false discovery rate control at the 0.05 level for 161 species-level comparisons and 266 DNA pathway comparisons [38]. Analyses were performed in R version 4.0.0 (Vienna, Austria).

Table 1. Demographic factors among African-American pancreatic cancer cases and controls.

	Cases ($n = 122$)	Controls ($n = 354$)
Age, years (mean ± SD)	59.5 ± 9.3	59.3 ± 9.1
Sex (*n*, %)		
Female	94 (77.0)	274 (77.4)
Male	28 (23.0)	80 (22.6)
Smoking status (*n*, %)		
Never	43 (35.2)	131 (37.0)
Former	33 (27.0)	94 (26.6)
Current	46 (37.7)	129 (37.7)
≥20 pack-years[a]	28 (23.0)	71 (20.1)
Body mass index (*n*, %)		
<25 kg/m^2	25 (20.5)	72 (20.3)
25–29 kg/m^2	48 (39.3)	127 (35.9)
30–34 kg/m^2	23 (18.9)	75 (21.2)
≥35 kg/m^2	26 (21.3)	80 (22.6)
Diabetes diagnosis (%)		
No	83 (68.0)	261 (73.7)
Yes	39 (32.0)	92 (26.0)
Missing	0 (0.0)	1 (0.3)
Alcohol intake (%)		
Non-drinker	68 (55.7)	197 (55.6)
1–3 drinks/week	37 (30.3)	109 (30.8)
≥4 drinks/week	17 (13.9)	48 (13.6)
Red meat intake, g/day (*n*, %)		
Quartile 1	23 (18.9)	93 (26.3)
Quartile 2	32 (26.2)	88 (24.9)
Quartile 3	31 (25.4)	75 (21.2)
Quartile 4	29 (23.8)	80 (22.6)
Missing	7 (5.7)	18 (5.1)
Tooth loss (*n*, %)		
None	12 (9.8)	51 (14.4)
1–4 teeth	23 (18.9)	58 (16.4)
5–10 teeth	32 (26.2)	42 (11.9)
>10–<32 teeth	6 (4.9)	31 (8.8)
All teeth	9 (7.4)	20 (5.6)
Missing	40 (32.8)	152 (42.9)
Gingivitis (*n*, %)		
No	74 (60.7)	197 (55.6)
Yes	14 (11.5)	40 (11.3)
Missing	34 (27.9)	117 (33.1)

[a]Among current and past smokers.

RESULTS

As expected for the oral cavity, the most abundant phylum, genus and species were *Firmicutes*, *Streptococcus*, and *Rothia mucilaginosa*, respectively. Unlike the prior study [10], microbial taxonomic profiles were similar between cases and controls (Supplemental Fig. S1). There were no differences in alpha or beta diversity or richness between cases and controls for taxa or DNA pathways (Supplemental Figs. S2–S5). Cases were more likely than controls to have diabetes at baseline (32.0 vs. 26.0%), consume levels of red meat above the median (49.2 vs. 43.8%), and have any teeth

Table 2. Associations of periodontal bacteria with risk of pancreatic cancer among African Americans.

Bacterial microbes	Cases	Controls	OR[a]	95% CI[a]
Aggregatibacter actinomycetemcomitans				
Non-carrier	116	331	1.00	
Carrie	6	23	0.74	(0.29–1.90)
Low carrier	4	12	0.95	(0.29–3.15)
High carrier	2	11	0.51	(0.11–2.38)
Log_{10} relative abundance[b]			–	
Filifactor alocis				
Non-carrier	67	198	1.00	
Carrier	55	156	1.08	(0.70–1.67)
Low carrier	33	76	1.31	(0.79–2.15)
High carrier	22	80	0.79	(0.43–1.45)
Log_{10} relative abundance[b]			0.78	(0.45–1.38)
Fusobacterium nucleatum				
Non-carrier	14	42	1.00	
Carrier	108	312	1.06	(0.54–2.09)
Low carrier	58	159	1.09	(0.54–2.21)
High carrier	50	153	1.03	(0.49–2.16)
Log_{10} relative abundance[b]			0.92	(0.69–1.23)
Porphyromonas gingivalis				
Non-carrier	46	134	1.00	
Carrier	76	220	1.04	(0.66–1.64)
Low carrier	37	106	1.05	(0.63–1.74)
High carrier	39	114	1.03	(0.59–1.79)
Log_{10} relative abundance[b]			1.17	(0.86–1.61)
Prevotella intermedia				
Non-carrier	59	167	1.00	
Carrier	63	187	0.97	(0.63–1.52)
Low carrier	29	93	0.90	(0.53–1.55)
High carrier	34	94	1.04	(0.62–1.78)
Log_{10} relative abundance[b]			1.30	(0.84–2.02)
Tannerella forsythia				
Non-carrier	38	115	1.00	
Carrier	84	239	1.09	(0.69–1.72)
Low carrier	41	120	1.05	(0.63–1.76)
High carrier	43	119	1.14	(0.66–1.97)
Log_{10} relative abundance[b]			0.88	(0.61–1.27)
Treponema denticola				
Non-carrier	60	182	1.00	
Carrier	62	172	1.12	(0.73–1.72)
Low carrier	38	86	1.35	(0.83–2.20)
High carrier	24	86	0.83	(0.46–1.49)
Log_{10} relative abundance[b]			0.75	(0.47–1.21)

[a]Odds ratios (ORs) and 95% confidence intervals (CIs), conditioned on matched sets (matched on: cohort [BWHS, SCCS], batch, age (5-year age groups), smoking status (never, former, current), timing of oral wash collection (±12 months) and for SCCS only, sex, race, and community health centre/mail enrolment) and adjusted for smoking intensity (<20, ≥20 pack-years).
[b]Calculated within carriers.

Table 3. Associations of periodontal bacteria with risk of pancreatic cancer, stratified by smoking status, among African Americans.

Bacterial microbes	Never smoker				Ever smoker			
	Cases	Controls	OR[a]	95% CI[a]	Cases	Controls	OR[a]	95% CI[a]
Aggregatibacter actinomycetemcomitans								
Non-carrier	41	123	1.00		75	208	1.00	
Carrier	2	8	0.75	(0.16–3.53)	4	15	0.73	(0.22–2.39)
Log_{10} relative abundance[b]			–				–	
Filifactor alocis								
Non-carrier	25	81	1.00		42	117	1.00	
Carrier	18	50	1.25	(0.62–2.53)	37	106	0.98	(0.56–1.71)
Log_{10} relative abundance[b]			1.01	(0.36–2.78)			0.73	(0.37–1.45)
Fusobacterium nucleatum								
Non-carrier	5	11	1.00		9	31	1.00	
Carrier	38	120	0.80	(0.26–2.45)	70	192	1.25	(0.54–2.90)
Log_{10} relative abundance[b]			0.73	(0.47–1.13)			1.09	(0.72–1.66)
Porphyromonas gingivalis								
Non-carrier	14	55	1.00		32	79	1.00	
Carrier	29	76	1.69	(0.80–3.56)	47	144	0.79	(0.44–1.41)
Log_{10} relative abundance[b]			1.30	(0.83–2.03)			0.97	(0.63–1.50)
Prevotella intermedia								
Non-carrier	19	67	1.00		40	100	1.00	
Carrier	24	64	1.40	(0.69–2.85)	39	123	0.76	(0.43–1.35)
Log_{10} relative abundance[b]			1.56	(0.75–3.26)			1.14	(0.64–2.04)
Tannerella forsythia								
Non-carrier	14	51	1.00		24	64	1.00	
Carrier	29	80	1.36	(0.66–2.77)	55	159	0.91	(0.50–1.67)
Log_{10} relative abundance[b]			0.94	(0.51–1.75)			0.85	(0.54–1.34)
Treponema denticola								
Non-carrier	28	82	1.00		32	100	1.00	
Carrier	15	49	0.93	(0.46–1.88)	47	123	1.23	(0.71–2.13)
Log_{10} relative abundance[b]			1.70	(0.40–7.22)			0.64	(0.38–1.10)

[a]Odds ratios (ORs) and 95% confidence intervals (CIs), conditioned on matched sets (matched on: cohort [BWHS, SCCS], batch, age (5-year age groups), smoking status (never, former, current), timing of oral wash collection (±12 months) and for SCCS only, sex, race, and community health centre/mail enrolmentage (5-year age groups), smoking status (never, former, current), timing of oral wash collection (±12 months) and for SCCS only, sex, race, and community health centre/mail enrolment) and adjusted for smoking intensity (<20, ≥20 pack-years).
[b]Calculated within carriers.

missing (57.4 vs. 42.7%; Table 1). Demographics of the White participants from SCCS are provided in Supplemental Table S1.

No individual microbial taxa—bacteria, fungi, viruse, or archaea—were differentially abundant between African-American cases and controls (Supplemental Table S2) or the full set of cases and controls (Supplemental Table S3), after accounting for multiple comparisons. Further, little of the beta-diversity variance was explained by case–control status (Supplemental Table S4). In the seven bacterial microbes that were targeted a priori based on previous study findings, there was little to no association with pancreatic cancer risk among African Americans overall (Table 2). For example, carriers of *P. gingivalis* had no increased pancreatic cancer risk (OR = 1.04, 95% CI: 0.66–1.64), compared to non-carriers. However, among never smokers (Table 3), ORs were elevated for carriers of *P. gingivalis* (OR = 1.69, 95% CI: 0.80–3.56), *P. intermedia* (OR = 1.40, 0.69–2.85), and *T. forsythia* (OR = 1.36, 95% CI: 0.66–2.77).

Fungal microbes also showed little association between carrier status and pancreatic cancer risk (Table 4), although the OR for carriers of *Saccharomyces* at high levels (i.e. above the median relative abundance) was 1.87 (95% CI: 0.81–4.35). Results were similar by smoking status (data not shown).

In analyses stratified by race (Supplemental Table S5), the OR for *P. gingivalis* in relation to pancreatic cancer risk was 1.68 (95% CI: 0.60–4.68) in Whites, compared to a null association among African Americans. Other ORs were similar across races. In analyses stratified by sex (Supplemental Table S6), all ORs for microbial taxa were close to 1.0 among women. Among men, elevated ORs for carriers relative to noncarriers were observed for *A.*

Table 4. Associations of periodontal fungi with risk of pancreatic cancer among African Americans.

Fungal microbes	Cases	Controls	OR[a]	95% CI[a]
Ascomycota				
Non-carrier	58	161	1.00	
Carrier	64	193	0.93	(0.61–1.42)
Low carrier	29	102	0.79	(0.47–1.32)
High carrier	35	91	1.11	(0.67–1.85)
Log_{10} relative abundance[b]			0.89	(0.62–1.28)
Saccharomyces				
Non-carrier	107	312	1.00	
Carrier	15	42	1.10	(0.57–2.13)
Low carrier	4	24	0.54	(0.18–1.62)
High carrier	11	18	1.87	(0.81–4.35)
Log_{10} relative abundance[b]			–	

[a]Odds ratios (ORs) and 95% confidence intervals (CIs), conditioned on matched sets (matched on: cohort [BWHS, SCCS], batch, age (5-year age groups), smoking status (never, former, current), timing of oral wash collection (±12 months) and for SCCS only, sex, race, and community health centre/mail enrolmentage (5-year age groups), smoking status (never, former, current), timing of oral wash collection (±12 months) and for SCCS only, sex, race, and community health centre/mail enrolment) and adjusted for smoking intensity (< 20, ≥ 20 pack-years).
[b]Calculated within carriers.

actinomycetemcomitans (OR = 2.33, 95% CI: 0.55–9.79), *T. denticola* (OR = 1.73, 95% CI: 0.58–5.14), *P. intermedia* (OR = 1.56, 95% CI: 0.52–4.68), *F. alocis* (OR = 1.34, 95% CI: 0.53–3.41), and *T. forsythia* (OR = 1.30, 95% CI: 0.35–4.87).

Functional profiling was assessed for all phenotypes and strata: 28 MetaCyc pathways were nominally differential between cases and controls, but effect sizes were small and q values were large adjusting for multiple comparisons ($0.25 > q > 0.05$; Table 5 and Supplemental Tables S7 and S8). The majority of these metabolic pathways include the expected taxonomic range of bacteria and are involved in biosynthesis processes characteristic of the oral cavity (e.g. aerobic sugar metabolism).

DISCUSSION

In the present study of African Americans, there was little to no association between carriage of periodontal bacterial or fungal microbes and pancreatic cancer risk. However, among never smokers, there was evidence of an association of several known oral bacterial pathogens (i.e. *P. gingivalis*, *P. intermedia*, *T. forsythia*) with elevated pancreatic cancer risk, with increases ranging from 36 to 69%.

These results are in contrast to several case–control studies, which have reported differences in oral microbial composition and diversity, as detected by 16S rRNA sequencing, between pancreatic cancer cases and controls [39–41]. Another case–control study reported that relative abundance of *Lactobacillus* was lower, and *Fusobacteria* higher, in tissue samples from pancreatic cancer cases compared to non-cases [42]. In a study of the gut microbiome, 14 bacterial features that discriminated between pancreatic cancer cases and controls were identified [43]. However, these studies are susceptible to reverse causation, as oral, faecal or tissue samples were obtained at, or after, the time of cancer diagnosis.

The present study is the first study of the oral microbiome in relation to pancreatic cancer risk in an African-American population. The results differed from previously reported results in a White population, which was based on combined data from two other prospective cohort studies [10]. In that study, *P. gingivalis* and *A. actinomycetemcomitans* were associated with increased pancreatic cancer risk and *Fusobacteria* with a decreased risk [10]. Stronger associations were observed among never smokers. In addition, in a prospective study from Europe, higher antibody levels to *P. gingivalis* were associated with a 2-fold increased pancreatic cancer risk [44].

The most likely explanations for the difference in findings are (1) sampling variation and small numbers in both studies or (2) differences in important characteristics of the study populations, or some combination of the two. The study populations differed in two major respects: race, which is a marker for shared cultural experiences, including the legacy of racism and racial discrimination, and prevalence of cigarette smoking.

Although the current report is focused on African Americans, we also analysed data from the smaller proportion of available White participants. Of note, among the 26 White pancreatic cancer cases and 87 White controls, carriage of *P. gingivalis* was associated with a 68% increased risk. This estimate is nearly identical to results from Fan et al., in which the OR for carriage of *P. gingivalis* was 1.60 [10]. Results from White participants in the present study were also similar to results for *Fusobacteria* and *F. nucleatum* in the prior study [10].

With regard to cigarette smoking, 38% of African-American participants in the present study were current smokers, whereas in the prior study [10], only 7% of participants were current smokers. We observed evidence of positive associations for several microbes, including *P. gingivalis*, among never smokers. Never smokers are an ideal group in which to assess associations of the oral microbiome with pancreatic cancer risk, given that cigarette smoking has an independent association with pancreatic cancer risk and also has an impact on oral health. It is possible that the high proportion of current smokers in our study population made it more difficult to detect an association, if present, as smoking reduces the host response to oral pathogens, such as *P. gingivalis* [45]. While smokers are more likely to have periodontitis, smokers paradoxically have reduced markers of clinical inflammation [46, 47]. Experimental studies have demonstrated that *P. gingivalis* cells in oral biofilms grown in the presence of extracts from cigarette smoke exhibit a lower proinflammatory capacity (i.e. lower levels of tumour necrosis factor-α, interleukin-6 and interleukin-12) than control oral biofilms [45, 48, 49]. This would not have been an issue in the prior study since 93% of participants were not current smokers.

Microbiome data are inherently compositional, which necessitates careful consideration of statistical approaches. However, if the use of compositionally corrected methods is necessary—or if these methods result in improved model performance—are a source of debate [50–57]. Recent findings suggest that compositionally corrected methods may not always outperform non-compositionally corrected methods [55–57]. In the current study, we utilised the MaAsLin2 method, which uses total sum scaling for normalisation. MaAsLin2 was recently compared to various compositionally corrected methods, naive methods, non-microbial analysis methods and experimental methods [57]. Compositionally corrected methods (e.g. analysis of the composition of microbiomes) or compositionally corrected normalisation (e.g. centred log-ratio transformation) did not improve performance over non-compositional approaches. While there is no method that is best in all scenarios, MaAsLin2 was the only multivariable method that controlled the false discovery rate and performed well in all scenarios [57]. A recent study also indicates that model goodness of fit may depend on whether the data come from 16S or shotgun metagenomic sequencing, due to different count data structures [56]. The differences in data structures may be due to several factors, including sequencing depth, taxonomic classification between technologies (i.e. metagenomic sequences vs. clusters of amplicon sequences) and

Table 5. Differential DNA pathways between African-American pancreatic cancer cases and controls.

Pathway code	Pathway common name	Expected taxonomic range[a]	Ontology-based classification[a]	Coefficient	Std. dev.	*P* value	*Q* value[b]
PWY0-1061	Superpathway of L-alanine biosynthesis	Bacteria	Biosynthesis, superpathways	−0.00063	0.00022	0.00492	0.06592
PWY66-400	Glycolysis VI (metazoan)	Metazoan	Generation of precursor metabolites and energy	−0.00045	0.00016	0.00547	0.06836
PWY-621	Sucrose degradation III (sucrose invertase)	Archaea, Bacteria, Eukaryota	Degradation/utilisation/ assimilation	−0.00055	0.00020	0.00627	0.07122
PWY-7357	Thiamine formation from pyrithiamine and oxythiamine (yeast)	Fungi	Biosynthesis	−0.00070	0.00026	0.00703	0.07423
LACTOSECAT-PWY	Lactose and galactose degradation I	Firmicutes	Degradation/utilisation/ assimilation	−0.00073	0.00029	0.01091	0.09848
PWY-7197	Pyrimidine deoxyribonucleotide phosphorylation	Archaea, Bacteria, Eukaryota	Biosynthesis, metabolic clusters	−0.00048	0.00019	0.01297	0.10778
ANAGLYCOLYSIS-PWY	Glycolysis III (from glucose)	Bacteria, Eukaryota	Generation of precursor metabolites and energy	−0.00034	0.00014	0.01354	0.11104
ASPASN-PWY	Superpathway of L-aspartate and L-asparagine biosynthesis	Bacteria	Biosynthesis, superpathways	0.00034	0.00014	0.01503	0.11760
PWY-7199	Pyrimidine deoxyribonucleosides salvage	Amoebozoa, Archaea, Bacteria, Metazoa, Viridiplantae	Biosynthesis	−0.00046	0.00019	0.01760	0.12650
MET-SAM-PWY	Superpathway of S-adenosyl-L-methionine biosynthesis	Bacteria	Superpathways	−0.00043	0.00018	0.02046	0.13856
FUCCAT-PWY	Fucose degradation	Bacteria	Degradation/utilisation/ assimilation	0.00002	0.00001	0.02437	0.15342
PWY0-1319	CDP-diacylglycerol biosynthesis II	Proteobacteria, Viridiplantae	Biosynthesis	−0.00047	0.00021	0.02463	0.15395
PWY-5097	L-Lysine biosynthesis VI	Archaeoglobaceae, Bacteroidetes, Chlamydiae, Chloroflexi, Cyanobacteria, Desulfuromonadales, Firmicutes, Magnoliopsida, Methanobacteriaceae, Methanococci, Spirochaeta	Biosynthesis	0.00036	0.00016	0.02879	0.16494
PWY-6125	Superpathway of guanosine nucleotides de novo biosynthesis II	Bacteria	Biosynthesis, superpathways	−0.00040	0.00018	0.02951	0.16699
COBALSYN-PWY	Superpathway of adenosylcobalamin salvage from cobinamide I	Proteobacteria	Biosynthesis, superpathways	0.00012	0.00006	0.02967	0.16741
HOMOSER-METSYN-PWY	L-methionine biosynthesis I	Bacteria	Biosynthesis	−0.00047	0.00022	0.03170	0.17724
METSYN-PWY	Superpathway of L-homoserine and L-methionine biosynthesis	Bacteria	Biosynthesis, superpathways	−0.00039	0.00018	0.03181	0.17735
PWY-7228	Superpathway of guanosine nucleotide de novo biosynthesis I	Archaea, Bacteria, Eukaryota	Biosynthesis, superpathways	−0.00036	0.00017	0.03581	0.19325
PWY-7208	Superpathway of pyrimidine nucleobase salvage	Archaea, Bacteria, Fungi, Viridiplantae	Biosynthesis, superpathways	−0.00047	0.00022	0.03673	0.19520
PWY-7220	Adenosine deoxyribonucleotides de novo biosynthesis II	Bacteria	Biosynthesis	−0.00042	0.00020	0.03682	0.19520
PWY-7222		Bacteria	Biosynthesis	−0.00042	0.00020	0.03682	0.19520

Table 5 continued

Pathway code	Pathway common name	Expected taxonomic range[a]	Ontology-based classification[a]	Coefficient	Std. dev.	*P* value	*Q* value[b]
	Guanosine deoxyribonucleotides de novo biosynthesis II						
FOLSYN-PWY	Superpathway of tetrahydrofolate biosynthesis and salvage	Bacteria, Fungi	Biosynthesis, superpathways	0.00028	0.00014	0.04193	0.21331
ARGSYNBSUB-PWY	L-Arginine biosynthesis II (acetyl cycle)	Archaea, Bacteria, Fungi, Viridiplantae	Biosynthesis	0.00035	0.00018	0.04361	0.21717
PWY-6612	Superpathway of tetrahydrofolate biosynthesis	Bacteria, Fungi, Viridiplantae	Biosynthesis, superpathways	0.00023	0.00011	0.04526	0.22040
SER-GLYSYN-PWY	Superpathway of L-serine and glycine biosynthesis I	Archaea, Bacteria, Eukaryota	Biosynthesis, superpathways	0.00025	0.00012	0.04598	0.22150
PWY-6703	preQ$_0$ biosynthesis	Bacteria	Biosynthesis	−0.00037	0.00019	0.04604	0.22150
PWY-724	Superpathway of L-lysine, L-threonine and L-methionine biosynthesis II	Viridiplantae	Biosynthesis, superpathways	0.00026	0.00013	0.04811	0.22586
PWY-6126	Superpathway of adenosine nucleotides de novo biosynthesis II	Archaea, Bacteria, Eukaryota	Biosynthesis, superpathways	−0.00038	0.00019	0.04987	0.23155

[a]Expected taxonomic range and ontology-based classification from MetaCycs.
[b]Benjamini–Hochberg procedure for false discovery rate control at the 0.05 level.

bioinformatic methods used for data preprocessing. The prior microbiome studies used 16S rRNA sequencing [10, 39–41], while the current study utilised shotgun metagenomic sequencing. Thus, data structure differences between 16S rRNA sequencing and shotgun metagenomic sequencing could partially account for different results between the prior studies [10, 39–43] and our current study.

Shotgun metagenomic sequencing permits functional profiling, substantially greater specificity in taxonomic profiling, and the detection (albeit at low levels) of non-bacterial organisms [58, 59]. A recent experimental study demonstrated that pancreatic tumours harbour ~3000-fold more fungi than normal pancreatic tissue and were specifically enriched for *Malassezia* [30]. Ablation of the fungi was protective, while repopulation with *Malassezia* accelerated pancreatic oncogenesis [30]. In addition, prior studies have reported that individuals with *Candida* infection are at higher risk of pancreatic cancer [60]. In the current study, only one fungal phylum (*Ascomycota*) and one genus within that phylum (*Saccharomyces*) were detected, and neither were differentially abundant between cases and controls. Without targeted enrichment protocols, fungi are generally difficult to detect in typical human-associated communities by any means, due to the relatively low abundance of oral fungi and lack of well-characterised reference genomes [61, 62]. By nature of shotgun metagenomics, bacteria tend to dominate in abundance and in the resulting sequence data, due to much higher biomass prevalence in the samples. Thus, fungi may have been filtered out with whole-genome sequencing due to sparsity. An elevated OR was observed for high levels of *Saccharomyces*, but this may simply be an indicator of an increased fungal population—not *Saccharomyces* specifically.

The mechanisms underlying the potential association between the oral microbiota and pancreatic cancer are still speculative. In a prior study, high *P. gingivalis* antibody levels were associated with an increased risk of orodigestive cancer (including pancreatic) even among individuals without the overt periodontal disease [8], suggesting that the role of the oral microbiota is not dependent on oral disease. In a tissue-based study, oral bacterial species were identified in the pancreatic duct [42, 63], suggesting that oral bacteria can migrate and have direct effects on pancreatic carcinogenesis. This has also been the case in other gastrointestinal cancers [64] and inflammatory bowel disease [65], although it is not clear whether such taxa are causal or responsive to tumorigenesis. Thus, several hypotheses have been proposed for the biological pathways linking oral health and microbiota to pancreatic risk, including (1) direct somatic pathways, whereby oral microbes migrate to the pancreas through ingestion or circulation following tooth brushing [66, 67] and (2) systemic inflammation due to periodontal disease increasing levels of proinflammatory cytokines, which may promote the development of pancreatic cancer [68, 69]. Another potential pathway involves oral bacteria producing carcinogenic metabolic byproducts from tobacco smoke and alcohol, both known pancreatic cancer risk factors [70–74].

In the current study, we characterised the functional potential of the oral microbiome in relation to pancreatic cancer risk. While we identified several pathways that were differential between cases and controls, significant values were large after adjusting for multiple comparisons and the effect sizes were small. Some of these pathways (e.g. fructose degradation) are known to play a role in cancer development at the cellular level [75]. However, there is not a clear mechanism for fructose degradation in the oral microbiota to influence pancreatic cancer development.

Multiple comparisons need to be considered when discussing the study results, as there were 161 taxa and 266 pathway comparisons within the main analyses. Thus, there is a possibility that some results arose due to chance. Adjusting for multiple comparisons reduces the likelihood of detecting a false-positive association, but it also reduces the power for detecting a true association if one exists. Thus, we chose to focus on associations based on biological plausibility and

consistency with published results [76]. Our a priori aim was to examine seven oral pathogens that have been associated with periodontal disease or tooth loss. To examine these, we interpreted the magnitude of the effect estimate and present CIs [77]. To examine featurewise analysis, MaAsLin2 was used to preserve statistical power while accounting for the nuances of microbiome features and controlling false discovery rates [57, 78]. We provide the output of the differential abundance analysis for cases and controls in Supplemental Tables S2-S3 and S7-S8.

The primary limitations of the current study were the one-time sample collection, oral wash method and sample size. Whether a single sample—vs. repeated sampling—accurately reflects the relevant oral microbiota for pancreatic cancer development is unknown. However, several studies have shown that the oral microbiome ascertained from oral wash samples is stable over time [79–82]. In addition, within-person variation in the oral microbiome is consistently lower than between-person variation [83]. This study included a relatively large number of African-American cases for a prospective study of rare cancer, yet still had only 122 cases of incident pancreatic cancer in this group. In particular, statistical power for stratified analyses (e.g. by smoking status or sex) was limited. The BWHS study population is more highly educated than the general US population and study participants are from across the United States, while the SCCS study population is less highly educated and participants reside in the southeastern US. While these two studies are not population-based, they encompass a wide range of demographics and geographic locals within the US. Thus, these results are likely generalisable to other African Americans, but this cannot be assumed as neither study population represents all African-American individuals in the US.

Strengths include the study design, sequencing method and control of confounding. As pancreatic cancer can be rapidly fatal and case–control studies are susceptible to reverse causation, we conducted a study nested within two prospective cohorts and excluded cases that occurred within 2 years of provision of the oral wash sample. We utilised shotgun metagenomic sequencing for the detection of microbial taxa and functional profiling [58, 59]. Finally, detailed data collected in both cohort studies allowed for consideration of a wide range of potential confounders.

In conclusion, this prospective study of African Americans provides evidence of associations between known oral bacterial pathogens and pancreatic cancer risk among non-smokers. However, the findings did not reproduce previously reported overall associations in a White population.

DATA AVAILABILITY

The oral microbiome data produced in the current study are available via the database of Genotypes and Phenotypes (dpGaP, accession number: phs002454.v1.p1).

REFERENCES

1. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: Incidence—SEER 18 Regs Research Data, Nov 2018 Sub (2000–2015) <Katrina/Rita Population Adjustment>—Linked To County Attributes—Total U.S., 1969–2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2019, based on the November 2018 submission. www.seer.cancer.gov.
2. Gerlovin H, Michaud DS, Cozier YC, Palmer JR. Oral health in relation to pancreatic cancer risk in African American women. Cancer Epidemiol Biomark Prev. 2019;28:675–9.
3. Michaud DS, Joshipura K, Giovannucci E, Fuchs CS. A prospective study of periodontal disease and pancreatic cancer in US male health professionals. J Natl Cancer Inst. 2007;99:171–5.
4. Hiraki A, Matsuo K, Suzuki T, Kawase T, Tajima K. Teeth loss and risk of cancer at 14 common sites in Japanese. Cancer Epidemiol Biomark Prev. 2008;17:1222–7.
5. Stolzenberg-Solomon RZ, Dodd KW, Blaser MJ, Virtamo J, Taylor PR, Albanes D. Tooth loss, pancreatic cancer, and *Helicobacter pylori*. Am J Clin Nutr. 2003;78:176–81.
6. Hujoel PP, Drangsholt M, Spiekerman C, Weiss NS. An exploration of the periodontitis-cancer association. Ann Epidemiol. 2003;13:312–6.
7. Arora M, Weuve J, Fall K, Pedersen NL, Mucci LA. An exploration of shared genetic risk factors between periodontal disease and cancers: a prospective co-twin study. Am J Epidemiol. 2010;171:253–9.
8. Ahn J, Segers S, Hayes RB. Periodontal disease, Porphyromonas gingivalis serum antibody levels and orodigestive cancer mortality. Carcinogenesis. 2012;33:1055–8.
9. Maisonneuve P, Amar S, Lowenfels AB. Periodontal disease, edentulism, and pancreatic cancer: a meta-analysis. Ann Oncol. 2017;28:985–95.
10. Fan X, Alekseyenko AV, Wu J, Peters BA, Jacobs EJ, Gapstur SM, et al. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. Gut. 2018;67:120–7.
11. Como DH, Stein Duker LI, Polido JC, Cermak SA. The persistence of oral health disparities for African American children: a scoping review. Int J Environ Res Public Health. 2019;16:710. https://doi.org/10.3390/ijerph16050710.
12. Rosenberg L, Adams-Campbell L, Palmer JR. The Black Women's Health Study: a follow-up study for causes and preventions of illness. J Am Med Women's Assoc. 1995;50:56–8.
13. Signorello LB, Hargreaves MK, Blot WJ. The Southern Community Cohort Study: investigating health disparities. J Health Care Poor Underserved. 2010;21:26–37.
14. Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, et al. Southern community cohort study: establishing a cohort to investigate health disparities. J Natl Med Assoc. 2005;97:972–9.
15. Cozier YC, Palmer JR, Rosenberg L. Comparison of methods for collection of DNA samples by mail in the Black Women's Health Study. Ann Epidemiol. 2004;14:117–22.
16. Garcia-Closas M, Egan KM, Abruzzo J, Newcomb PA, Titus-Ernstoff L, Franklin T, et al. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer Epidemiol Biomark Prev. 2001;10:687–96.
17. Yang Y, Cai Q, Shu XO, Steinwandel MD, Blot WJ, Zheng W, et al. Prospective study of oral microbiome and colorectal cancer risk in low-income and African American populations. Int J Cancer. 2019;144:2381–9.
18. Yang Y, Zheng W, Cai Q, Shrubsole MJ, Pei Z, Brucker R, et al. Racial differences in the oral microbiome: data from low-income populations of African ancestry and European ancestry. mSystems. 2019;4:e00639–19. https://doi.org/10.1128/mSystems.00639-19.
19. Adams-Campbell LL, Dash C, Palmer JR, Wiedemeier MV, Russell CW, Rosenberg L, et al. Predictors of biospecimen donation in the Black Women's Health Study. Cancer Causes Control. 2016;27:797–803.
20. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159:789–99.
21. McIver LJ, Abu-Ali G, Franzosa EA, Schwager R, Morgan XC, Waldron L, et al. bioBakery: a meta'omic analysis environment. Bioinformatics. 2018;34:1235–7.
22. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012;9:811–4.
23. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res. 2020;48:D445–53.
24. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol. 2017;35:1077–86.
25. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.
26. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: Community Ecology Package. R package version 2.5-6. 2019, https://cran.r-project.org/web/packages/vegan/index.html. Accessed October 15, 2020.
27. Hadley W. Ggplot2. New York, NY: Springer Science+Business Media, LLC; 2016.
28. Holt SC, Ebersole JL. *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*: the "red complex", a prototype polybacterial pathogenic consortium in periodontitis. Periodontology 2000. 2005;38:72–122.
29. Aruni AW, Mishra A, Dou Y, Chioma O, Hamilton BN, Fletcher HM. Filifactor alocis—a new emerging periodontal pathogen. Microbes Infect. 2015;17:517–30.
30. Aykut B, Pushalkar S, Chen R, Li Q, Abengozar R, Kim JI, et al. The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. Nature. 2019;574:264–7.
31. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
32. Buchowski MS, Schlundt DG, Hargreaves MK, Hankin JH, Signorello LB, Blot WJ. Development of a culturally sensitive food frequency questionnaire for use in the Southern Community Cohort Study. Cell Mol Biol. 2003;49:1295–304.
33. Kumanyika SK, Mauger D, Mitchell DC, Phillips B, Smiciklas-Wright H, Palmer JR. Relative validity of food frequency questionnaire nutrient estimates in the Black Women's Health Study. Ann Epidemiol. 2003;13:111–8.

34. 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.
35. Yu G, Phillips S, Gail MH, Goedert JJ, Humphrys MS, Ravel J, et al. The effect of cigarette smoking on the oral and nasal microbiota. Microbiome. 2017;5:3.
36. Huang C, Shi G. Smoking and microbiome in oral, airway, gut and some systemic diseases. J Transl Med. 2019;17:225.
37. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. Am Statistician. 2019;73:235–45.
38. Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57:289–300.
39. Vogtmann E, Han Y, Caporaso JG, Bokulich N, Mohamadkhani A, Moayyedkazemi A, et al. Oral microbial community composition is associated with pancreatic cancer: a case-control study in Iran. Cancer Med. 2020;9:797–806.
40. Farrell JJ, Zhang L, Zhou H, Chia D, Elashoff D, Akin D, et al. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. Gut. 2012;61:582–8.
41. Lu H, Ren Z, Li A, Li J, Xu S, Zhang H, et al. Tongue coating microbiome data distinguish patients with pancreatic head cancer from healthy controls. J Oral Microbiol. 2019;11:1563409.
42. Del Castillo E, Meier R, Chung M, Koestler DC, Chen T, Paster BJ, et al. The microbiomes of pancreatic and duodenum tissue overlap and are highly subject specific but differ between pancreatic cancer and noncancer subjects. Cancer Epidemiol Biomark Prev. 2019;28:370–83.
43. Half E, Keren N, Reshef L, Dorfman T, Lachter I, Kluger Y, et al. Fecal microbiome signatures of pancreatic cancer patients. Sci Rep. 2019;9:16801.
44. Michaud DS, Izard J, Wilhelm-Benartzi CS, You DH, Grote VA, Tjonneland A, et al. Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large European prospective cohort study. Gut. 2013;62:1764–70.
45. Hanioka T, Morita M, Yamamoto T, Inagaki K, Wang PL, Ito H, et al. Smoking and periodontal microorganisms. Jpn Dent Sci Rev. 2019;55:88–94.
46. Scott DA, Singer DL. Suppression of overt gingival inflammation in tobacco smokers - clinical and mechanistic considerations. Int J Dent Hyg. 2004;2:104–10.
47. Palmer RM, Wilson RF, Hasan AS, Scott DA. Mechanisms of action of environmental factors–tobacco smoking. J Clin Periodontol. 2005;32:180–95.
48. Bagaitkar J, Williams LR, Renaud DE, Bemakanakere MR, Martin M, Scott DA, et al. Tobacco-induced alterations to *Porphyromonas gingivalis*-host interactions. Environ Microbiol. 2009;11:1242–53.
49. Bagaitkar J, Daep CA, Patel CK, Renaud DE, Demuth DR, Scott DA. Tobacco smoke augments *Porphyromonas gingivalis-Streptococcus gordonii* biofilm formation. PLoS ONE. 2011;6:e27386.
50. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.
51. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013;10:1200–2.
52. Quinn TP, Erb I, Richardson MF, Crowley TM. Understanding sequencing data as compositions: an outlook and review. Bioinformatics. 2018;34:2870–8.
53. Quinn TP, Crowley TM, Richardson MF. Benchmarking differential expression analysis tools for RNA-Seq: normalization-based vs. log-ratio transformation-based methods. BMC Bioinform. 2018;19:274.
54. Calle ML. Statistical analysis of metagenomics data. Genomics Inf. 2019;17:e6.
55. Hawinkel S, Mattiello F, Bijnens L, Thas O. A broken promise: microbiome differential abundance methods do not control the false discovery rate. Brief Bioinform. 2019;20:210–21.
56. Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data. Genome Biol. 2020;21:191.
57. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. bioRxiv:2021.2001.2020.427420 [Preprint] 2021. Available from: https://doi.org/10.1101/2021.01.20.427420.
58. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. Front Microbiol. 2016;7:459.
59. Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, et al. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. Nat Rev Microbiol. 2015;13:360–72.
60. Li S, Fuhler GM, Bn N, Jose T, Bruno MJ, Peppelenbosch MP, et al. Pancreatic cyst fluid harbors a unique microbiome. Microbiome. 2017;5:147.
61. Coker OO, Nakatsu G, Dai RZ, Wu WKK, Wong SH, Ng SC, et al. Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. Gut. 2019;68:654–62.
62. Klimesova K, Jiraskova Zakostelska Z, Tlaskalova-Hogenova H. Oral bacterial and fungal microbiome impacts colorectal carcinogenesis. Front Microbiol. 2018;9:774.
63. Swidsinski A, Schlien P, Pernthaler A, Gottschalk U, Barlehner E, Decker G, et al. Bacterial biofilm within diseased pancreatic and biliary tracts. Gut. 2005;54:388–95.
64. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. PLoS Comput Biol. 2016;12:e1004977.
65. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569:655–62.
66. Crasta K, Daly CG, Mitchell D, Curtis B, Stewart D, Heitz-Mayfield LJ. Bacteraemia due to dental flossing. J Clin Periodontol. 2009;36:323–32.
67. Lockhart PB, Brennan MT, Sasser HC, Fox PC, Paster BJ, Bahrani-Mougeot FK. Bacteremia associated with toothbrushing and dental extraction. Circulation. 2008;117:3118–25.
68. Garlet GP. Destructive and protective roles of cytokines in periodontitis: a re-appraisal from host defense and tissue destruction viewpoints. J Dent Res. 2010;89:1349–63.
69. Cardoso EM, Reis C, Manzanares-Cespedes MC. Chronic periodontitis, inflammatory cytokines, and interrelationship with other chronic diseases. Postgrad Med. 2018;130:98–104.
70. Salaspuro M. Acetaldehyde as a common denominator and cumulative carcinogen in digestive tract cancers. Scand J Gastroenterol. 2009;44:912–25.
71. Kurkivuori J, Salaspuro V, Kaihovaara P, Kari K, Rautemaa R, Gronroos L, et al. Acetaldehyde production from ethanol by oral streptococci. Oral Oncol. 2007;43:181–6.
72. Shapiro KB, Hotchkiss JH, Roe DA. Quantitative relationship between oral nitrate-reducing activity and the endogenous formation of N-nitrosoamino acids in humans. Food Chem Toxicol. 1991;29:751–5.
73. Risch HA. Pancreatic cancer: *Helicobacter pylori* colonization, N-nitrosamine exposures, and ABO blood group. Mol Carcinogen. 2012;51:109–18.
74. Duell EJ. Epidemiology and potential mechanisms of tobacco smoking and heavy alcohol consumption in pancreatic cancer. Mol Carcinogen. 2012;51:40–52.
75. Keeley TS, Yang S, Lau E. The diverse contributions of fucose linkages in cancer. Cancers. 2019;11:1241, https://doi.org/10.3390/cancers11091241.
76. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990;1:43–6.
77. Silva-Aycaguer LC, Suarez-Gil P, Fernandez-Somoano A. The null hypothesis significance test in health sciences research (1995–2006): statistical analysis and interpretation. BMC Med Res Methodol. 2010;10:44.
78. Mallick H, Rahnavard A, McIver L. Maaslin2: Maaslin2. R package version 1.3.0. 2020. http://huttenhower.sph.harvard.edu/maaslin. Accessed October 15, 2020.
79. Vogtmann E, Hua X, Zhou L, Wan Y, Suman S, Zhu B, et al. Temporal variability of oral microbiota over 10 months and the implications for future epidemiologic studies. Cancer Epidemiol Biomark Prev. 2018;27:594–600.
80. 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.
81. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. Genome Biol. 2014;15:R89.
82. Stahringer SS, Clemente JC, Corley RP, Hewitt J, Knights D, Walters WA, et al. Nurture trumps nature in a longitudinal survey of salivary bacterial communities in twins from early adolescence to early adulthood. Genome Res. 2012;22:2146–52.
83. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207–14.

ACKNOWLEDGEMENTS

Pathology data were obtained from the following state cancer registries (AZ, CA, CO, CT, DE, DC, FL, GA, IL, IN, KY, LA, MD, MA, MI, NJ, NY, NC, OK, PA, SC, TN, TX, VA), and results reported do not necessarily represent their views. We thank participants and staff of the BWHS and SCCS for their contributions.

AUTHOR CONTRIBUTIONS

Statistical analysis, interpretation of data, drafting of the paper and critical revision of the paper for important intellectual content: JLP. Statistical analysis, interpretation of data and critical revision of the paper for important intellectual content: JEW and HG. Study concept and design, analysis and interpretation of data, and critical revision of the paper for important intellectual content: DSM and CH. Interpretation of data and critical revision of the paper for important intellectual content: QC, BMW, EAR-N, JL, YY, and WEJ. Study concept and design, acquisition

of data, analysis and interpretation of data and critical revision of the paper for important intellectual content: LBS and X-OS. Study concept and design, acquisition of data, analysis and interpretation of data, drafting of the paper and critical revision of the paper for important intellectual content: JRP. All authors approved the final draft submitted.

FUNDING

This work was supported by National Institutes of Health grants U01 CA164974, U01 CA187508 and R01 CA058420; the Karin Grunebaum Cancer Research Foundation and Boston University Peter Paul Career Development Professorship.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All study participants provided written informed consent. The Institutional Review Boards of Boston University (Boston, MA) and Vanderbilt University Medical Center and Meharry Medical College (Nashville, TN) approved the BWHS and SCCS, respectively, and reviewed the studies annually.

COMPETING INTERESTS

BMW has received research funding from Celgene and Eli Lily and also serves as a consultant for BioLineRx, Celgene and Grail. The other authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-021-01578-5.

Correspondence and requests for materials should be addressed to Jessica L. Petrick or Julie R. Palmer.