HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2020 May 15.

Published in final edited form as:

Int J Cancer. 2019 May 15; 144(10): 2381–2389. doi:10.1002/ijc.31941.

Prospective Study of Oral Microbiome and Colorectal Cancer Risk in Low-income and African American Populations

Yaohua Yang[1], **Qiuyin Cai**[1], **Xiao-Ou Shu**[1], **Mark D. Steinwandel**[1], **William J. Blot**[1], **Wei Zheng**[1], and **Jirong Long**[1]

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN

Abstract

Oral microbiome may play an important role in cancer pathogenesis. However, no study has prospectively investigated the association of the oral microbiome with subsequent risk of developing colorectal cancer (CRC).

We conducted a nested case-control study including 231 incident CRC cases and 462 controls within the Southern Community Cohort Study with 75% of the subjects being African-Americans. The controls were individually matched to cases based on age, race, smoking, season-of-study enrollment and recruitment method. Oral microbiota were assessed using 16S rRNA gene sequencing in pre-diagnostic mouth rinse samples.

Multiple bacterial taxa showed an association with CRC risk at $P<0.05$. Oral pathogens *Treponema denticola* and *Prevotella intermedia* were associated with an increased risk of CRC, with odds ratios (ORs) and 95% confidence intervals (CIs) of 1.76(1.19–2.60) and 1.55(1.08–2.22), respectively, for the individuals carrying these bacteria compared to non-carriers. In the phylum *Actinobacteria*, *Bifidobacteriaceae* was more abundant among CRC patients than among controls. In the phylum *Bacteroidetes*, *Prevotella denticola* and *Prevotella sp. oral taxon 300* were associated with an increased CRC risk, while *Prevotella melaninogenica* was associated with a decreased risk of CRC. In the phylum *Firmicutes*, *Carnobacteriaceae*, *Streptococcaceae*, *Erysipelotrichaceae*, *Streptococcus*, *Solobacterium*, *Streptococcus sp. oral taxon 058* and *Solobacterium moorei* showed associations with a decreased risk of CRC. Most of these associations were observed among both African- and European-Americans. Most of the associations were not significant after Bonferroni correction for multiple testing, which may be conservative.

Our study suggests that the oral microbiome may play a significant role in CRC etiology.

Keywords

Colorectal cancer; oral microbiome; 16S rRNA

Correspondence to: Jirong Long, PhD, Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, TN 37203-1738, Tel: 615-343-6741, Fax: 615-343-0719, jirong.long@vumc.org.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors have declared no conflicts of interest.

INTRODUCTION

Colorectal cancer (CRC) is the third leading cause of cancer death in the US [1]. Epidemiological studies have shown that lifestyle factors, such as obesity, smoking and alcohol drinking are associated with an increased CRC risk [2, 3]. Systemic inflammation is also associated with an increased risk of CRC [3].

The oral microbiome plays a critical role in the occurrence of oral diseases [4, 5], and may also play an important role in maintaining systemic health [6, 7]. Some studies have suggested an increased risk of CRC [8] in individuals with oral diseases, though null associations were reported in other studies [9, 10]. Chronic inflammation caused by oral microbes has been suggested to play a role in CRC development [11]. Two research teams reported that the *Fusobacterium nucleatum*, one of the predominant subgingival microbial species present in periodontitis [12], is more abundant in colorectal carcinoma tissues than in normal colorectal tissues [13, 14]. In addition, a recent study [15] proposed that colon conditions could provide an anaerobic environment similar to subgingival space. In this anaerobic environment, periodontal pathogen species, such as *Prevotella intermedia*, *Fusobacterium nucleatum* and *Prevotella gingivalis*, are adept at coaggregation, which may result in consistent and chronic inflammation, and then promote the development of CRC [15]. These results suggest that the oral microbiome may play a direct role in the pathophysiology of CRC. However, to our knowledge, no study has prospectively and systematically investigated the oral microbiome in relation to the risk of CRC. Herein, we carried out a prospective nested case-control study within the Southern Community Cohort Study (SCCS) to investigate the association of the oral microbiome with the subsequent risk of developing CRC.

MATERIALS AND METHODS

Study participants and data collection

The SCCS is an ongoing prospective cohort study investigating risk factors for cancer and other chronic diseases. Details on the methodology of the study have been described elsewhere [16, 17]. Briefly, approximately 86,000 adults were recruited between 2002 and 2009 from 12 states in the southeastern US, two-thirds of whom were African-Americans. Approximately 86% of them were recruited from community health centers (CHCs), institutions providing basic health care and preventative services in underserved areas. As a result, the cohort includes a substantial number of individuals of low income and educational status. The remaining 14% of the cohort members were recruited through mail-based general population sampling. Mouth rinse samples were collected from ~34,100 participants at the time of enrollment. The SCCS was reviewed and approved by the institutional review boards at Vanderbilt University Medical Center and Meharry Medical College. Written informed consent was obtained from all study participants.

During the study enrollment, participants completed a baseline survey with a comprehensive questionnaire, which collected information on demographics, anthropometric characteristics, lifestyle factors, disease history, medication use, and other characteristics. As part of the active follow-up surveys initiated in 2008, participants were asked about their personal medical histories, including their oral health. Incident CRC cases diagnosed after entry into

the SCCS were identified via linkage to state cancer registries operating in the 12-state study area and national mortality registries. CRC was defined according to the International Classification of Diseases (ICD-10), codes C18-C21.

We conducted a nested case-control study of incident CRC among SCCS participants who donated mouth rinse samples. Individuals who received antibiotics treatment during the seven days prior to sample collection were excluded. For each of the 231 incident CRC cases, two controls were randomly selected and individually matched to cases by age of enrollment (± 5 years), race (African-American/European-American/other), gender, smoking status (current smoker/former smoker/ never smoker), season-of-study enrollment (spring/ summer/fall/winter) and recruitment method (CHC/general population).

DNA extraction and 16S rRNA gene sequencing

DNA was isolated from mouth rinse samples using Qiagen's QIAmp DNA kit. Sequencing libraries were prepared using the NEXTflex 16S V4 Amplicon-Seq Kit (Bioo Scientific 4201–05), following the protocol provided by the manufacturer. Sequencing was performed at paired-end 250bp using Illumina HiSeq System at the BGI Americas. Each 96-well plate was sequenced with two duplicate quality control samples (QCs). In total, six duplicate samples from the same single subject were sequenced and very similar microbiome profiles were observed: coefficient of variability (CV) for the Shannon index and the Simpson index (measurement of microbial community diversity within each sample) among the six samples were 1.7% and 0.3%, respectively; CV for the relative abundance of the four most abundant phyla, i.e. *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, among the six samples ranged from 5.1% to 9.4%; Pearson correlation coefficients of phylum-level microbial relative abundance between the six samples ranged from 97.8% to 99.9%.

Sequence data processing and QC

Raw sequencing data were trimmed and filtered to remove bases and reads of low quality by the use of Sickle [18]. Then, BayesHammer [19] was utilized for sequencing error correction and PANDAseq [20] was used to stitch the paired-end reads together [21]. Clean reads were then clustered into Operational Taxonomic Units (OTUs) at 97% sequence identity using the closed-reference OTU picking strategy with the Human Oral Microbiome Database [22] (HOMD) as reference via the Quantitative Insights Into Microbial Ecology (QIIME), v1.9.1 [23]. We removed one sample with less than 20,000 sequencing reads. Finally, data from 231 CRC patients and 461 controls were included in the downstream analysis.

Statistical analysis

To evaluate the relationship between overall oral microbiota composition and CRC risk, we calculated the Bray-Curtis dissimilarity [24], as well as unweighted and weighted UniFrac distance matrices [25]. PERMANOVA-S [26] was used to estimate whether there was a difference regarding these three distance metrics between CRC cases and controls. Statistical analyses were conducted within matched case-control sets.

To estimate the association between individual bacterial taxon and CRC risk, first we investigated five pre-defined oral pathogens, including three "red complex" periodontal

pathogens [27], *Porphyromonas gingivalis*, *Treponema denticola* and *Tannerella forsythia*, and two additional periodontal pathogenic species, *Fusobacterium nucleatum* [12] and *Prevotella intermedia* [28]. For each pathogen, study participants were categorized into two groups according to whether they carried the pathogen or not. The association between prevalence of the pathogen and CRC risk was analyzed through conditional logistic regression modeling.

We also investigated other bacterial taxa in relationship to CRC risk. Taxa of different taxonomic ranks are correlated evolutionarily through a phylogenetic tree. Taxa with a median relative abundance of >0.01% among control subjects, including seven phyla, 24 families, 35 genera and 71 species, were treated as "common taxa" in this study and included in downstream analyses. At each taxonomic level, the raw sequencing count for each taxon was normalized using the centered log-ratio transformation after adding 1 as a pseudocount [29, 30]. Then for each taxon, a conditional logistic regression analysis was performed to test the association of taxa abundance with CRC risk. For taxa with a median relative abundance of ≤0.01% among controls, they were treated as "rare taxa" in this study. Only those rare taxa that were observed in at least 20% of the control subjects were included in the statistical analyses, in which study participants were grouped into carriers and non-carriers. Conditional logistic regression analyses were conducted for each taxon to evaluate the association of taxon prevalence with CRC risk. This included three phyla, 22 families, 54 genera and 145 species.

During all of these analyses, smoking amount (pack-year of smoking) and alcohol consumption status were adjusted. For analyses of prevalence, sequencing depth was additionally adjusted. Considering the fact that taxa of different taxonomic ranks are correlated, we used Galwey's method [31], implemented in the function "meff" of the R package "poolR" (https://github.com/ozancinar/poolR/) to estimate the total independent tests. Then, Bonferroni correction was used to adjust for multiple testing. Stratified analyses by race and gender were conducted for both carriage and abundance. In addition, we did sensitivity analyses by excluding the cases diagnosed within two years (N=47) after enrollment and the corresponding controls (N=94). We also conducted analyses based on relative abundance after rarefying the OTU table at a subsampling depth of 20,000 to account for variations in sequencing depth among samples. The results didn't change materially. All of these analyses were carried out using R (v3.3.1) and Python (v2.7.8).

RESULTS

Table 1 presents the distribution of selected demographic characteristics of the study participants. About 74% of the study participants were African-Americans, 60% female, 30% overweight and over 40% obese. Most participants had a low education level and low income, with only about 18% having a college education and 15% having an annual household income of at least $50,000. The CRC patients and control subjects were generally similar for these characteristics, and for smoking status, because of the matched study design. No significant differences were observed for tooth loss or tooth decay, but CRC cases were slightly more likely to have poor oral health status.

Associations of overall microbiome composition with CRC risk

We did not find any significant difference of overall microbiome composition between CRC cases and matched controls, as measured by Bray-Curtis dissimilarity, weighted UniFrac and unweighted UniFrac distance matrices and tested by PERMANOVA-S [26].

Associations of oral pathogens with CRC risk

All five of the investigated oral pathogens were more prevalent among CRC patients than among controls, but only differences for *Treponema denticola* and *Prevotella intermedia* reached a $P<0.05$ (Table 2). Specifically, the carriage of *Treponema denticola* was associated with an increased risk of CRC with an OR (95% CI) of 1.76 (1.19–2.60) and a *P* value of 4.45×10^{-3}. For these five pathogens, there were three independent tests, and after Bonferroni correction, *Treponema denticola* still showed an association with CRC risk. Carriages of *Treponema denticola* and *Tannerella forsythia* were correlated with a spearman correlation coefficient of 0.57. After adjusting for *Tannerella forsythia, Treponema denticola* was still associated with an increased risk of CRC with an OR (95% CI) of 1.95(1.25–3.04) and a *P* value of 3.14×10^{-3}. Similarly, *Prevotella intermedia* was also associated with an increased risk of CRC with an OR (95% CI) being 1.55(1.08–2.22) and a *P* value of 0.02 (Table 2). *Fusobacterium nucleatum* was present in most study participants, with a nearly equal prevalence observed in CRC patients and control subjects.

Associations of other bacterial taxa with CRC risk

We evaluated the differences of abundance for common taxa ($N=137$) between CRC patients and control participants. Multiple taxa were associated with CRC risk at a $P<0.05$ (Table 3). In the phylum *Actinobacteria,* the family *Bifidobacteriaceae* was associated with an increased risk of CRC with an OR (95% CI) of 1.10(1.01–1.19) and a *P* value of 0.03. In the phylum *Bacteroidetes*, the species *Prevotella melaninogenica* was associated with a decreased CRC risk with an OR (95% CI) of 0.91(0.84–0.99) and a *P* value of 0.04. Within this phylum, another two species, *Prevotella denticola* and *Prevotella sp. oral taxon 300*, were associated with an increased risk, with ORs (95% CIs) of 1.11(1.02–1.20) and 1.13(1.01–1.26) respectively, and *P* values of 0.02 and 0.04, respectively. In the phylum *Firmicutes*, seven taxa were found to be associated with CRC risk and all of them showed a protective effect on the risk of developing CRC. Among them, the most abundant taxon at the species level, *Streptococcus sp. oral taxon 058*, showed the most significant association with an OR (95% CI) of 0.79(0.67–0.94) and a *P* value of 7.87×10^{-3} (Table 3). Among the 137 common taxa included in association analyses of taxa abundance and CRC risk, there were 53 independent tests and no taxa maintained a significant association with CRC risk after Bonferroni correction (Table 3).

Among those 224 rare taxa, 16 showed an association with CRC risk at $P<0.05$, and all of them were associated with an increased risk of CRC (Table 4). The most significant taxon was the phylum *SR1*. It was observed among ~55% of cases and ~42% of controls. Participants carrying this species had a higher risk than non-carriers with an OR (95% CI) of 1.76(1.25–2.47) and a *P* value of 1.09×10^{-3}. Within this phylum, two species were associated with CRC risk. Carrying *SR1_[G-1] sp. oral taxon 874* was associated with a 75% increased risk of CRC with a *P* value of 1.58×10^{-3}. Carrying *SR1_[G-1] sp. oral taxon*

345 was associated with a 50% increased CRC risk with a *P* value of 0.04. In the phylum *Actinobacteria*, the species *Bifidobacterium dentium* was associated with an increased risk of CRC. In the phylum *Bacteroidetes*, in addition to the oral pathogen *Prevotella intermedia* (Table 2), another species, *Prevotella sp. oral taxon 304*, also showed an association with increased CRC risk. Three genera and two species in the phylum *Firmicutes*, namely *Peptococcus*, *Anaeroglobus*, *Mitsuokella*, *Lactobacillus salivarius and Eubacterium yurii*, were associated with an increased CRC risk. In the phylum *Proteobacteria*, four taxa were associated with an increased risk of CRC, including *Burkholderiaceae*, *Lautropia*, *Neisseria oralis* and *Campylobacter sp. oral taxon 044*. Among these four taxa, strong correlations were observed between *Burkholderiaceae* and *Lautropia.* After mutual adjustments, neither of these two taxa showed an association with CRC risk. In the phylum *Spirochaetes*, in addition to the oral pathogen *Treponema denticola*, two more species, *Treponema lecithinolyticum* and *Treponema sp. oral taxon 250*, were also associated with an increased CRC risk. Among the 224 rare taxa included in the investigation of taxa prevalence and CRC risk, there were 92 independent tests. After Bonferroni correction, none of these associations reached $P<0.05$.

We further evaluated the associations between the oral taxa and CRC risk, presented in Tables 2–4, stratified by race and gender. Most associations were observed among both African-Americans and European-Americans, as well as in men and women. In general, the associations were stronger among African-Americans than among European-Americans, and among females than among males (Supplementary Table 1–5). For example, the oral pathogen *Treponema denticola* showed an association with an increased risk of CRC with ORs (95% CIs) of 1.79(1.12–2.86) and 1.55(0.70–3.41) among African-Americans and European-Americans, respectively, and 2.02(1.19–3.41) and 1.33(0.73–2.42) among females and males, respectively (Supplementary Table 1). Similar differential associations were also observed for *Prevotella intermedia* between African-Americans and European-Americans, and between females and males (Supplementary Table 1). We also found that some taxa showed stronger associations among European-Americans than among African-Americans, e.g. *Burkholderiaceae*, *Lautropia* and *Lactobacillus salivarius* (Supplementary Table 4). However, a formal test of multiplicative interaction failed to show statistical significance. We also did sensitivity analyses to exclude the CRC patients who were diagnosed within two years ($N=47$) after enrollment and the corresponding controls ($N=94$), and the results did not change materially.

DISSCUSSION

In this prospective study of oral microbiome and CRC risk, we found that two of five oral pathogens, *Treponema denticola* and *Prevotella intermedia,* were associated with increased CRC risk. In addition, 11 common taxa and 16 rare taxa were also associated with the risk of CRC. Our findings warrant further investigation in larger studies to comprehensively estimate the potentiality of utilizing the oral microbiota for CRC early detection or manipulating it for CRC prevention.

Two studies have reported the increased abundance of the oral pathogen *Fusobacterium nucleatum* in colorectal carcinoma tissues than in normal colorectal tissues [13, 14]. In the

present study, *Fusobacterium nucleatum* was nearly universally prevalent in the oral cavity, and neither abundance nor prevalence of this bacteria was associated with CRC risk. Similarly, in two recently published studies with cross-sectional designs, no associations were observed between oral *Fusobacterium nucleatum* and CRC risk [32, 33]. In the present study, sequencing data of the V4 region of the 16S rRNA gene were used to quantify the abundance of *Fusobacterium nucleatum*. This may cause misclassification of microbial composition [34], hence led to the lowered statistical power to investigate the relationship of this bacteria with CRC risk.

All of the other four periodontal pathogens, *Porphyromonas gingivalis, Tannerella forsythia, Treponema denticola* and *Prevotella intermedia*, were associated with an increased risk of CRC, though only the associations of the latter two taxa reached nominal significance. Although there is no existing direct evidence linking them to CRC risk, several studies suggested that they were associated with other cancers. For example, *Treponema denticola* was found to be more prevalent in cancerous esophageal tissues than in normal tissues [35]. This pathogen was also reported to promote stemness and migration in oral squamous carcinoma [36]. In a recent prospective case-control study, *Treponema forsythia* showed a suggestive significant association with a higher risk of esophageal adenocarcinoma [37]. *Prevotella intermedia* was suggested to cooperate with other oral pathogens, such as *Fusobacterium nucleatum* and *Porphyromonas gingivalis*, to colonize and persist as a community in the colon and form an inflammatory microenvironment, which may promote CRC development [15]. However, in the present study, we did not see any correlation of the *Prevotella intermedia* with *Fusobacterium nucleatum* or *Porphyromonas gingivalis*, with the maximum Pearson correlation coefficient of <0.47. Studies have also suggested that certain oral bacteria may transfer to other body sites, including the gastrointestinal tract [38, 39], while further research will be needed to elucidate the underlying biological mechanism of this transmission.

Multiple studies investigating oral health status in association with CRC risk have been conducted and inconsistent results were reported [8–10]. In 2016, we carried out a meta-analysis of oral health with risk of CRC and did not find any associations [9]. However, the results need to be interpreted with caution because there were many limitations in these studies, e.g., an inconsistent definition of oral health, self-reported oral health (potentially prone to recall errors) and many others. On the other hand, the association of the oral pathogen *Fusobacterium nucleatum* with CRC has been consistently reported; however, the mechanism underlying the association is not clear. In the present study, we found associations for the other oral pathogenic bacteria with CRC, however, the results need to be replicated in larger studies.

In the phylum *Actinobacteria*, the genus *Bifidobacteriaceae* was more abundant and the species *Bifidobacterium dentium* was more prevalent among CRC cases than among controls. No study had linked these two taxa to the risk of developing CRC. In a recent oral microbiome study, the genus *Rothia* of *Actinobacteria* was found to be associated with CRC risk [32]. However, in the present study, this taxon was only slightly more abundant among CRC patients (7.8%) than among controls (7.3%).

The phylum *Firmicutes* showed a lower abundance in CRC cases than in controls. Specifically, the genus *Streptococcus* was also found to be significantly less abundant among CRC patients than among controls in a recent oral microbiome study [33]. Similarly, in a gut microbiome study, the genus *Streptococcus* and the species *Streptococcaceae* and *Streptococcus sp. oral taxon 058,* showed a lower relative abundance in CRC patients than in normal controls [40].In the present study, the family *Erysipelotrichaceae*, along with one of its genera, *Solobacterium*, and one of its species, *Solobacterium moorei*, were also associated with a decreased CRC risk. However, in two gut microbiome studies [41, 42], *Erysipelotrichaceae* and *Solobacterium moorei* were significantly enriched in CRC patients compared with healthy controls. This inconsistency may be derived from the different roles of oral and gut microbiota in CRC etiology, considering that for *Fusobacterium nucleatum*, previous gut microbiome studies had associated this species with increased CRC risk while in the present oral microbiome study no association was found between this species and CRC risk. We also found a lower abundance of *Carnobacteriaceae* and a higher prevalence of three genera and two species among CRC cases than among controls, while no study has investigated them in association with CRC risk.

We found that in the phylum *Bacteroidetes*, in addition to the oral pathogen *Prevotella intermedia*, abundance of *Prevotella denticola* and *Prevotella sp. oral taxon 300* and prevalence of *Prevotella sp. oral taxon 304* were also associated with an increased CRC risk, while abundance of *Prevotella melaninogenica* was associated with a decreased CRC risk. Among them, *Prevotella melaninogenica* and *Prevotella sp. oral taxon 300* showed independent associations based on mutual adjustment analyses. There are no reports available regarding the associations of these four species with any cancers. We also found that prevalence of nine taxa from other three phyla, namely *Proteobacteria*, *Spirochaetes* and *SR1*, were associated with an increased risk for CRC. However, no additional studies have investigated these taxa in relationship with CRC risk.

Most associations were observed in both African-Americans and European-Americans, while generally the associations were slightly more significant among African-Americans, e.g. the oral pathogens *Treponema denticola* and *Prevotella intermedia*. Considering that the sample size of African-Americans (N=515) was three times that of European-American participants (N=159) in the present study, these differential associations were not unexpected. However, several studies have demonstrated that there may be racial disparity in the microbiome of different human habitats, including the oral cavity [43]. Hence, the differential associations among African-Americans and European-Americans in this study may be also derived from the racial disparity in oral microbiome. For example, despite the smaller sample size of European-Americans, associations of the genus *Lautropia* and the species *Lactobacillus salivarius* with CRC risk were stronger among European-Americans than among African-Americans.

To the best of our knowledge, the present study is the first prospective study to investigate the influence of the oral microbiome on the risk of developing CRC. We conducted sensitivity analyses to exclude those patients who were diagnosed with CRC within two years after the collection of biological samples and the corresponding controls. The results did not change materially, indicating that our findings were not likely the result of reverse

causation. In the present study, 231 cases and 461 controls were included. We have 89% statistical power to detect an association for CRC risk, with an OR of 1.76 for bacterial taxa with the increment of taxa abundance. However, the power for rare bacteria taxa is limited. Especially for some bacterial species, there is a misclassification based on 16S rRNA sequencing data, which will further lower the statistical power. Nevertheless, the prospective design of the present study with stored pre-diagnostic oral rinse samples provides some of the first opportunities to examine whether the oral microbiota are predictive of near-term risk of one of the most common cancers afflicting the US population, especially African-Americans.

In summary, in this prospective nested case-control study, we found that multiple oral bacterial taxa were associated with subsequent CRC risk. These results raise the possibility that the oral microbiome may play an important role in CRC etiology. Further studies with a larger sample size are needed to confirm the identified associations and estimate the potential utilization of the oral microbiota for CRC early detection or prevention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors wish to thank all of the individuals who took part in the study, and all of the researchers, clinicians, technicians, and administrative staff who have enabled this work to be carried out. We thank Regina Courtney, Jie Wu, Jing He, and Marshal Younger for their help with sample preparation, statistical analysis and technical support for the project. The data analyses were conducted using the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University. Sample preparation was conducted at the Survey and Biospecimen Shared Resources, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485). The SCCS was supported by NIH grants R01CA92447 and U01CA202979. This project was also supported by the development fund from the Department of Medicine at Vanderbilt University Medical Center and the NIH supported grants R01CA207466 and U54CA163072. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abbreviations Used:

CRC	colorectal cancer
OR	odds ratio
CI	confidence interval
SCCS	Southern Community Cohort Study
CHC	community health centers
ICD	International Classification of Diseases
QC	quality control
OTU	Operational Taxonomic Units
HOMD	Human Oral Microbiome Database

QIIME	Quantitative Insights Into Microbial Ecology
ACCRE	Advanced Computing Center for Research and Education

REFERENCES

1. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RG, Barzi A, Jemal A. Colorectal cancer statistics, 2017. CA: a cancer journal for clinicians 2017;67: 177–93. [PubMed: 28248415]
2. Correa Lima MP, Gomes-da-Silva MH. Colorectal cancer: lifestyle and dietary factors. Nutr Hosp 2005;20: 235–41. [PubMed: 16045124]
3. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, Berry DA. Meta-analyses of colorectal cancer risk factors. Cancer Cause Control 2013;24: 1207–22.
4. Struzycka I The Oral Microbiome in Dental Caries. Pol J Microbiol 2014;63: 127–35. [PubMed: 25115106]
5. Liu B, Faller LL, Klitgord N, Mazumdar V, Ghodsi M, Sommer DD, Gibbons TR, Treangen TJ, Chang Y-C, Li S. Deep sequencing of the oral microbiome reveals signatures of periodontal disease. PloS one 2012;7: e37919. [PubMed: 22675498]
6. Wade WG. The oral microbiome in health and disease. Pharmacol Res 2013;69: 137–43. [PubMed: 23201354]
7. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. The human oral microbiome. J Bacteriol 2010;192: 5002–17. [PubMed: 20656903]
8. Ahn J, Segers S, Hayes RB. Periodontal disease, Porphyromonas gingivalis serum antibody levels and orodigestive cancer mortality. Carcinogenesis 2012: bgs112.
9. Ren H, Luu H, Cai H, Xiang Y, Steinwandel M, Gao Y, Hargreaves M, Zheng W, Blot W, Long J. Oral health and risk of colorectal cancer: results from three cohort studies and a meta-analysis. Ann Oncol 2016;27: 1329–36. [PubMed: 27217540]
10. Hiraki A, Matsuo K, Suzuki T, Kawase T, Tajima K. Teeth loss and risk of cancer at 14 common sites in Japanese. Cancer Epidemiology and Prevention Biomarkers 2008;17: 1222–7.
11. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. Nature Reviews Microbiology 2012;10: 575–82. [PubMed: 22728587]
12. Signat B, Roques C, Poulet P, Duffaut D. Role of Fusobacterium nucleatum in periodontal health and disease. Curr Issues Mol Biol 2011;13: 25–36. [PubMed: 21220789]
13. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA, Holt RA. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res 2012;22: 299–306. [PubMed: 22009989]
14. Kostic AD, Chun EY, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, El-Omar EM, Brenner D, et al. Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. Cell Host Microbe 2013;14: 207–15. [PubMed: 23954159]
15. Flynn KJ, Baxter NT, Schloss PD. Metabolic and Community Synergy of Oral Bacteria in Colorectal Cancer. mSphere 2016;1: e00102–16. [PubMed: 27303740]
16. Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, Buchowski MS, Arnold CW, McLaughlin JK, Blot WJ. Southern community cohort study: establishing a cohort to investigate health disparities. Journal of the National Medical Association 2005;97: 972. [PubMed: 16080667]
17. Signorello LB, Hargreaves MK, Blot WJ. The Southern Community Cohort Study: Investigating Health Disparities. J Health Care Poor U 2010;21: 26–37.
18. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software]. Available at https:/githubcom/najoshi/sickle 2011.
19. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics 2013;14 Suppl 1: S7.

20. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. Bmc Bioinformatics 2012;13: 31. [PubMed: 22333067]
21. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic acids research 2015: gku1341.
22. Chen T, Yu W-H, 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 (Oxford) 2010;2010: baq013. [PubMed: 20624719]
23. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7: 335–6. [PubMed: 20383131]
24. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecological monographs 1957;27: 325–49.
25. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. The ISME journal 2011;5: 169. [PubMed: 20827291]
26. Tang Z-Z, Chen G, Alekseyenko AV. PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. Bioinformatics 2016;32: 2618–25. [PubMed: 27197815]
27. 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. [PubMed: 15853938]
28. Newman MG, Takei H, Klokkevold PR, Carranza FA. Carranza's clinical periodontologyed: Elsevier health sciences, 2011.
29. Fernandes AD, Reid JN, 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. [PubMed: 24910773]
30. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall L-I, McDonald D. Best practices for analysing microbiomes. Nature Reviews Microbiology 2018: 1.
31. Galwey NW. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. Genet Epidemiol 2009;33: 559–68. [PubMed: 19217024]
32. Kato I, Vasquez AA, Moyerbrailean G, Land S, Sun J, Lin H-S, Ram JL. Oral microbiome and history of smoking and colorectal cancer. Journal of Epidemiological Research 2016;2: p92.
33. Flemer B, Warren RD, Barrett MP, Cisek K, Das A, Jeffery IB, Hurley E, O'Riordain M, Shanahan F, O'Toole PW. The oral microbiota in colorectal cancer is distinctive and predictive. Gut 2017: gutjnl-2017–314814.
34. Shah N, Tang H, Doak TG, Ye Y. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics Biocomputing 2011 ed, 2011: 165–76.
35. Narikiyo M, Tanabe C, Yamada Y, Igaki H, Tachimori Y, Kato H, Muto M, Montesano R, Sakamoto H, Nakajima Y. Frequent and preferential infection of Treponema denticola, Streptococcus mitis, and Streptococcus anginosus in esophageal cancers. Cancer science 2004;95: 569–74. [PubMed: 15245592]
36. Kamarajan P, Ateia I, Shin JM, Fenno JC, Kapila YL. Treponema denticola, a periodontal pathogen, promotes stemness and migration in oral squamous cell carcinoma: AACR, 2016.
37. Peters BA, Wu J, Pei Z, Yang L, Purdue MP, Freedman ND, Jacobs EJ, Gapstur SM, Hayes RB, Ahn J. Oral Microbiome composition reflects prospective risk for esophageal cancers. Cancer research 2017;77: 6777–87. [PubMed: 29196415]
38. 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. [PubMed: 18541739]

39. Ahn J, Chen CY, Hayes RB. Oral microbiome and oral and gastrointestinal cancer risk. Cancer Causes & Control 2012;23: 399–404. [PubMed: 22271008]

40. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, Goedert JJ, Hayes RB, Yang L. Human gut microbiome and risk for colorectal cancer. Journal of the National Cancer Institute 2013;105: 1907–11. [PubMed: 24316595]

41. Candela M, Turroni S, Biagi E, Carbonero F, Rampelli S, Fiorentini C, Brigidi P. Inflammation and colorectal cancer, when microbiota-host mutualism breaks. World journal of gastroenterology: WJG 2014;20: 908. [PubMed: 24574765]

42. Yu J, Feng Q, Wong SH, Zhang D, yi Liang Q, Qin Y, Tang L, Zhao H, Stenvang J, Li Y. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 2015: gutjnl-2015–309800.

43. Li J, Quinque D, Horz H-P, Li M, Rzhetskaya M, Raff JA, Hayes MG, Stoneking M. Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. BMC microbiology 2014;14: 316. [PubMed: 25515234]

Novelty and Impact:

In this prospective nested case-control study, multiple oral bacterial taxa were found to be associated with subsequent risk of developing colorectal cancer (CRC), including two previously identified oral pathogenic bacteria, *Treponema denticola* and *Prevotella intermedia*. This study demonstrates that oral microbiome may play a significant role in CRC etiology.

Table 1.

Characteristics of study participants in the Southern Community Cohort Study

Characteristic	Group	Cases (*N*=231)	Controls (*N*=461)	*P* value [a]
Race				1.00
	European-Americans	53 (22.94%)	106 (22.99%)	
	African-Americans	172 (74.46%)	343 (74.41%)	
	Others	6 (2.60%)	12 (2.60%)	
Sex				1.00
	Male	93 (40.26%)	185 (40.13%)	
	Female	138 (59.74%)	276 (59.87%)	
Age at enrollment				0.93
	40 – 49	51 (22.08%)	111 (24.08%)	
	50 – 59	96 (41.56%)	186 (40.35%)	
	60 – 69	61 (26.41%)	116 (25.16%)	
	70 – 79	23 (9.96%)	48 (10.41%)	
BMI				0.96
	< 18.5	2 (0.88%)	3 (0.66%)	
	18.5 – 24.9	52 (23.01%)	103 (22.79%)	
	25.0 – 29.9	71 (31.42%)	136 (30.09%)	
	≥ 30	101 (44.69%)	210 (46.46%)	
Education				0.86
	<High school	59 (25.65%)	125 (27.35%)	
	High/Vocational school	85 (36.96%)	164 (35.89%)	
	Some college	45 (19.57%)	80 (17.51%)	
	College	41 (17.83%)	88 (19.26%)	
Annual household income ($)				0.30
	<15,000	117 (51.54%)	226 (50.11%)	
	15,000–24,999	34 (14.98%)	90 (19.96%)	
	25,000–49,999	42 (18.50%)	69 (15.30%)	
	50,000–100,000	27 (11.89%)	44 (9.76%)	
	>100,000	7 (3.08%)	22 (4.88%)	
Tobacco smoking				1.00
	Current	64 (27.71%)	128 (27.77%)	
	Former	74 (32.03%)	148 (32.10%)	
	Never	93 (40.26%)	185 (40.13%)	
Smoking pack-year [b]		21.99 ± 19.34	25.36 ± 28.98	0.17
Alcohol consumption [c]				0.03
	None	137 (60.35%)	223 (49.78%)	
	Light	52 (22.91%)	139 (31.03%)	
	Moderate	18 (7.93%)	52 (11.61%)	
	Heavy	20 (8.81%)	34 (7.59%)	

Characteristic	Group	Cases (*N*=231)	Controls (*N*=461)	*P* value [a]
Tooth loss				0.28
	None	16 (6.93%)	53 (11.50%)	
	1–4	38 (16.45%)	88 (19.09%)	
	5–10	21 (9.09%)	82 (17.79%)	
	>10, not all	28 (12.12%)	82 (17.79%)	
	All	28 (12.12%)	57 (12.36%)	
	Unknown	100 (43.29%)	99 (21.48%)	
Tooth Decay				0.17
	0	58 (25.11%)	157 (34.06%)	
	1–2	28 (12.12%)	70 (15.18%)	
	3–5	18 (7.79%)	58 (12.58%)	
	≥6	2 (0.87%)	24 (5.21%)	
	No teeth	29 (12.55%)	63 (13.67%)	
	Unknown	96 (41.56%)	89 (19.31%)	
Gingivitis				1.00
	With	31 (13.42%)	87 (18.87%)	
	Without	106 (45.89%)	290 (62.91%)	
	Unknown	94 (40.69%)	84 (18.22%)	

[a] *P* values were calculated through two-sided χ^2 test or t-test with missing values excluded

[b] Mean ± SD were reported for smoking pack year among current- and former-smokers

[c] Light, <1 drink per day; Moderate, 1–2 drink per day; Heavy, >2 drinks per day

Table 2.

Associations of periodontal pathogens prevalence with CRC risk, from the Southern Community Cohort Study

Periodontal pathogens	Cases (N=231)	Controls (N=461)	OR (95% CI) [a]	*P* value [a]	*P* value [b]
Treponema denticola					
Non-carriers	78 (33.77%)	201 (43.60%)	Ref		
Carriers	153 (66.23%)	260 (56.40%)	1.76 (1.19–2.60)	4.45×10^{-3}	0.01
Prevotella intermedia					
Non-carriers	82 (35.50%)	201 (43.60%)	Ref		
Carriers	149 (64.50%)	260 (56.40%)	1.55 (1.08–2.22)	0.02	0.05
Porphyromonas gingivalis					
Non-carriers	81 (35.06%)	165 (35.79%)	Ref		
Carriers	150 (64.94%)	296 (64.21%)	1.05 (0.73–1.49)	0.80	1.00
Tannerella forsythia					
Non-carriers	81 (35.06%)	172 (37.31%)	Ref		
Carriers	150 (64.94%)	289 (62.69%)	1.11 (0.76–1.61)	0.58	1.00
Fusobacterium nucleatum					
Non-carriers	1 (0.43%)	2 (0.43%)	Ref		
Carriers	230 (99.57%)	459 (99.57%)	1.12 (0.1–12.65)	0.93	1.00

[a] For each pathogen, individuals were categorized into carriers and non-carriers according to whether they carried the pathogen or not. The association of pathogen prevalence with CRC risk was evaluated using conditional logistic regression. Smoking pack-years, alcohol consumption status and sequencing depth were adjusted in the models.

[b] Bonferroni-corrected *P* values, adjusted for three independent tests.

Table 3.

Common bacterial taxa associated with colorectal cancer risk [a]

Taxa	Median relative abundance		OR (95% CI) [b]	*P* value [b]	*P* value [c]
	Cases (N=231)	Controls (N=461)			
Phylum *Actinobacteria*					
Family *Bifidobacteriaceae*	0.09%	0.05%	1.10 (1.01–1.19)	0.03	1.00
Phylum *Bacteroidetes*					
Species *Prevotella denticola*	0.14%	0.07%	1.11 (1.02–1.20)	0.02	1.00
Species *Prevotella melaninogenica*	2.03%	2.12%	0.91 (0.84–0.99)	0.04	1.00
Species *Prevotella sp. oral taxon 300*	0.04%	0.02%	1.13 (1.01–1.26)	0.04	1.00
Phylum *Firmicutes*					
Family *Carnobacteriaceae*	1.22%	1.39%	0.85 (0.72–0.99)	0.04	1.00
Family *Streptococcaceae*	33.45%	36.23%	0.73 (0.56–0.96)	0.02	1.00
Genus *Streptococcus*	33.40%	35.92%	0.73 (0.55–0.96)	0.02	1.00
Species *S. sp. oral taxon 058*	16.66%	18.59%	0.79 (0.67–0.94)	7.87×10^{-3}	0.42
Family *Erysipelotrichaceae*	0.09%	0.09%	0.87 (0.76–0.99)	0.04	1.00
Genus *Solobacterium*	0.07%	0.07%	0.87 (0.76–0.98)	0.02	1.00
Species *S. moorei*	0.07%	0.07%	0.87 (0.77–0.99)	0.03	1.00

[a] Common taxa were defined as those with a median relative abundance of >0.01% among control subjects

[b] For each sample, centered log-ratio transformation was used to normalize taxa counts at each taxonomic level after adding a pseudocount of 1. The associations of taxon abundance with CRC risk was evaluated using conditional logistic regression. Smoking pack-years and alcohol consumption status were adjusted in the models

[c] Bonferroni-corrected *P* values, adjusted for 53 independent tests

Table 4.

Rare bacterial taxa associated with colorectal cancer risk [a]

Taxa	Prevalence		OR (95% CI) [b]	*P* value [b]	*P* value [c]
	Cases (N=231)	Controls (N=461)			
Phylum *Actinobacteria*					
Species *Bifidobacterium dentium*	58.87%	50.54%	1.46 (1.04–2.07)	0.03	1.00
Phylum *Bacteroidetes*					
Species *Prevotella sp. oral taxon 304*	33.77%	26.03%	1.59 (1.09–2.33)	0.02	1.00
Phylum *Firmicutes*					
Genus *Peptococcus*	60.61%	52.06%	1.46 (1.02–2.08)	0.04	1.00
Genus *Anaeroglobus*	69.70%	60.74%	1.48 (1.04–2.10)	0.03	1.00
Genus *Mitsuokella*	44.16%	36.01%	1.52 (1.08–2.14)	0.02	1.00
Species *Lactobacillus salivarius*	46.75%	38.18%	1.46 (1.03–2.08)	0.03	1.00
Species *Eubacterium yurii*	37.66%	29.93%	1.46 (1.01–2.10)	0.04	1.00
Phylum *Proteobacteria*					
Family *Burkholderiaceae*	66.67%	55.97%	1.62 (1.14–2.30)	7.40×10^{-3}	0.68
Genus *Lautropia*	66.23%	54.45%	1.72 (1.20–2.45)	2.88×10^{-3}	0.26
Species *Neisseria oralis*	42.42%	34.71%	1.42 (1.01–2.00)	0.04	1.00
Species *Campylobacter sp. oral taxon 044*	51.52%	42.08%	1.58 (1.12–2.24)	0.01	0.92
Phylum *Spirochaetes*					
Species *Treponema lecithinolyticum*	44.59%	33.62%	1.76 (1.23–2.53)	2.11×10^{-3}	0.19
Species *Treponema sp. oral taxon 250*	25.97%	20.17%	1.66 (1.07–2.56)	0.02	1.00
Phylum *SR1*	54.98%	41.65%	1.76 (1.25–2.47)	1.09×10^{-3}	0.10
Species *SR1_[G-1] sp. oral taxon 345*	33.77%	26.90%	1.50 (1.01–2.23)	0.04	1.00
Species *SR1_[G-1] sp. oral taxon 874*	41.13%	29.07%	1.75 (1.24–2.48)	1.58×10^{-3}	0.14

[a] Rare taxa were defined as those with a median relative abundance of ≤0.01% and a carriage >20% among control subjects.

[b] For each taxon, individuals were categorized into carriers and non-carriers according to whether they carried the taxon or not. The association of taxon prevalence and CRC risk was evaluated using conditional logistic regression. Smoking pack-years, alcohol consumption status and sequencing depth were adjusted in the models.

[c] Bonferroni-corrected *P* values, adjusted for 92 independent tests.