Title

An international report on bacterial communities in esophageal squamous cell carcinoma

Journal

International Journal of Cancer, 151(11)

ISSN

0020-7136

Authors

Nomburg, Jason
Bullman, Susan
Nasrollahzadeh, Dariush
et al.

Publication Date

2022-12-01

DOI

10.1002/ijc.34212

Peer reviewed

HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2024 May 17.

Published in final edited form as:

Int J Cancer. 2022 December 01; 151(11): 1947–1959. doi:10.1002/ijc.34212.

An international report on bacterial communities in esophageal squamous cell carcinoma

Jason Nomburg[1,2,3], **Susan Bullman**[4], **Dariush Nasrollahzadeh**[5,6], **Eric A. Collisson**[7,8], **Behnoush Abedi-Ardekani**[6], **Larry O. Akoko**[9], **Joshua R. Atkins**[6], **Geoffrey C. Buckle**[7,8], **Satish Gopal**[10], **Nan Hu**[11], **Bongani Kaimila**[12], **Masoud Khoshnia**[5], **Reza Malekzadeh**[5], **Diana Menya**[13], **Blandina T. Mmbaga**[14,15], **Sarah Moody**[16], **Gift Mulima**[17], **Beatrice P. Mushi**[9], **Julius Mwaiselage**[18], **Ally Mwanga**[9], **Yulia Newton**[19], **Dianna L. Ng**[7,20], **Amie Radenbaugh**[19], **Deogratias S. Rwakatema**[14,15], **Msiba Selekwa**[9], **Joachim Schüz**[21], **Philip R. Taylor**[11], **Charles Vaske**[19], **Alisa Goldstein**[11], **Michael R. Stratton**[16], **Valerie McCormack**[21], **Paul Brennan**[6], **James A. DeCaprio**[1,3,22], **Matthew Meyerson**[1,2,22,23,*], **Elia J. Mmbaga**[9,24,*], **Katherine Van Loon**[7,8,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA

[2]Broad Institute of MIT and Harvard, Cambridge, MA

[3]Harvard Program in Virology, Harvard Medical School, Boston, MA

[4]Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

[*]**Correspondence to:** Katherine Van Loon - Katherine.VanLoon@ucsf.edu, Elia J. Mmbaga - ejmmbaga@medisin.uio.no, Matthew Meyerson - matthew_meyerson@dfci.harvard.edu.

Author Contributions

Conceptualization - K.V.L., E.J.M., M.M.

Data curation - J.N., Y.N., C.V.

Formal analysis - J.N.

Funding acquisition - K.V.L., S.B., J.A.D., M.M., M.R.S., P.B.

Investigation - J.N.

Methodology - J.N., K.V.L., M.M., E.J.M.

Project Administration - J.N., K.V.L.

Resources - D.N., E.A.C., B.A., L.O.A., J.R.A., G.C.B., S.G., N.H., B.K., M.K., R.M., D.M., B.T.M., S.M., G.M., B.P.M., J.M., A.M., Y.N., D.L.N., A.R., D.S.R., M.S., J.S., P.R.T., C.V., A.G., M.R.S., V.M., P.B., E.J.M., K.V.L.

Software - J.N.

Supervision - K.V.L., E.J.M., M.M., J.A.D.

Validation - J.N.

Visualization - J.N.

Writing - original draft - J.N.

Writing - review and editing - J.N., S.B., D.N., E.A.C., B.A., L.O.A., J.R.A., G.C.B., S.G., N.H., B.K., M.K., R.M., D.M., B.T.M., S.M., G.M., B.P.M., J.M., A.M., Y.N., D.L.N., A.R., D.S.R., M.S., J.S., P.R.T., C.V., A.G., M.R.S., V.M., P.B., J.A.D., M.M., E.J.M., K.V.L.

The work reported in the paper has been performed by the authors, unless clearly specified in the text.

CONFLICT OF INTEREST

M.M. receives research support from Bayer, Janssen, and Ono, has patents licensed to Bayer and Labcorp, is a consultant for Bayer, Interline, and Isabl, and has equity in Interline and Isabl. J.A.D. has received research support from Rain Therapeutics, Inc. J.A.D. has served as consultant to Rain Therapeutics, Inc. and Takeda, Inc. S.B. has consulted for GlaxoSmithKline and BioMx, and is on the cancer program scientific advisory board for BioMx. D.L.N. receives research funding from Cepheid, Inc., unrelated to this article. E.A.C. received expert consulting fees arising from an unrelated matter from Nant Bio in 2021 and 2022. C.V. is a shareholder of NantHealth. The other authors declare no conflict of interest exist.

Ethics Statement

This paper describes data that has already been collected. All studies were previously approved by the relevant ethical review boards and all patients acknowledged their informed consent, details of which can be found in the relevant publications.

[5]Digestive Oncology Research Center, Digestive Disease Research Institute, Tehran University of Medical Sciences, Shariati Hospital. Tehran Iran.

[6]International Agency for Research on Cancer (IARC/WHO), Genomic Epidemiology Branch, Lyon, France

[7]University of California, San Francisco (UCSF) Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA

[8]Division of Hematology/Oncology, Department of Medicine, UCSF, San Francisco, California, USA

[9]Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

[10]University of North Carolina, Chapel Hill, North Carolina, USA

[11]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

[12]UNC Project - Lilongwe, Malawi

[13]School of Public Health, Moi University, Eldoret, Kenya

[14]Kilimanjaro Clinical Research Institute, Kilimanjaro Christian Medical Centre, Moshi, Tanzania

[15]Kilimanjaro Christian Medical University College, Moshi, Tanzania

[16]Cancer, Ageing and Somatic Mutation, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK

[17]Kamuzu Central Hospital, Lilongwe, Malawi

[18]Ocean Road Cancer Institute, Dar es Salaam, Tanzania

[19]NantOmics/NantHealth, Inc., El Segundo, California, USA

[20]Department of Pathology, UCSF, San Francisco, CA, USA

[21]International Agency for Research on Cancer (IARC/WHO), Environment and Lifestyle Epidemiology Branch, Lyon, France

[22]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

[23]Department of Genetics, Harvard Medical School, Boston, MA

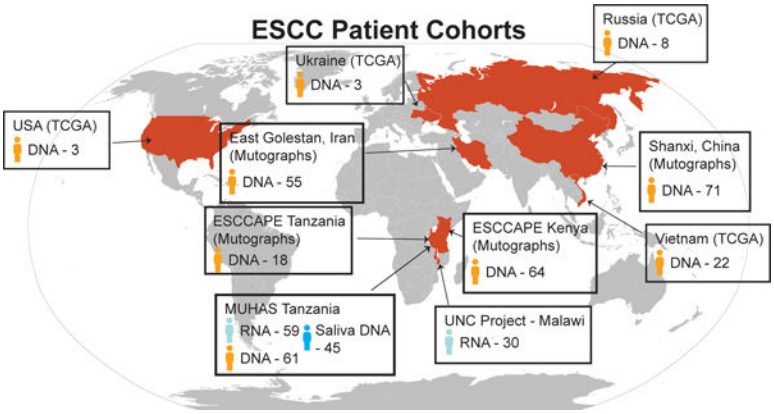
[24]Department of Community Medicine and Global Health, University of Oslo, Norway

Abstract

The incidence of esophageal squamous cell carcinoma (ESCC) is disproportionately high in the eastern corridor of Africa and parts of Asia. Emerging research has identified a potential association between poor oral health and ESCC. One possible link between poor oral health and ESCC involves the alteration of the microbiome. We performed an integrated analysis of four independent sequencing efforts of ESCC tumors from patients from high- and low-incidence regions of the world. Using whole genome sequencing (WGS) and RNA sequencing (RNAseq) of ESCC tumors from 61 patients in Tanzania, we identified a community of bacteria, including members of the genera *Fusobacterium, Selenomonas, Prevotella, Streptococcus,*

Porphyromonas, Veillonella, and *Campylobacter*, present at high abundance in ESCC tumors. We then characterized the microbiome of 238 ESCC tumor specimens collected in two additional independent sequencing efforts consisting of patients from other high-ESCC incidence regions (Tanzania, Malawi, Kenya, Iran, China). This analysis revealed similar ESCC-associated bacterial communities in these cancers. Because these genera are traditionally considered members of the oral microbiota, we next explored whether there was a relationship between the synchronous saliva and tumor microbiomes of ESCC patients in Tanzania. Comparative analyses revealed that paired saliva and tumor microbiomes were significantly similar with a specific enrichment of *Fusobacterium* and *Prevotella* in the tumor microbiome. Together, these data indicate that cancer-associated oral bacteria are associated with ESCC tumors at the time of diagnosis and support a model in which oral bacteria are present in high abundance in both saliva and tumors of some ESCC patients.

Graphical Abstract



Keywords

ESCC; microbiome; *Fusobacterium*

INTRODUCTION

Esophageal cancer is the sixth most common cause of cancer-related death worldwide (1). There are two histologic subtypes of esophageal cancer with distinct biological characteristics, geographic distributions, and risk factors (2). Esophageal adenocarcinoma is the most common histologic form of esophageal cancer in high-income countries and is associated with factors including gastroesophageal reflux disease, Barrett's esophagus, and obesity (3, 4). By contrast, esophageal squamous cell carcinoma (ESCC) represents more than 90% of worldwide esophageal cancer cases and is the dominant histology in low-resource settings. In particular, there are two main regions where ESCC is endemic: (1) the Asian esophageal cancer belt, extending from western/northern China to central and southeast Asia; and (2) the eastern corridor of Africa, extending from Ethiopia to South Africa (5, 6).

Emerging research has identified a possible association between poor oral health and ESCC. Studies from Asia, Europe, Latin America, Kenya, and Iran have reported associations of ESCC with poor oral hygiene, chronic periodontal disease, dental decay, and tooth loss (7–16). Recently, three parallel case-control studies in Kenya and Tanzania, conducted as part of the African Esophageal Cancer Consortium (AfrECC) and ESCCAPE (esccape.iarc.fr) collaborations, reported possible associations of poor or infrequent oral hygiene with increased risk for ESCC in East Africa (17–20).

Alterations of the oral microbiome due to poor oral health is one proposed biological pathway that could explain the link between oral health and ESCC. Many bacterial genera associated with gastrointestinal cancers contain species that are traditionally associated with healthy or diseased oral microbiomes. For example, *Helicobacter pylori* was discovered to be associated with gastric cancers and mucosa-associated lymphoid tissue (MALT) lymphomas, indirectly by promoting gastric inflammation and directly by influencing cellular signaling (21). Similarly, bacteria of the genera *Fusobacterium*, *Selenomonas*, and *Prevotella* are enriched in colorectal cancers (22–24) and can be visualized invasively within tumor tissue (25). *Fusobacterium*, in particular, has been reported to promote carcinogenesis through the selective expansion or inhibition of certain classes of immune cells (26) and may drive cellular proliferation by stimulating Wnt/β-catenin signaling (27, 28). Other bacterial genera such as *Porphyromonas, Campylobacter,* and *Streptococcus* have emerging associations with various human gastrointestinal cancers (29–35).

As part of ongoing investigation into the microbiome's association with ESCC, we performed an integrated analysis of four independent sequencing efforts including ESCC tumors from patients from both high- and low-incidence regions of the world. In addition, we investigated the relationship between the microbiomes of matched ESCC tumors and saliva specimens in a subset of ESCC cases.

MATERIALS AND METHODS

Samples and sequencing efforts

The origin and number of samples is as follows: MUHAS Tanzania cohort (n=61) (36), UNC Project - Malawi cohort (n=30) (37). The Mutographs study (n=210) (38) originated from patients in Golestan, Iran (n=55), ESCCAPE case-control studies in Tanzania (n=18) (19) and Kenya (n=64) (17), and patients in Shanxi, China (n=71). While there are over 500 samples in the full Mutographs study, our investigation was limited to a subsample of 210 samples. TCGA ESCC (n=36) and COAD samples (n=51) have been previously described (39, 40). The TCGA ESCC cohort includes tumors from patients in United States (n=3), Ukraine (n=3), Vietnam (n=22), and Russia (n=8), regions which have lower incidence of ESCC. Information on sample acquisition and sequencing for the UNC Project - Malawi, Mutographs, and TCGA patient samples can be identified in their relevant publications (19, 36–40). Details on each sequencing cohort are also listed in Table 1.

We provide sequencing statistics on studied samples in Supplementary Table 1. This table includes the following information for each sample: the total number of reads, the total number of reads that passed quality and complexity filtration during GATK-PathSeq

analysis, the number of human-mapped reads, the number of reads mapped to the microbial reference database, and the number of unmapped reads.

MUHAS Tanzania sample collection and sequencing

DNA and RNA were extracted from tumor samples using the Qiagen AllPrep kit. Saliva samples were collected in the Oragene DISCOVER ORG-500 tube and DNA was extracted using the Oragene PrepIT L2P extraction kit. Saliva samples were collected at the time of tumor collection. RNA integrity was assessed using an Agilent bioanalyzer, and RNA and DNA quantity was assessed by Nanodrop and PicoGreen (Invitrogen) methods. DNA sequencing libraries were prepared with the KAPA Hyper Prep kit, and samples were sequenced to a target depth of 60x coverage (for tumor) or 30x coverage (for saliva) on an Illumina NovaSeq. RNA isolations with RIN > 7 were prepared with the KAPA Stranded RNAseq with RiboErase kit and sequenced on an Illumina HiSeq or NovaSeq to a target of 200 million 150bp paired reads. Additional methodologic details can be found in the associated publication (36).

Bacterial identification and quantification

GATK-PathSeq (41) was used to conduct computational subtraction of human-mapping reads from input RNAseq and WGS datasets. GATK-PathSeq initially maps reads to a host reference database consisting of the human genome grch38 and various supplemental human reference sequences. Next, non-human reads are mapped against a comprehensive microbial database, and microbe read assignments are reported for further study.

For all analyses, we used the GATK-PathSeq “score”, which is roughly equivalent to read count. The GATK-PathSeq score is calculated by: 1) distributing a read count to multiple species if the read maps to multiple species (e.g. if a read maps to two species, each species gets +0.5 of a read); and 2) normalizing the read count to genome length (only in WGS data). The score is calculated at the taxonomic level of the entry in the reference (usually a species or strain). The genus score is the sum of scores assigned to each species in that genus.

Bacterial abundance analyses and plotting were conducted in R (v3.5.1). To calculate relative abundance at a phylogenetic level (e.g., phylum or genus), GATK-PathSeq results were filtered for taxa at the level, and relative abundance was calculated for each taxon as follows: (# of taxon reads)/(total # reads at the selected phylogenetic level). The rows of all bacterial abundance heatmaps are arranged according to the mean abundance across all samples. The sample order of relative abundance stacked barplots were determined based on *Fusobacterium* genus relative abundance except where noted. In Figure 2D, if any cohort contained more than 50 samples, 50 samples were randomly selected for plotting. The distribution of relative abundances of genera of interest in all samples can be found in Figure S2, where width of each violin represents the relative distribution of observed bacterial relative abundance for all patients in each patient cohort.

Jaccard distance between RNAseq and WGS data from each ESCC tumor was calculated in R based on bacterial genera with at least 1% relative abundance. The qualitative Jaccard

index was used in this case because the comparison was between DNA and RNA analytes which would not be expected to be quantitatively identical.

Tumor-saliva similarity

Only tumor-saliva pairs from the MUHAS Tanzania cohort with at least 10,000 reads mapped to the bacterial superkingdom were analyzed for similarity (N=21). Bray-Curtis dissimilarity metrics between tumor-oral pairs were calculated using the R package vegan (42). Figure 3A presents the Bray-Curtis *similarity* (1 – Bray-Curtis dissimilarity), for each tumor-oral pair.

To determine the correlation between the relative abundance of specific genera between tumor and saliva microbiomes, common-abundant genera with at least 1% abundance in at least 3 tumor-saliva pairs were identified. This resulted in the identification of 16 common-abundant genera. Correlations represent a two-sided Pearson correlation coefficient. To determine tumor-saliva enrichment of common-abundant genera, the difference in relative abundance of each genus between each tumor-saliva pair was plotted (Figure 3C). For the relative abundance bar plots of tumor-saliva pairs (Figure 3D), bacterial genera that had been highlighted in previous figures are labeled.

Bacterial contaminant identification

We identified potential bacterial contaminants with two approaches. First, we used a bacteria genera blacklist identified by Salter et al. and Poore et al. (43, 44) as bacteria commonly present in negative-blank controls. Second, we implemented the Decontam program (45). Decontam works on the assumption that the relative abundance of contaminants should be inversely proportional to the library concentration, and thus incorporates information on library concentrations prior to pooling for sequencing. We ran Decontam on samples from the Mutographs cohort with an aggressive p-value threshold of $p < 0.2$ and the “frequency” method. The code used to run Decontam can be found in running_decontam.Rmd. We then removed the blacklist and Decontam-called bacterial genera from each dataset, and re-generated Figure 2 as Figure S3B–E. The code used to do this analysis is present at Fig_S3_blacklist_and_decontam.Rmd.

We implemented Snippy (v4.4.5) (https://github.com/tseemann/snippy) for variant calling, and then FastTree (46) for constructing phylogenetic trees of *Fusobacterium* isolates. We first ran snippy-multi, followed by rerunning snippy-core on samples with sufficient single nucleotide polymorphisms (SNPs) relative to the reference. We implemented Snippy only on Fusobacterium-high samples that had sufficient SNPs, resulting in 4 samples from MUHAS Tanzania and 5 from ESCCAPE Kenya. As the reference *Fusobacterium* strain, we used *Fusobacterium nucleatum* subspecies *polymorphum* (accession NZ_LN831027.1). We included WGS reads from genetically-similar tumor-oral *Fusobacterium* pairs previously sequenced by Abed et al. (47) in each run of snippy-core.

Batch effects

Batch effects arise during separate collection, processing, and sequencing of samples. In the context of microbiome studies, this will result in differences in the amount and identity

of sequencing reads from contaminating genera. As noted above, because we do not have the controls necessary to concretely identify contaminating bacteria, we do not take action to remove contaminants. We acknowledge that batch effect may influence the relative abundance of non-contaminant bacteria (e.g. as a consequence of DNA or RNA extraction), meaning that there may be some minor quantitative differences as a result.

RESULTS

Study Population

To evaluate the potential role of the host microbiota in ESCC, we investigated the microbiome of 299 ESCC specimens from patients in five different countries with a high incidence of ESCC. Specimens were collected through four independent sequencing efforts (Figure 1A). Specimens consisted of whole genome sequencing (WGS) and RNA sequencing (RNAseq) data from the tumor and saliva of 61 patients from Tanzania (the "MUHAS Tanzania" cohort) (36), RNAseq data from the tumors of 30 ESCC patients in Malawi (the "UNC Project – Malawi" cohort) (37), and WGS from 208 additional samples of tumors from patients in high ESCC incidence regions, including specimens from ESCC patients in Tanzania (n=18) and Kenya (n=64) that were collected in the ESCCAPE studies (esccape.iarc.fr) and specimens from ESCC patients in East Golestan, Iran (n=55) and Shanxi, China (n=71) that were sequenced as part of the Cancer Research UK Mutographs project ("Mutographs" cohorts) (38). In addition, we analyzed WGS data of ESCC from The Cancer Genome Atlas (39), which includes a small number of tumors from patients in low-incidence geographic regions including the United States (n=3), Ukraine (n=3), Vietnam (n=22), and Russia (n=8) (the "TCGA" cohort). Patient characteristics are shown in Table 2.

Bacterial populations are abundant and diverse in ESCC tumors

We used the metagenomic analysis tool GATK-PathSeq (41) to process the RNAseq and WGS data. GATK-PathSeq uses a sequential mapping strategy to assign reads to human and microbial reference genomes, resulting in detailed information on sequencing reads of human and microbial origin (Figure S1A). We likewise used GATK-PathSeq to process WGS data sets from 50 colon adenocarcinoma (COAD) specimens available from TCGA (40) for comparison, as there is strong evidence of microbial associations with COAD (22–25).

The bacterial burden of ESCC tumors ranged from 10 to 1000 bacterial reads per million human reads, similar to numbers observed in TCGA COAD (Figure 1B). Furthermore, the Shannon diversity of bacterial populations at the genus level ranged from 2 to 3 (Figure 1C). By comparison, ESCC-associated bacterial communities are as diverse or more diverse than TCGA COAD. At the phylum level, ESCC bacterial populations generally consist of *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Actinobacteria*, and *Fusobacteria* (Figure 1D, Figure S1B). Of note, the higher than expected abundance of the phylum *Actinobacteria* specifically in the TCGA ESCC samples is attributable, in particular, to a very high abundance of the genus *Tetrasphaera* (Figure S1C). This is evidenced by a depressed Shannon diversity of *Actinobacteria* genera in these samples (Figure S1D) and may indicate

contamination of the TCGA ESCC samples. *Actinobacteria* have been reported as a source of contaminating reads in TCGA gastrointestinal cancer samples (48).

Bacterial genera associated with carcinogenesis are observed at high relative abundance in ESCC tumors from Tanzania

To determine if bacteria with known associations with cancer are present in ESCC, we first analyzed the sequencing series of the 61 ESCC cases from the MUHAS Tanzania cohort with both WGS and RNAseq data. The paired WGS and RNAseq data from these tumors allowed investigation of bacterial communities at the DNA and RNA levels. Both WGS and RNAseq data revealed high relative abundance of bacterial genera previously associated with carcinogenesis in these ESCC tumors (Figure 2A, 2B). The high relative abundance of the *Fusobacterium* genus was particularly notable. Other bacterial genera of interest include *Streptococcus*, *Porphyromonas*, *Campylobacter*, *Prevotella*, *Veillonella*, and *Selenomonas*, many of which have been associated with gastrointestinal malignancies alongside or independently of *Fusobacterium* (25, 29, 32, 34, 49). The mean Jaccard similarity index between tumor RNAseq and WGS data from the same tumor is 0.54, greater than the average Jaccard similarity index of random RNAseq-WGS pairs (0.36), indicating that bacterial populations inferred from WGS and RNAseq data are generally consistent (Figure 2C).

Next, we attempted to determine if similar bacterial genera were also present in ESCC from patients in high-incidence countries beyond Tanzania. Investigation of RNA sequencing data from patients in Malawi, WGS data from patients in Kenya, China, and Iran, as well as from the independent ESCCAPE Tanzania patient group revealed pervasive evidence of similar bacterial genera in the tumors of these patients (Figure 2D, Figure S2). To investigate if similar microorganisms were found in ESCC tumors from patients in low-incidence regions, we investigated WGS data from ESCC tumors originating from USA, Ukraine, Vietnam, and Russia that were available through TCGA. While the number of ESCC tumors available from low-incidence regions was low and relies on a single sequencing effort, we found that the tumors of many of these patients contain similar bacterial genera (Figure 2D, Figure S2). Colon cancers from the TCGA COAD cohort revealed evidence of *Fusobacterium,* as expected; however, these COAD samples were notable for much lower relative abundance of the other genera of interest, when compared to ESCC tumors.

Assessing contamination

To understand if bacterial contaminants were influencing our results, we identified potential contaminating bacterial genera and assessed the impact of their removal. We identified potential contaminating genera in two ways: 1) bacteria that had been identified by Settler et al. and Poore et al. as potential contaminants present in “negative-blank” sequencing controls (43, 44); and 2) bacteria that were identified by the Decontam algorithm (45) as potential contaminants. An average of 1.58% of bacterial genus reads belonged to bacterial genera present in the blacklist (Figure S3A). We were able to implement the Decontam algorithm only on the Mutographs samples (ESCCAPE Tanzania; ESCCAPE Kenya; Shanxi, China; and Golestan, Iran). In these samples, Decontam-called bacterial genera represented an average of 12.2% of bacterial genus reads (Figure S3A). We then

assessed the impact of the removal of blacklist and Decontam-identified bacterial genera on our results and found that the removal of these bacterial genera only had a marginal effect that did not change our conclusions (Figure S3B–E).

While the identification of *Fusobacterium, Streptococcus*, *Porphyromonas*, *Campylobacter*, *Prevotella*, *Veillonella*, and *Selenomonas* in multiple independent sequencing efforts reduces the likelihood they are contaminants, we implemented a secondary approach to rule out contamination. We focused on *Fusobacterium* in *Fusobacterium*-high samples from the MUHAS Tanzania and ESCCAPE Kenya cohorts, reasoning that if *Fusobacterium* was a contaminant, it would be the same isolate (and thus genetically similar) in each sample within each sequencing effort. As a positive control, we included WGS data from two pairs of tumor and oral *Fusobacterium* isolates from colorectal cancers generated by Abed et al., who had found that these tumor and oral *Fusobacterium* pairs were genetically similar isolates (47). We found that while our approach groups the oral and tumor isolates as reported by Abed et al., *Fusobacterium* from patients in the MUHAS Tanzania (Figure S3F) and ESCCAPE Kenya (Figure S3G) patient groups were genetically distinct within each sequencing effort, indicating that the *Fusobacterium* is not a sequencing contaminant.

Evaluation of association between saliva and tumor microbiomes in ESCC patients from Tanzania

We next investigated the similarity between the saliva and tumor microbiomes of ESCC patients. Paired tumor-saliva samples were only available from patients in the MUHAS Tanzania cohort; these paired tumor-saliva specimens were analyzed to evaluate bacterial abundance as a proxy for the oral microbiome.

We first assessed the similarity between paired saliva and tumor microbiomes with the Bray Curtis similarity index (50). To avoid potential confounding due to low bacterial read counts in some tumor samples, we limited these analyses to the 21 tumor-saliva pairs that contain appreciable microbial sequencing depth (at least 10,000 bacterial reads each). We found that the saliva and tumor microbiomes from the same patient in the Tanzanian samples are significantly more similar than random saliva-tumor pairs ($p=0.0003$, Wilcoxon rank sum test) (Figure 3A). Next, we asked if there are bacterial genera whose relative abundance in the saliva correlates with their relative abundance in the tumor. For this analysis, we included only common-abundant bacterial genera with at least 1% relative abundance in at least three tumor-saliva pairs. The relative abundance of four bacterial genera (*Fusobacterium*, *Veillonella*, *Streptococcus*, and *Porphyromonas*) are strongly correlated between tumor and saliva microbiomes, while other common-abundant bacterial genera were not (Figure 3B). To assess if any bacterial genera are preferentially enriched in the tumor microbiome relative to the saliva microbiome, we next calculated the difference in the relative abundance of the common-abundant bacterial genera between tumor-saliva pairs. Several genera including *Porphyromonas* and *Veillonella* were at higher relative abundance in the saliva, while *Prevotella* and *Fusobacterium* were enriched in the tumor microbiome (Figure 3C). Finally, the relative abundance of tumor-associated bacteria including *Fusobacterium*, *Prevotella*, *Selenomonas*, *Veillonella*, *Streptococcus*, and *Campylobacter* are strikingly similar between the microbiomes of tumor and saliva pairs (Figure 3D).

DISCUSSION

This report provides a comprehensive analysis of bacterial communities present in ESCC tumors from nine countries from different regions of the world, which were previously sequenced in four independent studies. We found traditionally oral, cancer-associated, bacterial genera in tumors from patients in Tanzania, Malawi, Kenya, China, and Iran. These results provide evidence that these bacterial genera may be associated with ESCC in these high-incidence regions. We also identified similar bacterial genera in ESCC tumors from low-incidence regions, although this finding was based on a small sample size and only one sequencing cohort. Finally, in a sub-analysis of tumor and saliva pairs available from Tanzania, we demonstrated that the paired saliva and tumor microbiomes of ESCC patients were strikingly similar at the time of diagnosis; in particular, we identified a specific correlation between the saliva and tumor relative abundance of the bacterial genera *Fusobacterium*, *Veillonella*, *Streptococcus*, and *Porphyromonas*, with *Prevotella* and *Fusobacterium* significantly enriched in the tumor microbiome. Altogether, these data support the hypothesis that there is an association between the saliva and tumor microbiomes of ESCC patients in Tanzania. Additional studies are necessary to confirm this finding in additional patient groups and to clarify the significance.

Many of the bacterial genera identified in this study have been previously implicated in the carcinogenesis of gastrointestinal cancers. For example, studies have found that oral microbiota including *Fusobacterium*, *Prevotella*, *Selenomonas*, *Veillonella*, *Streptococcus*, and *Campylobacter* can be used to distinguish individuals with colorectal cancer from healthy controls (51), and that *Fusobacterium nucleatum* strains that colonize the oral cavity and tumors of patients with colorectal cancer are identical in some patients (52), raising the possibility that the oral cavity is a source of extra-oral cancer microbiota. Our group has previously shown that *Fusobacterium*, *Selenomonas*, and *Prevotella* can be visualized invasively within colorectal tumors and liver metastases (25). *Fusobacterium nucleatum* has been previously identified in esophageal cancers and is associated with shorter overall survival (53). Members of the genus *Porphyromonas* have been previously observed invasively within ESCC tumors (29) and have been reported to promote oral squamous cell carcinoma through a variety of mechanisms (30, 31). *Campylobacter jejuni* has been reported to promote tumorigenesis in mice (32), and *Streptococcus* species have been identified in human esophageal cancers (33). In addition, the striking association of *Streptococcus bovis* with colorectal cancer has led to the recommendation that colonoscopy be performed upon detection of *Streptococcus bovis* bacteremia or endocarditis (34, 35). Oral commensal bacteria such as *Veillonella* species have been previously implicated in pathogenesis of lung cancer (49). A prospective cohort of American patients (54) and a study of Japanese patients (55) likewise found that oral microbiome composition is associated with risk for development of esophageal cancer.

We found that bacterial genera including *Fusobacterium*, *Prevotella*, *Selenomonas*, *Veillonella*, *Streptococcus*, and *Campylobacter* are pervasive in the microbiome of ESCC tumors from patients in high-incidence regions. Moreover, the bacterial composition of ESCC tumors is remarkably similar across countries in those high-incidence regions, raising the possibility that particular bacterial genera may be involved in ESCC carcinogenesis

or that they may colonize tumors as a result of the common clinical presentation of patients with severe dysphagia. Notably, there are several alternative hypotheses that warrant mention. For example, it is possible that the ESCC-associated bacterial genera simply represent common members of the esophageal microbiome (56) and that the microbial populations we observed in these cancers are not significantly different from those found in normal esophagus tissue. A limitation of our study is a lack of normal esophageal tissue from ESCC cases or healthy controls in these settings, which would allow us to address this possibility. Another possible explanation is that ESCC tumors provide a favorable niche in which these bacteria are sequestered, in the setting of malignant obstruction, and allowed to colonize. Thus, it is plausible that ESCC-associated bacteria are not necessarily promoting ESCC carcinogenesis but rather represent passengers resulting from the sequestration of oral secretions proximal to an obstructing tumor. While the previous association of these bacterial genera with other cancers is consistent with the hypothesis that they influence carcinogenesis of ESCC, future studies are necessary to identify which, if any, direct influences these bacterial genera have upon ESCC carcinogenesis. Nevertheless, even if these bacterial genera do not have a role in increasing ESCC risk, but arise at the time of disease onset, they may have an important role to play as part of a biomarker for non-invasive early detection. Finally, a concern of all microbiome analyses is that observed bacteria can be a consequence of contamination at some step between tumor harvest and sequencing. However, our analysis of potential contaminants and the fact that we identify *Fusobacterium*, *Prevotella*, *Selenomonas*, *Veillonella*, *Streptococcus*, and *Campylobacter* in four independently collected sequencing efforts indicates that these finding are unlikely due to contamination.

While this study focused on the presence of bacteria with ESCC in high-incidence regions, we found evidence of similar cancer-associated bacteria in tumors in patients from low-incidence regions (USA, Ukraine, Vietnam, and Russia). A limitation of this assessment is the small sample size (n=36) and reliance on a single TCGA cohort that likely contains contaminants (48). Regardless, this finding does not exclude the possibility that the microbiome could be a factor driving patterns of ESCC incidence. For example, it is possible that the prevalence of ESCC-associated bacteria in people could vary across regions, which in turn could drive these differing rates of ESCC incidence. This is an important topic for future study.

We found that the structure of synchronous paired tumor and oral microbiomes were strikingly similar, although in a limited group of 21 patients. It is possible that this similarity is driven by transient contact of saliva and its associated microbiome with the tumor (e.g., during swallowing or tumor extraction). However, we found that only four of sixteen common-abundant bacterial genera correlate in abundance between the tumor and oral microbiomes, suggesting tumor-oral microbiome similarity is not driven exclusively by “in-trans” interactions between the saliva and tumor. We also found that genera including *Prevotella* and *Fusobacterium* are often specifically enriched in the tumor microbiome, supporting a model where specific oral bacterial preferentially colonize the tumor. A caveat of this study is that we infer oral bacterial populations from the saliva, despite diverse communities of bacteria throughout the oral cavity (57). However, we do observe *Fusobacterium* in the saliva despite its general association with periodontal

plaques (58), suggesting saliva is capable of detecting periodontal pathogens. Additionally, because the samples studied here are from patients with late-stage disease, it is possible that tumor-induced changes to upper-gastrointestinal physiology and dysphagia symptom-induced major dietary changes could themselves alter the oral microbiomes of these patients. The previous findings from the ESCCAPE studies in Kenya and Tanzania (17, 19) which found strong associations with dental staining (ORs > 10) and for which photographic validation studies suggest that most dental staining was not fluorosis, also point to a recent build-up of chromogenic bacteria. Studies of the oral microbiome of patients at earlier stages of ESCC and in prospective studies are necessary to address this possibility. We restricted our analysis to 21 tumor-saliva pairs that have a sufficient number of bacterial reads (at least 10,000). It is likely that excluded samples are not molecularly distinct from included samples but that the relatively low bacterial read counts in some tumors is simply reflective of low sequencing depth. Additional studies are necessary to understand the relationship between tumor and oral bacterial communities in additional patient groups in other high- and low- incidence regions.

Our observation of similar tumor and saliva microbiomes in ESCC patients is especially notable considering emerging evidence linking periodontal disease and poor oral health with increased risk of various cancers (17, 59, 60). This raises several important questions for future inquiry. It will be essential to determine if there is a difference in the oral prevalence of these identified cancer-associated bacteria between ESCC patients and non-patients earlier in the natural history of the disease, for example through comparisons of patients with esophageal squamous dysplasia and healthy controls. Because the prevalence of these bacteria may be associated with factors such as oral health, hygiene, and diet, studies of the impact of these factors on the oral microbiome in the general population would inform whether the oral microbiome is on a pathway linking oral hygiene to ESCC risk and may have a role in prevention.

In conclusion, we show that cancer-associated, traditionally-oral bacteria including the genera *Fusobacterium, Selenomonas, Prevotella, Streptococcus, Porphyromonas, Veillonella,* and *Campylobacter* are highly abundant within ESCC tumors from patients in regions with a high incidence of ESCC. We also show that there is a correlation between the genus composition of the saliva microbiome and the ESCC tumor microbiome of some ESCC patients. These findings will be foundational for future studies to understand if and how bacteria influence ESCC pathogenesis and to understand the role of the oral microbiome in this process. Finally, this study highlights the benefit of collaborative investigation to evaluate the international heterogeneity of this disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Liu et al. (37) for providing access to sequencing data from ESCC patients in Malawi (dbGaP accession phs001448.v1.p1). Portions of this research were conducted using the O2 High Performance Computing Cluster, supported by the Research Computing Group at Harvard Medical School. We thank Aleksandar Kostic for his

helpful discussions. We thank the teams from the MUHAS-ORCI-UCSF Cancer Collaboration, UNC-Malawi Project, Mutographs, ESCCAPE, and TCGA for specimen and data collection.

Funding

National Institutes of Health, National Cancer Institute Cancer Center Administrative Supplement to Promote Cancer Prevention and Control Research in Low and Middle Income Countries, A119617, [CA0082629] to K.V.L. S.B. received funding from NIH/NCI Grants: R00CA229984 and Cancer Center Support Grant P30 CA015704. EAC received funding from NIH/NCI Grants: U24CA210974, R01CA222862, R01CA227807, R01CA239604, R01CA230263. This work was supported in part by the US Public Health Service grants R35CA232128 and P01CA203655 to J.A.D. Content does not reflect the views of the National Cancer Institute or National Institutes of Health. This work was supported by a Cancer Grand Challenges OPTIMISTICC team award (C10674/A27140) to M.M. and a Cancer Grand Challenges Mutographs team award funded by Cancer Research UK [C98/A24032] to M.R.S. and P.B. R21CA191965 supported the collection of ESCCAPE Kenya tumors. The International Agency for Research on Cancer Section of Environment and Radiation supported the collection of ESCCAPE Tanzania tumors. This work was supported in part by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics.

Data Availability Statement

Raw sequencing data for all samples are available as indicated in the relevant manuscripts. The microorganism-mapped, non-human reads from the MUHAS Tanzania samples are available on the SRA under BioProject PRJNA810436.

All GATK-PathSeq output files and reproducible analysis and plotting R Notebooks are available.

Zenodo: https://doi.org/10.5281/zenodo.4750577

GitHub: https://github.com/jnoms/ESCC_microbiome

Furthermore, all analysis and figures can be automatically reproduced through a series of Google Colab documents.

Figure 1 and Supplementary Figure 1: https://github.com/jnoms/ESCC_microbiome/blob/main/collab/Figure1.ipynb

Figure 2 and Supplementary Figure 2: https://github.com/jnoms/ESCC_microbiome/blob/main/collab/Figure2.ipynb

Figure 3: https://github.com/jnoms/ESCC_microbiome/blob/main/collab/Figure3.ipynb

Other data that support the findings of this study are available from the corresponding author upon request

Abbreviations:

AFRECC	African Esophageal Cancer Consortium
COAD	Colon adenocarcinoma
ESCA	Esophageal adenocarcinoma
ESCC	Esophageal squamous cell carcinoma

ESCCAPE	Esophageal Squamous Cell Carcinoma African Prevention Research
MUHAS	Muhimbili University of Health and Allied Sciences
RNAseq	RNA sequencing
TCGA	The Cancer Genome Atlas
WGS	Whole genome sequencing

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians.n/a(n/a).
2. CC MRA, Dawsey S. Oesophageal cancer: A tale of two malignancies. World Cancer Report: Cancer Research for Cancer Prevention Lyon, France: International Agency for Research on Cancer Available from: http://publications.iarc.fr/586. 2020.
3. Coleman HG, Xie S-H, Lagergren J. The epidemiology of esophageal adenocarcinoma. Gastroenterology. 2018;154(2):390–405. [PubMed: 28780073]
4. Rustgi AK, El-Serag HB. Esophageal carcinoma. New England Journal of Medicine. 2014;371(26):2499–509. [PubMed: 25539106]
5. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. Gut. 2015;64(3):381–7. [PubMed: 25320104]
6. Cheng ML, Zhang L, Borok M, Chokunonga E, Dzamamala C, Korir A, et al. The incidence of oesophageal cancer in Eastern Africa: identification of a new geographic hot spot? Cancer epidemiology. 2015;39(2):143–9. [PubMed: 25662402]
7. Abnet CC, Kamangar F, Islami F, Nasrollahzadeh D, Brennan P, Aghcheli K, et al. Tooth loss and lack of regular oral hygiene are associated with higher risk of esophageal squamous cell carcinoma. Cancer Epidemiology and Prevention Biomarkers. 2008;17(11):3062–8.
8. Abnet CC, Qiao Y-L, Mark SD, Dong Z-W, Taylor PR, Dawsey SM. Prospective study of tooth loss and incident esophageal and gastric cancers in China. Cancer Causes & Control. 2001;12(9):847–54. [PubMed: 11714113]
9. Dar N, Islami F, Bhat G, Shah I, Makhdoomi M, Iqbal B, et al. Poor oral hygiene and risk of esophageal squamous cell carcinoma in Kashmir. British journal of cancer. 2013;109(5):1367–72. [PubMed: 23900216]
10. Chen X, Yuan Z, Lu M, Zhang Y, Jin L, Ye W. Poor oral health is associated with an increased risk of esophageal squamous cell carcinoma-a population-based case-control study in China. International journal of cancer. 2017;140(3):626–35. [PubMed: 27778330]
11. Sato F, Oze I, Kawakita D, Yamamoto N, Ito H, Hosono S, et al. Inverse association between toothbrushing and upper aerodigestive tract cancer risk in a Japanese population. Head & neck. 2011;33(11):1628–37. [PubMed: 21259377]
12. Liang H, Yang Z, Wang JB, Yu P, Fan JH, Qiao YL, et al. Association between oral leukoplakia and risk of upper gastrointestinal cancer death: a follow-up study of the Linxian general population trial. Thoracic cancer. 2017;8(6):642–8. [PubMed: 28929584]
13. Guha N, Boffetta P, Wünsch Filho V, Eluf Neto J, Shangina O, Zaridze D, et al. Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies. American journal of epidemiology. 2007;166(10):1159–73. [PubMed: 17761691]
14. Chen Q-L, Zeng X-T, Luo Z-X, Duan X-L, Qin J, Leng W-D. Tooth loss is associated with increased risk of esophageal cancer: evidence from a meta-analysis with dose-response analysis. Scientific reports. 2016;6(1):1–7. [PubMed: 28442746]

15. Sheikh M, Poustchi H, Pourshams A, Etemadi A, Islami F, Khoshnia M, et al. Individual and combined effects of environmental risk factors for esophageal cancer based on results from the Golestan Cohort Study. Gastroenterology. 2019;156(5):1416–27. [PubMed: 30611753]
16. Patel K, Wakhisi J, Mining S, Mwangi A, Patel R. Esophageal cancer, the topmost cancer at MTRH in the Rift Valley, Kenya, and its potential risk factors. International Scholarly Research Notices. 2013;2013.
17. Menya D, Maina SK, Kibosia C, Kigen N, Oduor M, Some F, et al. Dental fluorosis and oral health in the African Esophageal Cancer Corridor: Findings from the Kenya ESCCAPE case–control study and a pan-African perspective. International journal of cancer. 2019;145(1):99–109. [PubMed: 30582155]
18. Mmbaga EJ, Mushi BP, Deardorff K, Mgisha W, Akoko LO, Paciorek A, et al. A Case–Control Study to Evaluate Environmental and Lifestyle Risk Factors for Esophageal Cancer in Tanzania. Cancer Epidemiology and Prevention Biomarkers. 2020.
19. Mmbaga BT, Mwasamwaja A, Mushi G, Mremi A, Nyakunga G, Kiwelu I, et al. Missing and decayed teeth, oral hygiene and dental staining in relation to esophageal cancer risk: ESCCAPE case-control study in Kilimanjaro, Tanzania. International journal of cancer. 2020.
20. Buckle GC, Mmbaga EJ, Paciorek A, Akoko L, Deardorff K, Mgisha W, et al. Risk factors associated with early-onset esophageal cancer in Tanzania. JCO Global Oncology. 2022;8:e2100256. [PubMed: 35113655]
21. Ishaq S, Nunn L. Helicobacter pylori and gastric cancer: a state of the art review. Gastroenterology and hepatology from bed to bench. 2015;8(Suppl1):S6. [PubMed: 26171139]
22. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell host & microbe. 2013;14(2):207–15. [PubMed: 23954159]
23. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome research. 2012;22(2):292–8. [PubMed: 22009990]
24. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome research. 2012;22(2):299–306. [PubMed: 22009989]
25. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science. 2017;358(6369):1443–8. [PubMed: 29170280]
26. Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, et al. Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. Immunity. 2015;42(2):344–55. [PubMed: 25680274]
27. Rubinstein MR, Baik JE, Lagana SM, Han RP, Raab WJ, Sahoo D, et al. Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. EMBO reports. 2019;20(4):e47638. [PubMed: 30833345]
28. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell host & microbe. 2013;14(2):195–206. [PubMed: 23954158]
29. Gao S, Li S, Ma Z, Liang S, Shan T, Zhang M, et al. Presence of Porphyromonas gingivalis in esophagus and its association with the clinicopathological characteristics and survival in patients with esophageal cancer. Infectious agents and cancer. 2016;11(1):3. [PubMed: 26788120]
30. Whitmore SE, Lamont RJ. Oral bacteria and cancer. PLoS pathogens. 2014;10(3):e1003933. [PubMed: 24676390]
31. Inaba H, Sugita H, Kuboniwa M, Iwai S, Hamada M, Noda T, et al. P orphyromonas gingivalis promotes invasion of oral squamous cell carcinoma through induction of pro MMP 9 and its activation. Cellular microbiology. 2014;16(1):131–45. [PubMed: 23991831]
32. He Z, Gharaibeh RZ, Newsome RC, Pope JL, Dougherty MW, Tomkovich S, et al. Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. Gut. 2019;68(2):289–300. [PubMed: 30377189]

33. Narikiyo M, Tanabe C, Yamada Y, Igaki H, Tachimori Y, Kato H, et al. Frequent and preferential infection of Treponema denticola, Streptococcus mitis, and Streptococcus anginosus in esophageal cancers. Cancer science. 2004;95(7):569–74. [PubMed: 15245592]
34. Boleij A, Schaeps RM, Tjalsma H. Association between Streptococcus bovis and colon cancer. Journal of clinical microbiology. 2009;47(2):516-. [PubMed: 19189926]
35. Ferrari A, Botrugno I, Bombelli E, Dominioni T, Cavazzi E, Dionigi P. Colonoscopy is mandatory after Streptococcus bovis endocarditis: a lesson still not learned. Case report. World journal of surgical oncology. 2008;6(1):49. [PubMed: 18474093]
36. Van Loon K, et al. A Genomic Analysis of Esophageal Squamous Cell Carcinoma in Eastern Africa. (In Progress).
37. Liu W, Snell JM, Jeck WR, Hoadley KA, Wilkerson MD, Parker JS, et al. Subtyping sub-Saharan esophageal squamous cell carcinoma by comprehensive molecular analysis. JCI insight. 2016;1(16).
38. Moody S, Senkin S, Islam SMA, Wang J, Nasrollahzadeh D, Penha RCC, et al. Mutational signatures in esophageal squamous cell carcinoma from eight countries of varying incidence. medRxiv. 2021:2021.04.29.21255920.
39. Network CGAR. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017;541(7636):169. [PubMed: 28052061]
40. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330. [PubMed: 22810696]
41. Walker MA, Pedamallu CS, Ojesina AI, Bullman S, Sharpe T, Whelan CW, et al. GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. Bioinformatics. 2018;34(24):4287–9. [PubMed: 29982281]
42. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, et al. The vegan package. Community ecology package. 2007;10:631–7.
43. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020;579(7800):567–74. [PubMed: 32214244]
44. 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 biology. 2014;12(1):1–12. [PubMed: 24417977]
45. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6(1):1–14. [PubMed: 29291746]
46. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one. 2010;5(3).
47. Abed J, Maalouf N, Manson AL, Earl AM, Parhi L, Emgård JE, et al. Colon cancer-associated Fusobacterium nucleatum may originate from the oral cavity and reach colon tumors via the circulatory system. Frontiers in cellular and infection microbiology. 2020;10:400. [PubMed: 32850497]
48. Dohlman AB, Arguijo Mendoza D, Ding S, Gao M, Dressman H, Iliev ID, et al. The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. Cell Host & Microbe 2020.
49. Tsay J-CJ, Wu BG, Sulaiman I, Gershner K, Schluger R, Li Y, et al. Lower airway dysbiosis affects lung cancer progression. Cancer Discovery. 2020.
50. Ricotta C, Podani J. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. Ecological Complexity. 2017;31:201–5.
51. Flemer B, Warren RD, Barrett MP, Cisek K, Das A, Jeffery IB, et al. The oral microbiota in colorectal cancer is distinctive and predictive. Gut. 2018;67(8):1454–63. [PubMed: 28988196]
52. Komiya Y, Shimomura Y, Higurashi T, Sugi Y, Arimoto J, Umezawa S, et al. Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity. Gut. 2019;68(7):1335–7. [PubMed: 29934439]

53. Yamamura K, Baba Y, Nakagawa S, Mima K, Miyake K, Nakamura K, et al. Human microbiome Fusobacterium nucleatum in esophageal cancer tissue is associated with prognosis. Clinical Cancer Research. 2016;22(22):5574–81. [PubMed: 27769987]

54. Peters BA, Wu J, Pei Z, Yang L, Purdue MP, Freedman ND, et al. Oral microbiome composition reflects prospective risk for esophageal cancers. Cancer research. 2017;77(23):6777–87. [PubMed: 29196415]

55. Kawasaki M, Ikeda Y, Ikeda E, Takahashi M, Tanaka D, Nakajima Y, et al. Oral infectious bacteria in dental plaque and saliva as risk factors in patients with esophageal cancer. Cancer. 2021;127(4):512–9. [PubMed: 33156979]

56. Corning B, Copland AP, Frye JW. The esophageal microbiome in health and disease. Current gastroenterology reports. 2018;20(8):1–7. [PubMed: 29350301]

57. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu W-H, et al. The human oral microbiome. Journal of bacteriology. 2010;192(19):5002–17. [PubMed: 20656903]

58. Signat B, Roques C, Poulet P, Duffaut D. Role of Fusobacterium nucleatum in periodontal health and disease. Curr Issues Mol Biol 2011;13(2):25–36. [PubMed: 21220789]

59. Michaud DS, Lu J, Peacock-Villada AY, Barber JR, Joshu CE, Prizment AE, et al. Periodontal disease assessed using clinical dental measurements and cancer risk in the ARIC study. JNCI: Journal of the National Cancer Institute. 2018;110(8):843–54. [PubMed: 29342298]

60. Ahrens W, Pohlabeln H, Foraita R, Nelis M, Lagiou P, Lagiou A, et al. Oral health, dental care and mouthwash associated with upper aerodigestive tract cancer risk in Europe: the ARCAGE study. Oral oncology. 2014;50(6):616–25. [PubMed: 24680035]

Novelty and Impact:

Esophageal cancer is the sixth most common cause of cancer-related death worldwide. There are unexplained geographic patterns of esophageal squamous cell carcinoma (ESCC), with disproportionately higher incidence rates occurring in the eastern corridor of Africa and parts of Asia. We found that bacteria associated with other gastrointestinal cancers are associated with ESCC from patients in nine countries worldwide, highlighting a pressing need for future studies to understand the role of the microbiome in ESCC pathogenesis.

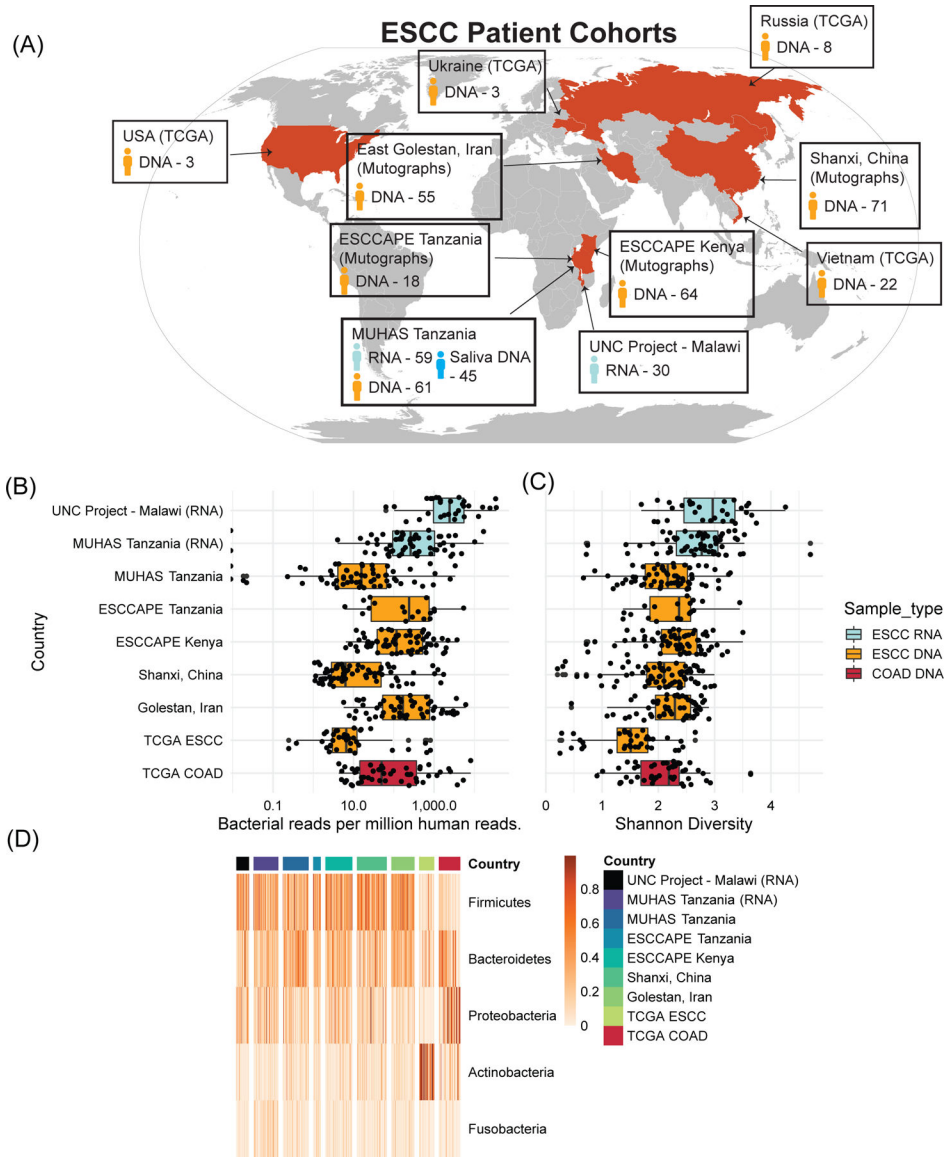


Figure 1. Microbiome structure and composition of ESCC tumors

A. Description of ESCC patients, and sample types, assessed in this study. TCGA – The Cancer Genome Atlas; ESCCAPE – Esophageal Squamous Cell Carcinoma African Prevention Research; Mutographs – Cancer Research UK Mutographs Project.

B. Bacterial burden of ESCC tumors for each patient cohort. Units are bacterial reads per million human reads as determined by GATK-PathSeq analysis. Each dot represents one sample. Analyte type (RNA or DNA) and tumor type (ESCC or COAD) are indicated by color.

C. Shannon diversity of ESCC tumors for each patient cohort. Shannon diversity was determined for each sample at the genus level based on genera that are at least 1% relative abundance. Each dot represents one sample. Analyte type (RNA or DNA) and tumor type (ESCC or COAD) are indicated by color.

D. Heatmap describing the relative abundance of the five top phyla sorted by average phylum relative abundance. Each column represents one sample. Rows represent the indicated phyla. Units are relative abundance. Samples from each cohort are WGS unless noted with “(RNA)”, in which case they are RNAseq.

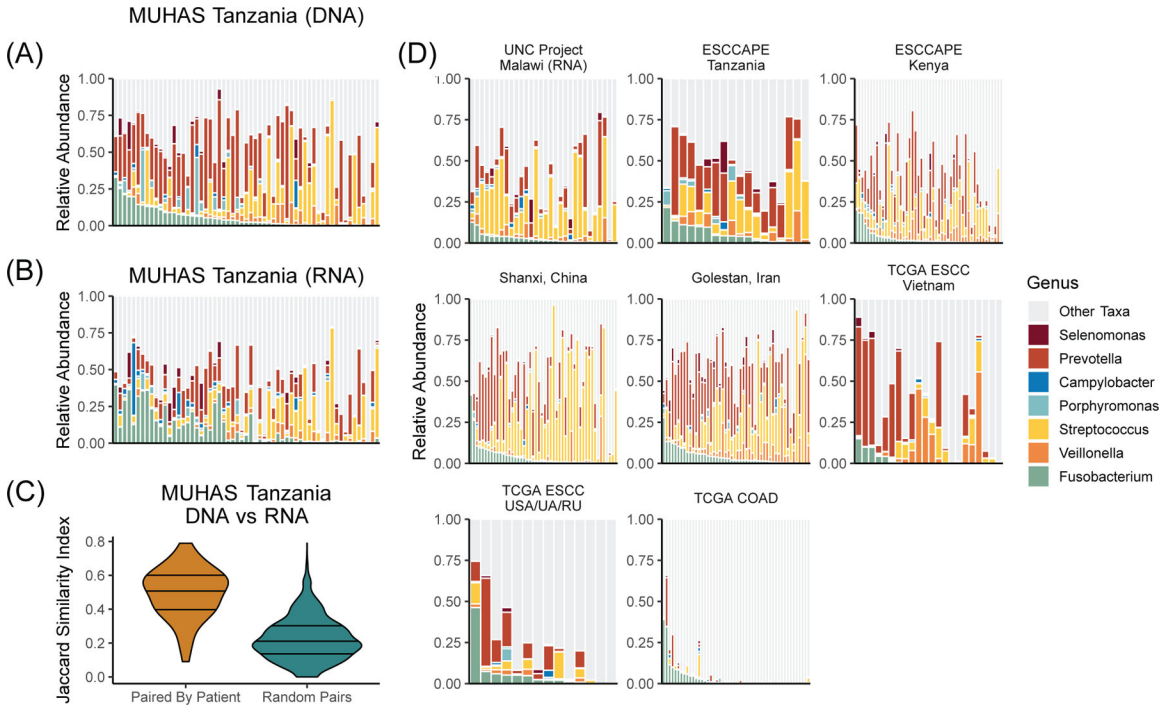


Figure 2. Identification of bacterial genera associated with carcinogenesis

A. Bacterial genera relative abundance of WGS data from the MUHAS Tanzania cohort. Each column represents a single sample. Samples are ordered by decreasing Fusobacterium relative abundance. Units are relative abundance of bacterial genus-mapping reads. Color indicates the genus, and seven genera are specified. Only patients with GATK-PathSeq analysis from both RNAseq and WGS tumor data are plotted (n=59).

B. Bacterial genera relative abundance of RNAseq data from the MUHAS Tanzania cohort. Each column represents a single sample. Here, column order is dictated according to the patient order in Figure 2A. Units are relative abundance of bacterial genus-mapping reads. Color indicates the genus, and seven genera are specified. Only patients with GATK-PathSeq analysis from both RNAseq and WGS tumor data are plotted (n=59). Samples are ordered in the same order as Figure 2A, which is by *Fusobacterium* genus relative abundance in the WGS data.

C. Jaccard index between RNAseq and WGS data of tumors from the MUHAS Tanzania cohort. For the “Paired by Sample” column, Jaccard indices were calculated only between the WGS and RNAseq data from the same tumor (n=59 comparisons). For the “Random Pairs” column, Jaccard indices were calculated between all possible WGS-RNAseq pairs independent of patient of origin to represent the expected random distribution of Jaccard indices (n=3,481 comparisons). Jaccard index was calculated from relative abundance at the genus level based on genera that are at least 1% relative abundance. The width of the violin represents the relative proportion of comparisons with each Jaccard index, and lines indicate 25th, 50th, and 75th percentiles.

D. Bacterial genera relative abundance of the remaining patient cohorts, including RNAseq and WGS data as indicated. Each column represents a single sample. Samples are ordered by decreasing Fusobacterium relative abundance within each patient cohort. Units are relative abundance of bacterial genus-mapping reads. Color indicates the genus, and seven

genera are specified. Here, if there were more than 50 samples in a patient cohort, 50 samples were randomly selected for visualization. USA – United States, UA – Ukraine, RU – Russia. All cohorts consist of WGS data, with the exception of the tumors from Malawi which are RNAseq. (Number of samples plotted: UNC Project - Malawi 30; ESCCAPE Tanzania 18; ESCCAPE Kenya 50; Shanxi, China 50; Golestan, Iran 50; TCGA ESCC Vietnam 22; TCGA ESCC USA/UA/RU 14).

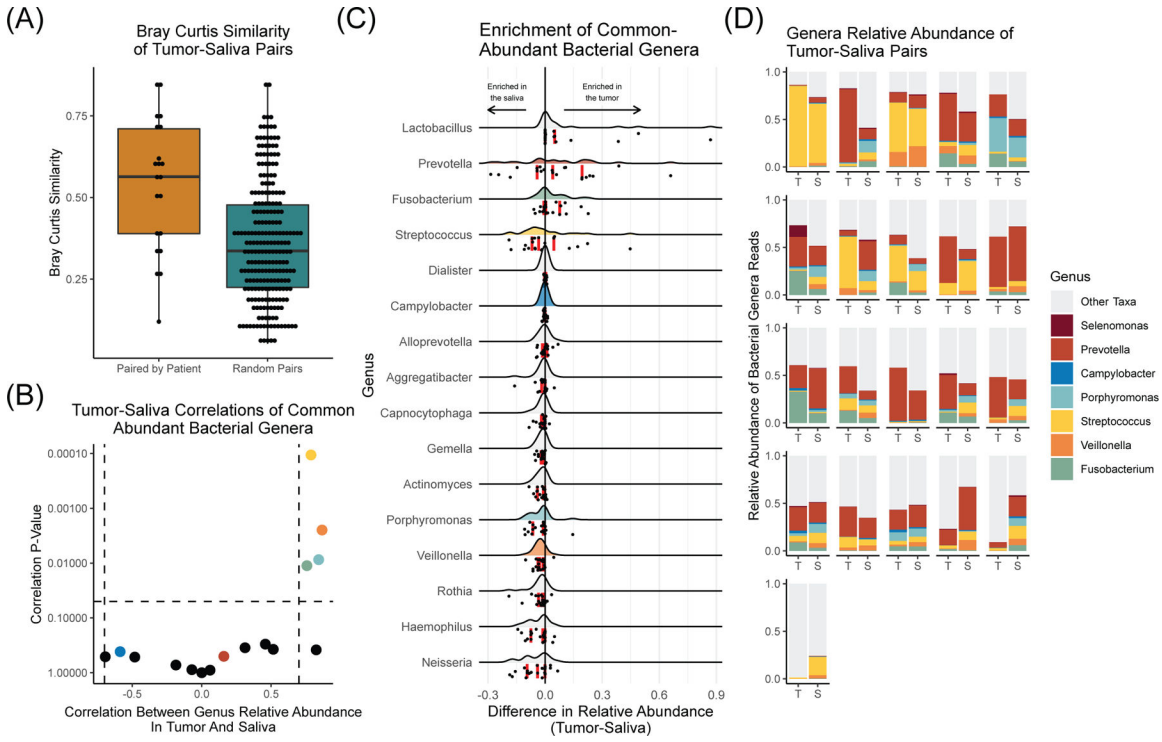


Figure 3. Association between synchronous saliva and tumor microbiomes in Tanzanian ESCC patients

A. Bray Curtis Similarity comparing tumor-saliva pairs from patients in the MUHAS Tanzania cohort. Analysis was restricted to the 21 tumor-saliva pairs that contained at least 10,000 bacterial reads. This analysis was conducted at the genus level and using relative abundance. For the “Paired by Patient” column, Bray Curtis Similarity was calculated only between the tumor and saliva WGS data from the same patient. For the “Random Pairs” column, Bray Curtis Similarity was calculated between all possible tumor-saliva pairs independent of patient of origin to represent the expected random distribution of Bray Curtis Similarity. (p=0.0003, Wilcoxon rank sum test).

B. Correlation between the relative abundance of common-abundant bacterial genera in paired saliva and tumor WGS data. Analysis was restricted to the 21 tumor-saliva pairs that contained at least 10,000 bacterial reads. Common-abundant bacterial genera are bacterial genera that are at least 1% abundance in at least 3 tumor-saliva pairs – 16 bacterial genera made this cutoff. Correlation represents a two-sided Pearson correlation. X-axis is the correlation coefficient, and Y axis is the correlation P-Value plotted on a log scale.

C. Enrichment of genera in the oral or tumor microbiome. Each row details one of the 16 common-abundant bacterial genera. Each row contains one data point per patient, for a total of 21 data points. The value of each point represents the difference in the relative abundance of the specified genus in the tumor and saliva microbiomes of one patient, with positive values indicating a genus is at higher relative abundance in a patient’s tumor. For example, if a genus is at a relative abundance of 0.7 (70%) in the tumor and 0.3 (30%) in the saliva of a patient, the plotted value for that genus and that patient is 0.4. Curves represent the distribution of this relative abundance difference across the tumor-saliva pairs, with dots indicating individual tumor-oral pairs. Vertical red lines indicate quartiles.

D. Relative abundance bar charts of tumor-saliva pairs. Analysis was restricted to the 21 tumor-saliva pairs that contained at least 10,000 bacterial reads. Units are relative abundance of bacterial genus-mapping reads. Color indicates the genus, and seven genera are specified. (abbreviations: T – tumor, S – saliva).

Table 1:

Sequencing effort characteristics

Sequencing Cohort	Number of patients	Reference	Sequencing Type
MUHAS Tanzania	61	(36)	DNA: WGS RNA: Human rRNA subtraction, random-hexamer priming
UNC Project - Malawi	30	(37)	RNA: Human rRNA subtraction, random-hexamer priming
Mutographs	210	(17, 19, 38)	DNA: WGS
TCGA	36	(39)	DNA: WGS

Int J Cancer. Author manuscript; available in PMC 2024 May 17.

Table 2.

Demographics and information on patient populations from all ESCC sequencing efforts

Study		**Tanzania**	**Malawi****	**ESCCAPE Tanzania*****	**ESCCAPE Kenya*****	**East Golestan, Iran******	**Shanxi, China******
No. cases included		61	30	18	65	55	71
Demographics							
Median age (IQR)		49 (44–62)	56	65 (61–73)	64 (53, 71)	62 (54,73)	56 (50, 64)
% male		67%	45.8%	61%	68%	55%	56%
Status at diagnosis							
Weight (kg), median (IQR)				44 (40–52)	52 (46, 60)		
Body mass index (kg/m^2) median (IQR)				15.8 (15.4, 19.1)	19.5 (15.6, 22.0)		
Median months ill before coming to endoscopy (IQR)				2 (1, 6)	3 (2, 4.5)		
HIV status	Positive Negative Not known	2 (3.2%) 36 (59.0%) 23 (37.7%)	10 (16.9%) 44 (74.6%) 5 (8.5%)	1 (5%) 10 (56%) 7 (39%)	5 (8%) 48 (74%) 12 (18%)		
Key lifestyle habits							
N (%) ever tobacco users				11 (61%)	38 (58%)	17 (31%)	35 (49%)
N (%) who brush teeth daily: With toothbrush With stick				12 (67%) 6 (33%)	16 (25%) * 10 (15%)		
Median no missing teeth (IQR)				3 (1, 5)	4 (1, 8)		

*N=22 (34%) brush once per week or never, n=17 (26%) brush 2 to 6 times/week

**Indicates demographics are from the entire patient population, consisting of both included and unincluded patients.

***Indicates demographic percentages are from the entire patient population, with discrete counts scaled to the number of cases included.

****Indicates demographic information is exclusively for included patients.