



OPEN ACCESS

EDITED BY

Paris Salazar-Hamm,
University of New Mexico, United States

REVIEWED BY

Alexander Link,
Otto von Guericke University Magdeburg,
Germany
Donald O. Natvig,
University of New Mexico, United States

*CORRESPONDENCE

Lary A. Robinson
✉ lary.robinson@moffitt.org

†PRESENT ADDRESS

Christine M. Pierce,
Epidemiology, MRL, Merck & Co., Inc.,
Kenilworth, NJ, United States

RECEIVED 30 June 2024

ACCEPTED 31 October 2024

PUBLISHED 21 November 2024


CITATION

Bailey A, Hogue S, Pierce CM, Paul S,
LaFuente N, Thapa R, Kim Y and Robinson LA
(2024) Metagenomic characterization of the
tracheobronchial microbiome in lung cancer.
Front. Microbiomes 3:1457537.
doi: 10.3389/frmbi.2024.1457537

COPYRIGHT

© 2024 Bailey, Hogue, Pierce, Paul, La Fuente,
Thapa, Kim and Robinson. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Metagenomic characterization of the tracheobronchial microbiome in lung cancer

Alexis Bailey¹, Stephanie Hogue², Christine M. Pierce^{2†},
Shirlene Paul², Natalie La Fuente², Ram Thapa³,
Youngchul Kim³ and Lary A. Robinson ^{1*}

¹Division of Thoracic Oncology, Moffitt Cancer Center, Tampa, FL, United States, ²Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, United States, ³Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, United States

Background: The tracheobronchial and oral microbiome may be associated with lung cancer, potentially acting as predictive biomarkers. Therefore, we studied the lung and oral bacteriome and virome in non-small cell lung cancer (NSCLC) patients compared to melanoma controls to discover distinguishable features that may suggest lung cancer microbial biomarkers.

Methods: In this pilot case-control study, we recruited ten patients with early-stage NSCLC (cases) and ten age-matched melanoma patients (controls) who both underwent tumor resection. Preoperative oral gargles were collected from both groups, who then underwent transbronchoscopic tracheal lavage after intubation. Lung tumor and adjacent non-neoplastic lung were sterilely collected after resection. Microbial DNA from all lung specimens underwent 16S rRNA gene sequencing. Lavage and gargle specimens underwent whole-genome shotgun sequencing. Microbiome metrics were calculated to compare both cohorts. T-tests and Wilcoxon rank sum tests were used to test for significant differences in alpha diversity between cohorts. PERMANOVA was used to compare beta diversity.

Results: No clear differences were found in the microbial community structure of case and control gargles, but beta diversity of case and control lavages significantly differed. Two species, *Granulicatella adiacens* and *Neisseria subflava*, which are both common oral commensal organisms, appeared in much higher abundance in case versus control lavages. Case lavages also maintained higher relative abundances of other oral commensals compared to controls.

Conclusions: Lung lavages demonstrated oral microbiota enrichment in cases compared to controls, suggesting microaspiration and resultant inflammation. The oral commensals *Granulicatella adiacens* and *Neisseria subflava* were more abundant in the tracheobronchial lavages of lung cancer versus melanoma patients, implicating these microorganisms as potential lung cancer biomarkers, warranting further validation studies.

KEYWORDS

lung cancer microbiome, non-small cell lung cancer, tracheal microbiota, oral microbiota, lung microbiome, microaspiration, lung inflammation

1 Introduction

Lung cancer is the most common cancer worldwide and the leading cause of cancer deaths, with 1.8 million deaths in 2020 (World Health Organization, 2022). In the US, lung cancer has the second highest incidence rate among both males and females, but it is the most common cause of death among both sexes (Henley et al., 2020). An estimated 130,450 Americans will die from lung cancer in 2024, exceeding the number of deaths expected from colon, breast, and prostate cancers combined (American Cancer Society, 2024). Despite enormous research and treatment efforts, the high fatality rate of this malignancy (75%) has changed little over the last few decades (American Lung Association, 2024). The high mortality rate of lung cancer is primarily due to delayed diagnosis, with 77% of cases not being recognized until later stages (Mao et al., 2016). Screening low dose chest computed tomography is underutilized, with only 5.5% of eligible individuals obtaining a scan despite its proven potential to detect early-stage diseases in high-risk individuals (Richards et al., 2019). Therefore, exploration of potential biomarkers of this disease is warranted.

As the affordability of next-generation sequencing techniques improves, the microbiome, or the collective genomic material of all microorganisms found within and on the body, is increasingly being investigated for associations with disease and potential therapeutic value. Most research has focused on the gut microbiome, the largest and most diverse microbiome in the human body, with relatively little investigation of the microbiota of other anatomic sites. Until recently, the lungs were considered sterile, but evidence indicates commensal microbes, including *Acinetobacter*, *Pseudomonas*, and *Ralstonia*, indeed colonize this organ (Yu et al., 2016). Furthermore, the composition and function of the microbiota in lung tissue are distinct from other anatomic sites, including the oral cavity (Yu et al., 2016).

Recent research has further shown associations between the local lung microbiome and various lung pathologies, such as asthma, cystic fibrosis, and chronic obstructive pulmonary disease (COPD) (Moffatt and Cookson, 2017; Mur et al., 2018). Additionally, hypotheses regarding an association between the lung microbiome and lung cancer, potentially mediated by chronic inflammation, have been suggested (Mur et al., 2018). Nevertheless, relatively little research on the lung microbiome in the context of lung cancer has been conducted.

Given the potential for the lung microbiome to be associated with lung cancer and to be utilized as a biomarker, this study aimed to characterize lung and oral bacteriomes and viromes in early-stage non-small cell lung cancer (NSCLC) patients compared to melanoma controls to identify distinguishable features in their oral, tracheal, and tumor microbiomes that may suggest a reliable

microbial biomarker for the presence of lung cancer. The results may help assess the potential of minimally invasive samples from the oral flora that might act as proxies for tumor microbiomes, potentially leading to the development of a reliable screening technique for high-risk individuals with early cancers. While the presence of specific microorganisms found in conjunction with lung cancer is intriguing, this study is not focused on correlating these findings with possible lung carcinogenesis.

2 Methods

2.1 Patients

This prospective, exploratory case-control study recruited ten early-stage NSCLC patients and ten control melanoma patients undergoing surgical resection of their tumor under general anesthesia at Moffitt Cancer Center between July 2015 and May 2016. Melanoma patients were chosen as the best possible proxy for controls because this cancer type has not shown clear evidence of microbial etiology (Woo et al., 2022) that would alter the normal oral and respiratory flora, and these patients were already undergoing anesthesia with intubation for major resection of locally-advanced peripheral melanomas on extremities, far distant from the lung. Theoretically, the best control would be bronchoscopy and tracheobronchial lavages on normal people without cancer or lung disease. However, it would be exceedingly difficult to recruit these individuals and we had ethical concerns about recommending this procedure with its inherent risks for this exploratory study. In general, the usual reasons for performing bronchoscopy on patients without diagnosed cancer are either for evaluation of lung infections and/or to biopsy lung abnormalities which might be cancer. Obviously, these patients would not be valid controls.

Lung cancer cases and melanoma controls were matched by age (\pm 10 years) and smoking status (current/former versus never-smokers). Eligible participants were at least 21 years of age, mentally competent, not pregnant, and received no chemotherapy within 1 year of surgery. Furthermore, participants could not have post-obstructive pneumonitis, current pneumonitis, purulent bronchitis, other acute respiratory infections, cystic fibrosis, clinically significant bronchiectasis, other inflammatory or fibrotic lung diseases, chronic or current corticosteroid use, antimicrobial therapy within 1 month or prebiotics/probiotics within 3 months of surgery. We performed this study per the ethical standards established in the 1964 Declaration of Helsinki and its later amendments. It was approved by the Liberty Institutional Review Board, Protocol 14.12.0036 (MCC 17976). We obtained informed consent from all participants.

2.2 Specimen collection

2.2.1 Tracheobronchial lavages

We collected intraoperative tracheobronchial lavages in all patients. After induction of general anesthesia and within two minutes of endotracheal intubation with a sterile single-use tube, LAR performed bronchoscopy with tracheal lavage using 50-100mL of

Abbreviations: COPD, Chronic obstructive pulmonary disease; DNA, Deoxyribonucleic acid; HPV, Human papillomavirus; HTLV-1, Human T-cell leukemia virus-1; ICI, Immune checkpoint inhibitors; NSCLC, Non-small cell lung cancer; PBS, Phosphate-buffered saline; PERMANOVA, Permutational multivariate analysis of variance; PCR, Polymerase chain reaction; 16S rRNA, 16S ribosomal ribonucleic acid; TBL, Tracheobronchial lavage; WGSS, Whole genome shotgun sequencing.

sterile 0.9% normal saline solution and an Olympus pediatric bronchoscope pre-cleaned and disinfected with Steris (Steris System 1E Liquid Chemical Sterilant Processing System, Steris Corporation, Mentor, OH) according to CDC guidelines (Rutala and Weber, 2008). An intravenous, preoperative, prophylactic antibiotic was started during bronchoscopy, so it would not have reached a therapeutic blood level when we obtained lavage samples. We collected approximately 20 mL of tracheobronchial lavage fluid into a sterile Lukens trap (Argyle™ Specimen Trap, Cardinal Health Inc., Dublin, OH), transported on ice to the laboratory, and processed by centrifugation at 3,000 x g for 15 minutes at 4°C to separate supernatant and cell pellet. 3.2mL of supernatant was pipetted between two cryovials. Cell pellets were re-suspended in 1.2mL of sterile PBS and aliquoted evenly between two cryovials, which were snap-frozen in liquid nitrogen. Cell pellets were snap frozen in liquid nitrogen (LN) and stored at -80°C.

2.2.2 Oral gargle samples

We also collected oral gargles from cases and controls in the preoperative area. Participants vigorously swished and gargled 15mL of disinfectant-free mouthwash for 15 seconds, then expectorated into a sterile 50mL conical tube. Specimens were centrifuged according to the same parameters as lavages. We collected 3.2mL of supernatant between two cryovials. The cell pellet was re-suspended in 20mL of PBS and centrifuged again at the same speed, duration, and temperature. The final cell pellet was re-suspended in 1.2mL PBS and aliquoted as two 0.6mL aliquots stored at -80°C.

2.2.3 Tissue samples

Only lung cancer patients provided tumor and adjacent non-neoplastic lung tissue specimens. Immediately after resection, the surgeon (LAR, the thoracic surgeon on all cases), while wearing a mask, took the resected specimen in a sterile container to the frozen section room and after the pathologist removed specimens needed for clinical pathology (diagnosis, margins, and lymph nodes), LAR removed 1cm³ from the tumor using sterile gloves and instruments in a sterile field. LAR also harvested a similar-sized, non-neoplastic lung specimen in the same manner at a distance from the tumor. Tissue specimens were transported to the laboratory and snap frozen in LN before undergoing macrodissection and long-term storage at -80°C.

2.3 DNA extraction

We extracted microbial DNA from all sample types. The MoBio® PowerSoil DNA isolation kit (Qiagen, Germantown, MD) was utilized in a modified protocol to extract bacterial DNA from 0.6mL cell pellets from lavages and gargles. Briefly, cell pellets were vortexed and spun down until the sample collected at the bottom of the tube. It was then added to a bead-beating tube with buffer and processed in the MP-Bio Fastprep™ 5G (MP Biomedicals, Irvine, CA) for 30 seconds at 6m/s for each of 2 cycles. Samples were centrifuged at 10,000 x g for 30 seconds at room temperature with the resulting supernatant collected. The supernatant was processed to remove PCR inhibitors and eluted with 100µL of buffer. DNA was quantitated using Qubit, and quality was checked using Nanodrop.

We used the Qiagen® DNeasy Blood and Tissue kit (Qiagen, Germantown, MD) to isolate DNA from tissue samples according to the manufacturer's protocol. Approximately 25mg was utilized, or about half of the total tissue volume. We briefly added the tissue to a bead-beating tube containing 360 µL of ATL buffer and 40 µL of proteinase K before being vortexed and incubated in the lytic step. Samples were bead-beat according to the same steps outlined above. Samples were then centrifuged at 20,000xg for 3 minutes, and the resulting supernatant was further processed and eluted in buffer AE.

2.4 16S rRNA gene sequencing

All samples underwent 16S rRNA gene sequencing with appropriate controls. Libraries were prepared using standard operating procedures (SOPs) from the Weinstock Lab at the Jackson Laboratory (The Jackson Laboratory, Farmington, CT). Briefly, high-performance liquid chromatography-purified primers and 4ng of DNA template were used to amplify the V1-V3 regions of the 16S rRNA gene. Libraries were screened for size and quantity as described in the SOP, and after pooling, they were quantified by qPCR using the Kapa Library Quantification Kit. The final libraries were sequenced with a 50% PhiX spike-in on an Illumina MiSeq v3 2x300 sequencing run. The raw reads were submitted to the NCBI Short Read Archive (SRA) under BioProject PRJNA1177881.

2.5 Metagenomic whole genome shotgun sequencing

DNA was isolated from all oral gargles and lung lavages. Whole genome shotgun DNA libraries were prepared from 100ng of DNA using the Illumina TruSeq Nano DNA kit following the manufacturer's protocol (Illumina, Inc., San Diego, CA). The libraries were sequenced on Illumina NextSeq High Output Kits v2 2x150 to about 80 to 260 million paired-end reads, depending on the percent alignment to microbial species. This method was utilized to resolve bacterial signatures to the species level and to identify viral signatures. Fungal sequences were not examined in the WGSS results.

2.6 Bioinformatics and statistical analyses

2.6.1 16s rRNA sequencing data analysis

Paired-end sequencing reads were cleaned using Trimmomatic v. 0.39 (Bolger et al., 2014) with the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 to remove adaptors and low-quality reads. Treatment samples with a minimum of 2,000 reads were kept for further downstream analysis. The chimeric reads were searched against the 16S rRNA Gold database with the default UCHIME (4.2) parameters (Edgar, 2016). Next, the cleaned reads were merged with PEAR (0.9.10) (Zhang et al., 2014) and operational taxonomic units (OTUs) were generated by open reference of QIIME1.9.1 pipeline (Caporaso et al., 2010). Only OTUs with a minimum observation count of 100 were retained. The database used for the taxonomic assignment was

Silva 128_97_otus_16S.fasta (Quast et al., 2013). Alpha- and Beta-diversity were analyzed using QIIME1.9.1. The taxonomy plots were based on the 25 most prevalent OTUs. PERMANOVA was used to compare beta diversity estimates.

Fold differences in the top 25 most abundant microbes, with relative abundances of at least 1% in one comparison group, were calculated by dividing the relative abundance of the microbe in the comparison groups. Similarly, fold differences in the top 25 most prevalent microbes with prevalence of at least 10% in either comparison group were calculated and organized into Venn diagrams. Student two-sample t-test and Wilcoxon rank sum test were used for differential abundance analysis between cases, controls, and sample types. Two-sided P values <0.05 were considered statistically significant. Statistical analysis was completed using the Phyloseq package in R software (v3.1.1 and v4.1.0, The R Foundation, Vienna, Austria).

2.6.2 WGSS data processing analysis for taxonomic classification methods

The CosmosID platform was used to process WGSS data and perform strain-level taxonomic classification. Briefly, their algorithm disambiguated short sequence reads into discrete genomes. The pipeline used pre-computation phases [using the CosmosID taxonomic reference databases containing bacteria, viruses, phages, virulence markers, and antimicrobial resistance markers curated by CosmosID (CosmosID, Inc., Germantown, MD) (CosmosID Inc, 2021)] with per-sample computation (searches short sequence reads or contigs from draft *de novo* assemblies against fingerprint sets), detect and classify microbial sequencing reads. The platform filtered reads using a filtering threshold derived from internal scores determined by analyzing many diverse metagenomes to exclude false positives.

3 Results

3.1 Patient characteristics

All 20 participants (ten NSCLC cases and ten melanoma controls) were Caucasian (Table 1). Cases had a higher percentage of females than controls (40% vs. 20%). Most lung cancer patients had stage I disease (80%), while most melanoma controls were advanced stage (50% had stage III disease). Most cases (90%) and controls (80%) had not received antibiotics within 2 months before their surgery. No significant differences were observed in the characteristics measured between cases and controls.

3.2 Microbial profiling

3.2.1 Cases vs. controls: tracheobronchial lavages

3.2.1.1 16S rRNA gene sequencing

The usual lower airway genera *Streptococcus* and *Prevotella* (CosmosID Inc, 2021) were identified in all lavages (Figure 1A). However, the oral commensals appeared more prevalent among cases versus controls: *Granulicatella* (100% versus 30%), *Leptotrichia* (100% versus 50%), *Moryella* (70% versus 20%), and *Neisseria* (80% versus

50%) respectively (Figure 2A). *Neisseria* was nearly eight-fold more abundant in tracheobronchial lavages of cases versus controls (Table 2; Figure 2B) (Lee et al., 2016; Huffnagle et al., 2017).

Lung cancer lavages appeared slightly more diverse than controls, but not significantly (Table 3; Figure 2C) (Lee et al., 2016; Gomes et al., 2019). Beta diversity, measured by Bray-Curtis dissimilarity, showed substantial differences but no clear separation between cases and controls (Table 3; Figure 2D).

3.2.1.2 Whole genome shotgun sequencing

The bacterial species *Granulicatella adiacens* and *Neisseria subflava* were more abundant in cases compared to controls by 6.18 and 15.93-fold, respectively (Table 2; Supplementary Figures S1A, B). Several species of *Prevotella* were more abundant in controls compared to cases. Alpha diversity estimates revealed no consistent pattern and, along with beta diversity, were not significantly different between cases and controls (Table 3; Supplementary Figures S1C, D).

The virome of the tracheal lavages was assessed through WGSS sequencing, too, identifying broadly similar prevalence and relative abundance between cases and controls (Table 2). Human betaherpesvirus 7 was far more abundant in case versus control lavages but was rare. Though not statistically significant, case lavages consistently showed higher viral alpha diversity than control lavages (Table 3). Beta diversity was similar.

3.2.2 Cases vs. controls: oral gargles

3.2.2.1 16S rRNA gene sequencing

Oral gargles from lung cancer and melanoma patients showed very little difference in prevalence (Figures 1B, 3A).

The genus *Prevotella* was more prevalent in controls (90%) compared to cases (50%), while *Granulicatella* was identified in all oral gargles from all patients (Figure 3A). Regarding relative abundance, *Streptococcus* and *Prevotella* (Huffnagle et al., 2017) were the most abundant genera in oral gargles from both cases and controls (Figure 3B). *Neisseria* was more abundant in controls compared to cases, while *Fusobacterium* was 2.5-fold higher in cases. Alpha and beta diversity showed no significant differences between gargles (Table 3; Figures 3C, D).

3.2.2.2 Whole genome shotgun sequencing

Neisseria subflava was twice as abundant in gargles from lung cancer cases compared to controls, and *Rothia dentocariosa* was slightly more abundant in controls versus cases (Table 2; Supplementary Figure S2B). Overall, bacterial abundance appeared quite similar among oral gargle samples. Neither alpha nor beta diversity indices demonstrated significant differences between cases and controls (Table 3; Supplementary Figures S2C, D). There were no significant differences in the case versus control viromes in the gargles (Supplementary Figure S3).

3.2.3 Cases: tumor versus normal (non-neoplastic) lung tissue

16S rRNA gene sequencing: Considering prevalence, *Propionibacterium*, *Atopobium*, and *Granulicatella* were identified in at least one tumor specimen but not in normal tissues (Figure 1C;

TABLE 1 Distribution of sample characteristics by lung cancer cases (n=10) versus melanoma control (n=10) status.

Characteristic	Categories	Cases, No. (%)	Controls, No. (%)	p value
Age, mean ± SD	–	71.2 ± 9.3	71.8 ± 12.0	0.791
Sex	Male	6 (60.0)	8 (80.0)	0.628
	Female	4 (40.0)	2 (20.0)	
Race	White	10 (100.0)	10 (100.0)	N/A
	Non-white	0 (0.0)	0 (0.0)	
Ethnicity	Hispanic	0 (0.0)	1 (10.0)	1.000
	Non-Hispanic	10 (100.0)	9 (90.0)	
Marital status	Married	7 (70.0)	7 (70.0)	1.000
	Divorced/separated	1 (10.0)	1 (10.0)	
	Widowed	1 (10.0)	2 (20.0)	
	Single	1 (10.0)	0 (0.0)	
Stage	I	8 (80.0)	3 (3.0)	0.212
	IA	3 (30.0)	2 (20.0)	
	IB	5 (50.0)	1 (10.0)	
	II	1 (10.0)	1 (10.0)	
	IIA	1 (10.0)	0 (0.0)	
	IIC	0 (0.0)	1 (10.0)	
	III	1 (10.0)	5 (50.0)	
	IIIA	1 (10.0)	2 (20.0)	
	IIIB	0 (0.0)	1 (10.0)	
	IIC	0 (0.0)	2 (20.0)	
Not staged	0 (0.0)	1 (10.0)		
Antibiotic use	< 2 Mo. before surgery	1 (10.0)	2 (20.0)	1.000
	> 2 Mo. before surgery	9 (90.0)	8 (80.0)	
Smoking status	Current/former smokers	5 (50.0)	5 (50.0)	N/A
	Never smokers	5 (50.0)	5 (50.0)	

Mo., months; N/A, not applicable; SD, standard deviation. Fisher's exact test and Wilcoxon rank sum test were used to determine if distributions of categorical and continuous variables differed according to case or control status, respectively.

Supplementary Figure S4A). The most abundant genus, albeit rare, in both the tumor and normal tissue was *Burkholderia*, though it is slightly more abundant in tumors (Table 2; Supplementary Figure S4B). The most abundant in both tissues were unclassified bacteria. Alpha and beta diversity were not significantly different between tissue types (Table 3; Supplementary Figures S4C, D), though normal tissue generally had lower alpha diversity than tumor tissue (Greathouse et al., 2018).

3.2.4 Cases: lavage versus gargle

3.2.4.1 16S rRNA gene sequencing

Several genera appeared slightly more prevalent in lavages than gargles (Figure 1D; Supplementary Figure S5A), including *Leptotrichia*, while genera like *Capnocytophaga* were more prevalent in gargles. The genera *Streptococcus*, *Prevotella*, and *Rothia* were more abundant in gargles than lavages, while *Leptotrichia* was >5-fold higher. By the Chao1 index, lavages maintained higher diversity than gargles (Table 3; Supplementary Figure S5C). Bacterial community structures significantly differed between gargles and lavages by both Bray Curtis

dissimilarity and unweighted UniFrac distance ($p=0.001$) (Supplementary Figure S5D).

3.2.4.2 Whole genome shotgun sequencing

Rothia dentocariosa was 2.7x more abundant in lavages versus gargles (Table 2; Supplementary Figure S6B). Interestingly, *E. coli* was not identified in the top 25 most prevalent species of lavages but was observed in all gargle samples. The species *R. mucilaginosa* (28.9% versus 17.9%), *Veillonella dispar* (5.9% versus 1.9%), and two species of *Prevotella* are more abundant in gargles versus lavages, respectively. On the other hand, *Porphyromonas* KLE 1280 (6.5% versus not within the top 25 most abundant species) and *G. adiaciens* (5.1% versus 2.5%) are more abundant in lavages than gargles, respectively (Supplementary Figure S6B). Shannon and Chao1 alpha diversity indices revealed gargles to be significantly more diverse compared to lavages (Table 3), significantly so by Shannon ($p=0.015$) and Chao1 ($p=0.004$) indices. Beta diversity by Bray Curtis dissimilarity ($p=0.005$) and unweighted UniFrac distance ($p=0.002$) showed significantly



differential bacterial community structures also between lavages and gargles (Table 3; Supplementary Figures S6C, D).

Considering viral signatures, prevalence appears different between these sample types (Supplementary Figures S7). For example, Human parainfluenza virus 3 and respiratory syncytial virus were identified in more lavages (10 and 6, respectively) than gargles (6 and 3, respectively). Human gammaherpesvirus 4 and beta herpesvirus 7 were identified in 5 and 6 gargle samples, but only 1 lavage specimen, respectively. Several non-human, plant and bacterial pathogens were identified in these samples as well. There also appeared to be a much higher proportion of unclassified viral taxa in lung cancer lavages (92.6%) versus gargles (74.6%). Alpha diversity by Chao1 was higher in gargles versus lavages (Table 3; Supplementary Figure S7C). Beta diversity was significant across Bray Curtis dissimilarity ($p=0.001$), weighted UniFrac distance ($p=0.004$), and unweighted UniFrac distance (Supplementary Figure S7D).

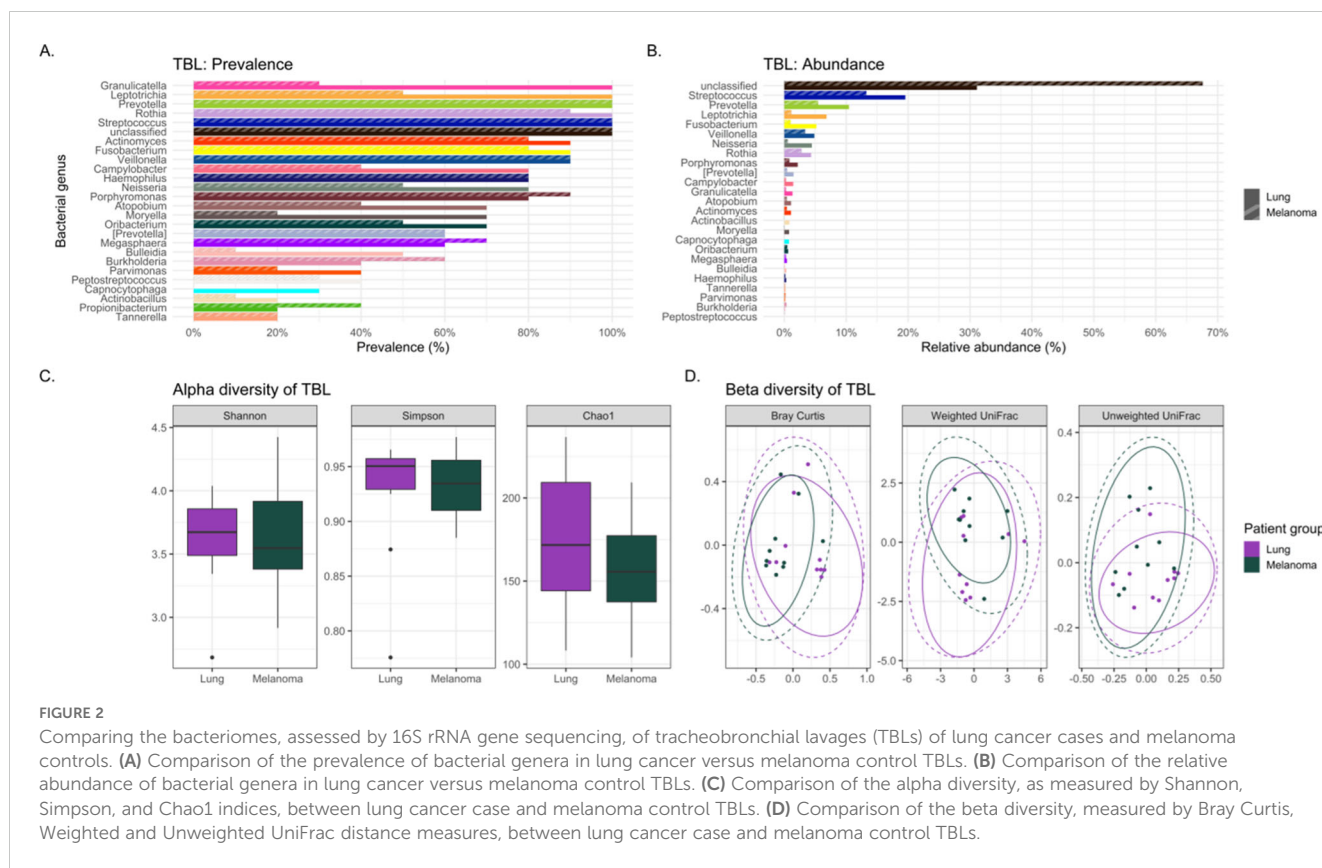
3.2.5 Cases: lavage versus tumor

16S rRNA gene sequencing: The genus *Burkholderia* is more prevalent in tumor tissue compared to lavages (90% versus 40%, respectively) (Supplementary Figure S8A). The tumor tissue mostly contained unclassified organisms but did maintain a higher relative abundance of *Burkholderia* than the lavage samples (1.4% versus 0.1%,

respectively). However, genera like *Streptococcus*, *Fusobacterium*, *Veillonella*, *Granulicatella*, *Neisseria*, *Leptotrichia*, *Prevotella*, and *Rothia*, amongst many others, were more abundant in lavages compared to tumor tissues. LEfSe showed that *Burkholderia* and its associated family were significantly differentially abundant in tumors as compared to lavages (data not shown). Many bacterial taxa significantly differentiated lavages from tumors, including *Granulicatella*, *Leptotrichia*, *Neisseria*, *Prevotella*, and *Rothia* (Supplementary Figures S8A, B). Alpha diversity was significantly different between lavages and tumor tissue by all three indices, Shannon ($p=0.0005$), Simpson ($p=0.0246$), and Chao1 ($p=0.0003$), whereby intra-sample diversity was consistently higher in lavages versus tumor tissue (Table 3; Supplementary Figure S8C). Similarly, all three Beta diversity measures showed that bacterial community structure significantly differs between the sample types ($p=0.001$ across all three indices) (Supplementary Figure S8D).

3.2.6 Cases: gargle versus tumor

16S rRNA gene sequencing: Most genera were more abundant in gargles versus tumor tissue. A substantially higher number of bacterial taxa were significantly discriminatory between gargles versus tumors, including *Granulicatella*, *Leptotrichia*, *Neisseria*, *Prevotella*, and *Rothia*. Alpha diversity was significantly different



between these two sample types for both Simpson ($p=0.007$) and Chao1 ($p=0.043$), indicating that oral gargles were more diverse in bacterial species richness and evenness compared to tumor tissue (Table 3). Beta diversity significantly differed across all three metrics, showing that these sample types maintain vastly disparate community structures.

4 Discussion

By a conservative estimate, there are at least a billion species of bacteria, but only 30,000 are formally named (Dykhuizen, 2005). Less than 2% of those can be cultured and identified in the laboratory (Wade, 2002). However, in the 1980's, the introduction of the polymerase chain reaction (PCR) targeting the highly conserved ribosomal genes (16S rRNA) of bacteria allowed for the identification of unculturable organisms (Wilson et al., 1997). Subsequent advances in DNA sequencing and other molecular techniques have allowed for inexpensive, rapid culture-independent identification of the vast array of resident microbiota in health (normobiosis or eubiosis) and in disease (dysbiosis), such as cancer. Despite the explosion in microbiome studies primarily in the gut looking for culprit bacteria that may cause malignancies, especially colon cancer and other gastrointestinal diseases, results have been disappointing since findings are inconsistent and not reproducible, likely due to variability of gut microbiota depending upon the sex, race, age, geographic location, and lifestyle factors of diet, exposures, drugs, and exercise (Pierrard and Seront, 2019).

Pierrard and associates and Nyein and associates observed that external factors such as antibiotics appeared to play a vital role in determining the bacterial composition in the gut microbiome which in turn affected immune checkpoint inhibitors (ICI) efficacy in cancer patients (Pierrard and Seront, 2019; Nyein et al., 2022). However, numerous human studies have failed to identify specific species or phyla associated with immunotherapy efficacy in cancer. Other factors, such as bacterial-dependent gut metabolite production that modifies blood metabolites and immune competence, may be essential to ICI efficacy (Tang, 2011). Although the taxonomy of gut microbiota has been under intense investigation to elucidate its role in various cancers, only a few studies have focused on the respiratory microbiome and its relationship to lung cancer. Therefore, we elected to perform this pilot study investigating possible microbial biomarkers indicating the presence of lung cancer and, if positive, then potentially they may be prognostic indicators of the efficacy of cancer treatment.

4.1 Oral gargles or tracheal lavage as potential microbial lung cancer biomarkers

4.1.1 Gargles

The primary focus of this study was to evaluate the oral and lower airway microbiome compositions of lung cancer cases compared to melanoma controls to reveal differences with potential applications as biomarkers for lung tumors. Therefore, we compared tracheobronchial lavages and oral gargles collected

TABLE 2 Relative abundance comparisons, using fold changes, of the top 25 most abundant (for taxa in >1% abundance) bacterial genera bacterial species, and viral taxa between case and control specimens and case tissue specimens.

Comparison		Sequencing Methodology		
		16S rRNA gene sequencing	WGSS (bacterial)	WGSS (viral)
Lavages: Lung cancer compared to melanoma	Higher abundance in Lung cancer	<ol style="list-style-type: none"> 1. <i>Neisseria</i> (7.82x) 2. <i>Leptotrichia</i> (5.94x) 3. <i>Campylobacter</i> (5.37x) 4. <i>Fusobacterium</i> (5.02x) 5. <i>Granulicatella</i> (4.55x) 6. <i>[Prevotella]</i> (3.10) 7. <i>Porphyromonas</i> and <i>Actinomyces</i> (2.60x) 8. <i>Atopobium</i> (2.50x) 9. <i>Prevotella</i> (1.91x) 10. <i>Rothia</i> (1.55x) 11. <i>Streptococcus</i> (1.47x) 	<ol style="list-style-type: none"> 1. <i>Gemella haemolysans</i> (26.17x) 2. <i>Neisseria subflava</i> (>15.93x) 3. <i>Porphyromonas KLE1280</i> (13.56x) 4. <i>Granulicatella adiacens</i> (>6.18x) 5. <i>Rothia dentocariosa</i> (1.80x) 6. <i>Rothia mucilaginosa</i> (1.39x) 	<ol style="list-style-type: none"> 1. Human betaherpesvirus 7 (>22.1x)
	Lower abundance in Lung Cancer	N/A	<ol style="list-style-type: none"> 1. <i>Megasphaera micronuciformis</i> (0.06x) 2. <i>Prevotella histicola</i> (0.07x) 3. <i>Veillonella dispar</i> (0.13x) 4. <i>Prevotella pallens</i> (0.18x) 	<ol style="list-style-type: none"> 1. Human respiratory syncytial virus (0.05x) 2. Tomato yellow leaf curl China betasatellite (0.37x) 3. Human parainfluenza virus 3 (0.61x)
Gargles: Lung cancer compared to melanoma	Higher abundance in Lung cancer	<ol style="list-style-type: none"> 1. <i>Fusobacterium</i> (2.47x) 2. <i>Atopobium</i> (2.09x) 3. <i>Leptotrichia</i> (1.80x) 4. <i>[Prevotella]</i> (1.77x) 5. <i>Porphyromonas</i> (1.34x) 6. <i>Granulicatella</i> (1.07x) 	<ol style="list-style-type: none"> 1. <i>Neisseria subflava</i> (2.05x) 2. <i>Prevotella ICM33</i> (1.31x) 	<ol style="list-style-type: none"> 1. Haemophilus phage HP2 (14.23x) 2. Haemophilus Phage HP1 (7.16x) 3. Human betaherpesvirus 7 (1.83x)
	Lower abundance in Lung cancer	<ol style="list-style-type: none"> 1. <i>Neisseria</i> (0.49x) 2. <i>Actinomyces</i> (0.51) 3. <i>Rothia</i> (0.76x) 4. <i>Veillonella</i> (0.89x) 5. <i>Prevotella</i> (0.93x) 6. <i>Streptococcus</i> (0.98x) 	<ol style="list-style-type: none"> 1. <i>Rothia dentocariosa</i> (0.23x) 2. <i>Prevotella melaninogenica</i> (0.57x) 3. <i>Veillonella dispar</i> (0.60x) 4. <i>Prevotella pallens</i> (0.80x) 5. <i>Rothia mucilaginosa</i> (0.93x) 	N/A
Lung cancer tumor compared to normal lung tissue	Higher abundance in Lung tumor	<ol style="list-style-type: none"> 1. <i>Burkholderia</i> (1.23x) 	N/A	N/A
	Lower abundance in Lung tumor	N/A	N/A	N/A
Lung cancer lavage compared to lung cancer gargle	Higher abundance in lung cancer lavage	<ol style="list-style-type: none"> 1. <i>Leptotrichia</i> (5.46x) 2. <i>Campylobacter</i> (2.50x) 3. <i>Fusobacterium</i> (1.13x) 4. <i>Neisseria</i> (1.04x) 	<ol style="list-style-type: none"> 1. <i>Rothia dentocariosa</i> (2.71x) 2. <i>Porphyromonas KLE1280</i> (2.53x) 3. <i>Gemella hemolysans</i> (1.79x) 4. <i>Granulicatella adiacens</i> (1.71x) 5. 	<ol style="list-style-type: none"> 1. Human parainfluenzavirus 3 (45.8x)
	Lower abundance in lung cancer lavage	<ol style="list-style-type: none"> 1. <i>Rothia</i> (0.35x) 2. <i>Atopobium</i> (0.52x) 3. <i>Porphyromonas</i> (0.54x) 4. <i>Streptococcus</i> (0.57x) 5. <i>Prevotella</i> (0.61x) 6. <i>Granulicatella</i> (0.77x) 7. <i>Actinomyces</i> (0.78x) 8. <i>Veillonella</i> (0.93x) 9. <i>[Prevotella]</i> (0.97x) 	<ol style="list-style-type: none"> 1. <i>Veillonella dispar</i> (0.27x) 2. <i>Prevotella pallens</i> (0.47x) 3. <i>Rothia mucilaginosa</i> (0.0.53x) 4. <i>Neisseria subflava</i> (0.94x) 	<ol style="list-style-type: none"> 1. Haemophilus phage HP1 (0.08x) 2. Haemophilus phage HP2 (0.06x) 3. Human betaherpesvirus 7 (0.68x)
Lung cancer lavage compared to lung tumor	Higher abundance in lung lavage	<ol style="list-style-type: none"> 1. <i>Streptococcus</i> (651.67x) 2. <i>Prevotella</i> (130.63x) 	N/A	N/A
	Lower abundance in lung lavage	<ol style="list-style-type: none"> 1. <i>Burkholderia</i> (0.08x) 	N/A	N/A
Lung cancer gargle compared to lung tumor	Higher abundance in lung gargle	<ol style="list-style-type: none"> 3. <i>Streptococcus</i> (1133.67x) 4. <i>Prevotella</i> (211.00x) 	N/A	N/A

(Continued)

TABLE 2 Continued

Comparison		Sequencing Methodology		
		16S rRNA gene sequencing	WGSS (bacterial)	WGSS (viral)
	Lower abundance in lung gargle	1. N/A	N/A	N/A

WGSS, whole genome shotgun sequencing; N/A, Not applicable. Oral commensal bacteria are shown in bold type. Fold changes were calculated as the relative abundance in the case sample divided by the control samples and control sample type one versus control sample type 2.

from lung cancer cases and controls. The results demonstrate that there are few significant differences in the oral gargles between lung cancer and melanoma patients in overall microbial composition using prevalence, abundance, and diversity measures, indicating unfortunately the readily available, noninvasively sampled oral gargle microbiome would *not* likely serve as a lung cancer biomarker.

4.1.2 Tracheal lavages

Beta diversity refers to the variation between the samples of one community (group) compared to another community, such that the microbiome composition of one group with a higher beta diversity indicates a more significant difference from the other group. By 16S rRNA gene sequencing data, beta diversity measured by Bray Curtis dissimilarity demonstrated significant differences ($p=0.022$) between case and control lavages, indicating that the bacterial communities in lavages from lung cancer versus melanoma were

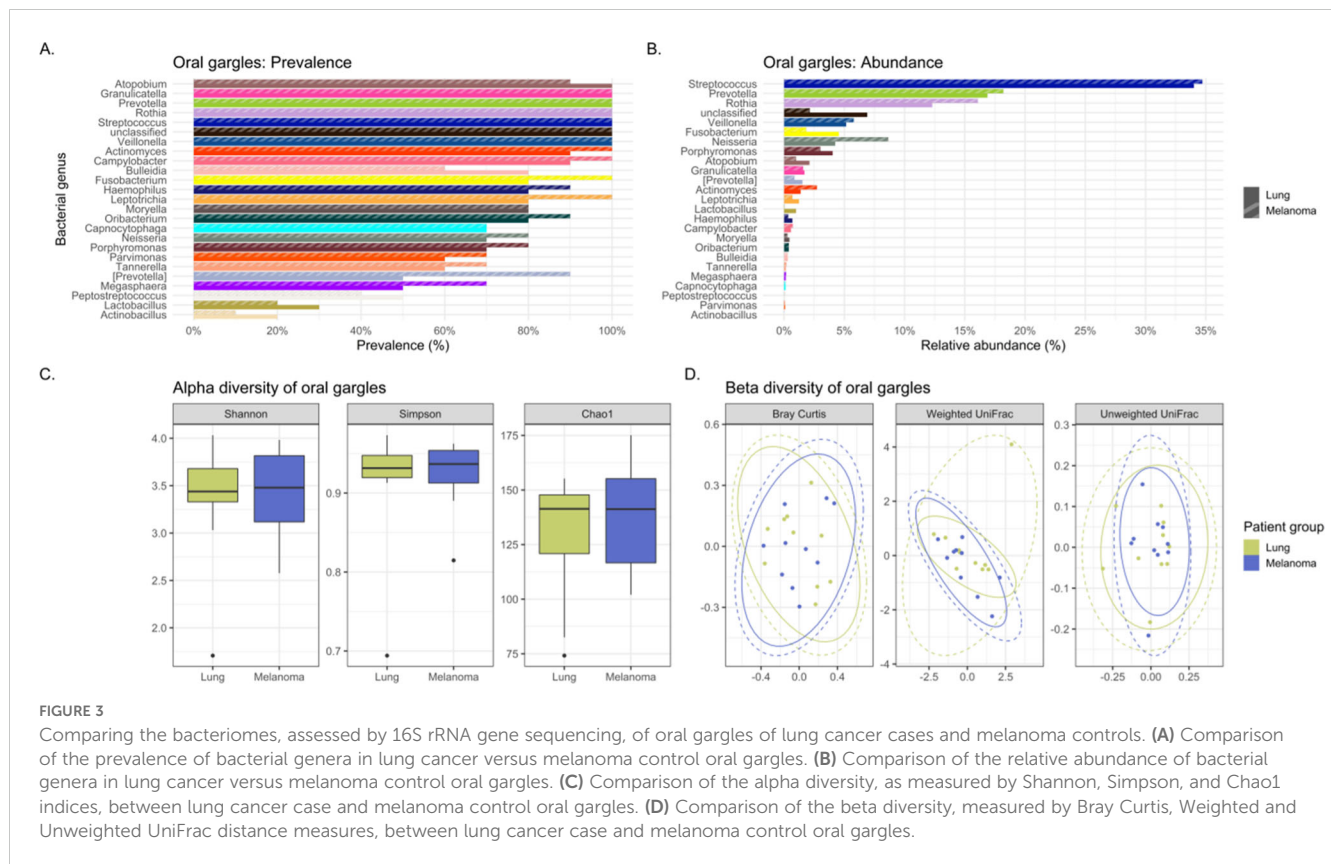
distinct. However, no such trend was observed for gargles. While the lung cancer and control tracheobronchial lavages were significantly different by 16S rRNA-derived beta diversity, it is difficult to say that the lavages will be able to distinguish lung cancer from non-lung cancer patients since WGSS did not clearly replicate these results.

Abundance denotes a specific bacterium's percentage of a sample's overall composition. In contrast, prevalence refers to the number (percentage) of cases in a specific group where a bacterium is detected. Significant trends were observed in abundance and prevalence of the lavages. Most noteworthy was *Granulicatella adiacens*, which was more prevalent and abundant in lung cancer cases. *Granulicatella adiacens* is a well-recognized oral commensal bacterium etiologically linked to endocarditis (Cincotta et al., 2015). We found this bacterium as one of the top 25 most abundant genera in lung cancer lavages, and it had a much higher prevalence appearing in virtually *all* lung cancer tracheal lavages (100%)

TABLE 3 Comparison of alpha and beta diversity of bacterial genera, bacterial species, and viral taxa between cases and controls and between sample types among cases.

Comparison	Sequencing methodology					
	16S rRNA gene sequencing		WGSS (bacterial)		WGSS (viral)	
	Alpha diversity (means)	Beta diversity	Alpha diversity (means)	Beta diversity	Alpha diversity (means)	Beta diversity
Lavages: Lung cancer compared to melanoma	1. lower 2. lower 3. higher	1. S 2. MS 3. MS	1. higher 2. lower 3. higher	1. NS 2. NS 3. NS	1. higher 2. higher 3. higher	1. NS 2. NS 3. NS
Gargles: Lung cancer compared to melanoma	1. lower 2. lower 3. lower	1. NS 2. NS 3. NS	1. lower 2. lower 3. lower	1. NS 2. NS 3. NS	1. higher 2. higher 3. lower	1. NS 2. NS 3. NS
Lung cancer tumor compared to normal lung tissue	1. higher 2. equal 3. higher	1. NS 2. NS 3. NS	N/A	N/A	N/A	N/A
Lung cancer lavage compared to lung cancer gargle	1. higher 2. higher 3. higher (S)	1. S 2. NS 3. S	1. lower (S) 2. lower 3. lower (S)	1. S 2. NS 3. S	1. lower 2. lower 3. lower (S)	1. S 2. S 3. S
Lung cancer lavage compared to lung tumor	1. higher (S) 2. higher (S) 3. higher (S)	1. S 2. S 3. S	N/A	N/A	N/A	N/A
Lung cancer gargle compared to lung tumor	1. higher (S) 2. higher (S) 3. higher (S)	5. S 6. S 7. S	N/A	N/A	N/A	N/A

Metric 1-3 for alpha diversity is Shannon index, Simpson index, and Chao1 index, respectively, and for beta diversity metric 1-3 is Bray-Curtis dissimilarity, weighted UniFrac distance, and unweighted UniFrac distance, respectively. S, significant ($p<0.05$), MS, marginally significant ($p=0.08$ or below); NS, not significant; WGSS, whole genome shotgun sequencing; N/A, Not applicable. Alpha diversity means were compared by T tests (Shannon) and Wilcoxon rank sum tests (Chao1 and Simpson).



versus only some control lavages (30%), despite being similarly abundant in gargle specimens of both groups. *Granulicatella adiacens* is the same organism that Cameron and associates found in the sputum of lung cancer patients but not controls in a recent pilot study, suggesting this as a potential novel lung cancer biomarker (Cameron et al., 2017). Replication of this finding in our study suggests that this microbe may be important to further investigate as a potential diagnostic biomarker (Cameron et al., 2017) and possibly even a predisposing factor to the development of lung cancer.

Additionally, lavage from the lower airways of our lung cancer cases harbored numerous supraglottic bacteria *Neisseria* (oral commensal), *Capnocytophaga* (oral commensal), *Leptotrichia* (oral commensal) and *Moryella* (oral and intestinal commensal) with twice the prevalence compared to control lavages. *Neisseria subflava*, which commonly colonizes the dorsum of the tongue, was also found in high abundance in lung cancer lavages.

LEfSe (linear discriminant analysis effect size) analysis is used to validate biomarkers by detailing features (bacterial taxa in lavages in this case) that distinguish two groups based on relative abundances. In our study, the LEfSe analysis did show several bacterial taxa, including Fusobacteria and *Neisseria* (especially the oral commensal *N. subflava*) to be significantly 8-fold differentially abundant in the tracheobronchial lavages of lung versus melanoma patients. These intriguing results strongly support continued research into the tracheal microbiota as potential biomarkers of lung cancer, especially the highly prevalent and abundant *Granulicatella adiacens* and *Neisseria subflava*.

4.2 Oral gargle or tracheal lavage as proxies of the tumor microbiome

Our study also investigated the potential utility of the oral gargle or tracheobronchial lavage microbiomes as proxies for the tumor microbiome in lung cancer. If the lavage and oral microbiomes were similar to the tumor microbiome, these less invasive sample types could be utilized to study the tumor microbiome more easily. Initially, lavages and gargles were compared to see if the gargle could mimic the lavage microbiota. However, significant differences were found between both bacterial and viral community structures (i.e., beta diversity) and alpha diversity in lavages and gargles. That is, the gargle microbiota were dissimilar from the lavages and cannot be used to represent the lavage microbiota.

Alpha diversity refers to the variation (how diverse it is) of bacteria within a single sample, such that a higher alpha diversity is usually associated with a more diverse, healthier microbiome. In our study, the alpha diversity of lavages versus gargles was likewise different, with gargles consistently maintaining higher bacterial and viral diversity by WGSS. LEfSe, performed on both 16S rRNA gene sequencing and WGSS data, also showed many differentially abundant bacterial taxa and some viral taxa between lavages and gargles. Unfortunately, as a result, these differences prevent oral gargles from acting as clinical proxies for tracheobronchial lavages. Further differences were identified between the tumor, gargles, and lavages, which precludes using these sample types as proxies of one another. This was not surprising, however, considering previous literature that has identified significant differences between lung tissue and oral microbiomes (Yu et al., 2016).

Despite these results, two recent studies have revealed the prognostic biomarker potential of the lung microbiome: one identified associations of the bronchoalveolar lavage microbiome with recurrence (Patnaik et al., 2021), and another identified *Enterobacter* in this same sample type associated with worse survival (Gomes et al., 2019), emphasizing the importance of continued investigation of the lung microbiome in lung cancer. It has already been hypothesized that Enterobacteriaceae, a bacterial family that expresses the common antigen lipopolysaccharide and identified in our study to be significantly more abundant in lavages and tumor tissue versus oral gargles in lung cancer cases, may induce inflammation in lung cancer that could be associated with poor prognosis (Gomes et al., 2019). Other studies have suggested that some microbiota may opportunistically invade damaged lung epithelium caused by smoking and drive tumorigenesis by producing free radicals like ROS/RNS that can damage the TP53 gene (Greathouse et al., 2018). Mouse models further suggest that lung microbiota may contribute to $\gamma\delta$ -T cell activation, which are cells that go on to release the cytokines IL-17A and IL-22 (Jin et al., 2019). These cytokines appeared to co-occur with tumor progression in the mice (Jin et al., 2019). Additional studies are needed to provide substantiated evidence of the mechanistic relationships between the microbiome, the immune system, and lung cancer.

4.3 The microbiome of tumor and non-neoplastic lung

Differences in the composition of the tumor and normal non-neoplastic tissue microbiomes of lung cancer patients were examined to highlight differences that might suggest a microbial contribution to lung carcinogenesis. If the microbiome signatures differed slightly but maintained similar microbial signatures between tumor and normal tissue, then it might indicate certain microbes from the typical lung environment that could have contributed to tumorigenesis or at least were opportunistic inhabitants of the tumor microenvironment. Indeed, sequencing revealed no significant differences in bacterial relative abundance and alpha or beta diversity between tumor and normal tissue samples. Interestingly, normal tissue had lower alpha diversity than tumor tissue, contrary to previously observed between tumor and healthy tissue controls (Mao et al., 2018). Finally, slight variations in bacterial prevalence were identified: a higher prevalence of the genera *Granulicatella* and *Burkholderia* in tumors was observed, and a higher prevalence of *Neisseria* and *Fusobacterium* in normal tissues.

The genus *Granulicatella*, in particular, has been found in a previous published study to inhabit the tumor microenvironment as it showed “lung cancer stage-specific increases in abundances” (Mur et al., 2018). As it becomes increasingly anaerobic, the production of useful metabolites for this genus increases (Mur et al., 2018). In the current study, *Granulicatella* was also identified as having a higher prevalence not only in tumors and normal tissue but also in tracheobronchial lavages of lung cancer patients versus melanoma

controls. Hosgood and associates also found a strong correlation between the finding of *Granulicatella* enriched in lung cancer patients’ oral and sputum samples compared to controls (Hosgood et al., 2014). This provides some intriguing preliminary data suggesting a possible carcinogenic role for some specific bacteria or at least opportunistic inhabitants of the tumor microenvironment, but testing in more extensive cohort studies is needed. The epidemiological relationship of *Granulicatella* may relate to the mechanism of elevated levels of microbial toxins and resultant inflammatory cytokines leading to chronic inflammation-associated carcinogenesis, such as that reported with *M. tuberculosis* and lung cancer (Elinav et al., 2013; Li et al., 2024).

Overall, tracheal lavages and gargles do not provide a consistent microbial signature for the tumor microbiome. Significant differences were observed between the lavage and tumor microbiomes. By all three alpha diversity indices, lavages maintained higher bacterial diversity than tumor tissue, and by all three beta diversity indices, bacterial communities are different between lavages and tumor tissue. LEfSe revealed many bacterial genera that are more abundant in lavages, like *Granulicatella* and *Neisseria*. However, one genus was more abundant in tumor tissue, namely *Burkholderia*, an important Gram-negative pathogen of lung infections in cystic fibrosis patients (Fauroux et al., 2004) and is the causative agent in the life-threatening respiratory illness melioidosis (Wiersinga et al., 2018). Beta diversity indicated significantly different bacterial community structures between oral gargles and tumor tissue. This was not surprising considering previous literature that has identified significant differences between lung tissue and oral microbiomes (Yu et al., 2016). Indices of alpha diversity also showed gargles to be significantly more diverse than tumor tissues.

Given their substantial differences, lavages and gargles cannot accurately represent the tumor microbiome. However, the genus *Burkholderia*, in particular, appeared more abundant and prevalent in tumor tissue than both lavages and gargles, suggesting a potential role in tumorigenesis or at least opportunistic inhabitants of the tumor microenvironment. However, further more mechanistic studies will need to be performed in order to suggest there is a connection between the mere presence of any of these microorganisms and lung carcinogenesis (Li et al., 2024).

4.4 Microaspiration

Silent microaspiration is the term used to describe the process in people when they have asymptomatic aspiration of small amounts of oropharyngeal or gastric secretions into their tracheobronchial tree and lungs, and it is felt to contribute to the genesis of many lung diseases such as pulmonary fibrosis (Lee et al., 2010). Microaspiration is a common event, occurring in as many as 50% of healthy people (Gleeson et al., 1997), although it is unknown how many have persistent colonization of the tracheobronchial tree with oral commensals. Previous studies by Segal and associates (Segal et al., 2016) demonstrated that the enrichment of oral commensals in the lower airways of normal individuals is associated with increased host

inflammatory tone and checkpoint inhibitor markers. Tsay and colleagues found this lower airway dysbiotic signature to distinguish between patients with lung cancer and benign lung nodules (Tsay et al., 2018).

Particularly notable differences in our study are the marked 16-, 6- and 6-fold higher abundance of the oral commensals *Neisseria subflava*, *Granulicatella adiacens*, and *Leptotrichia* in the lung cancer lavages versus controls. Also, the dysbiotic tracheal microbiome had extensive 2-3 times enrichment of oral microbiota (*Granulicatella*, *Capnocytophaga*, *Leptotrichia*, and *Neisseria*) in lung cancer patients compared to controls (Table 2), perhaps contributing to an inflammatory environment. The control lavages have a markedly reduced abundance of oral taxa, suggesting microaspiration and inflammation occur more significantly in lung cancer patients than in the control lavages. Indeed, beta diversity studies revealed significant differences ($p=0.022$) in bacterial community structures between the lung cancer and the control melanoma lavages.

Patnaik and associates also found oral aspiration as the source of lower airway microbiota in lung cancer, with the actual microbial community in bronchial lavage correlating with lung cancer recurrence after resection (Patnaik et al., 2021). Tsay and colleagues, using RNA-seq analysis of lower airway samples, found that the supraglottic predominant taxa found in the trachea were associated with upregulation of inflammatory pathways for p53 mutation, PI3K/PTEN, ERK and IL6/IL8, such that enrichment of the lower airway with oral commensals may increase local immune tone with upregulation of IL1, IL6, and ERK/MARK, in turn promoting tumor progression, hence suggesting microaspiration may be involved in lung cancer pathogenesis (Tsay et al., 2018).

4.5 Microvirome

While the characterization of the human microbiome's bacterial (and fungal) members, including the respiratory tree, has blossomed in the last decade, studies of the viral component still need to be completed. Many challenges exist with exploration of the microvirome due to its low biomass and enormous number of unclassified species. Even high throughput sequencing technologies are hampered by the small fraction of total DNA in the sample, often present in concentrations too low (< 1% of the reads) to be detected without amplification (Abbas, 2019). Additionally, contaminating human and bacterial DNA and RNA in samples are challenging. Finally, most reads with WGSS are commonly called "viral dark matter" since the results cannot be annotated into taxonomic categories (unclassified) due to the lack of species available in databases (Abbas, 2019). For example, as of 2022, the International Committee on Taxonomy of Viruses listed only 11,273 named viral species (International Committee on Taxonomy of Viruses, 2023) but they estimated there were unknown 100,939,140 viruses, excluding tens of millions of bacteriophage known and unknown species (Racaniello, 2013). As a result, only a relatively small number of viruses were identified in

our study, likely because of the low abundance and obvious inability to classify common though unknown viral organisms.

The virome of the tracheal lavages in our study was assessed through WGSS sequencing, which identified broadly similar prevalence and relative abundance between cases and controls among both lavages and gargles. Of the identified viruses, there was a higher abundance of human respiratory syncytial virus in the melanoma versus lung cancer patient lavages. Conversely, human beta-herpesvirus 7 was more abundant in lung cancer lavages versus controls. Oddly, many of the more prevalent viruses we identified in both cases and controls are plant pathogens (e.g., yellow vein viruses and tomato yellow leaf curl viruses), although yellow leaf curl is known to infect tobacco plants, which could conceivably enter the respiratory tree by cigarette smoking.

Viral signatures in the oral gargles demonstrated that bacteriophages targeting *Haemophilus* bacteria were more prevalent in melanoma controls versus lung cancer cases. Human-tropic viruses, such as endogenous retrovirus K and beta-herpesvirus 7, were similarly more prevalent between cases and controls. Although LEfSe identified several unclassified viral signatures as significantly different between cases and controls, neither alpha nor beta diversity indices demonstrated any significant differences between cases and controls. However, unclassified viral signatures were the most abundant in cases and controls.

A comparison of viral signatures between the oral gargle and lavages in the lung cancer cases demonstrated that the prevalence between these two sample types appears different. The most prevalent viral signature in lavages was human parainfluenza virus 3 and the human respiratory syncytial virus compared to gargles. However, human gamma-herpesvirus 4 was identified more commonly in gargles. LEfSe revealed several viral taxa that were significantly differentially abundant between gargles and lavages. Alpha diversity for viral signatures was higher across all three indices in gargles compared to lavages. Finally, beta diversity revealed significant differences in viral community structure between gargles and lavages.

Unfortunately, our WGSS and bioinformatics approaches left most of the viral taxa unclassified in lung cancer lavages (92.6%) and gargles (74.6%), thus hampering meaningful evaluation of the viral microbiome.

4.6 Proinflammatory microbiome

Ultimately, the question arises as to why a dysbiotic, inflammatory tracheobronchial microbiome appears to be uniformly associated with lung cancer, and perhaps the answer lies in the multifactorial nature of carcinogenesis as suggested by the human papillomavirus (HPV) and cervical cancer picture. HPV has been convincingly proven to cause 99.7% of cervical cancer (Walboomers et al., 1999). If a woman is found to have high risk HPV on her pelvic exam specimens, then she is at elevated risk for the malignancy, yet at most, only 8% of high risk, HPV-positive women ever develop either pre-cancerous cervical changes or frank

cancer (Rodríguez et al., 2010). Recent studies suggest that the primary factor determining the ability of HPV to transform cervical cells is the vaginal microbiota, such that a dysbiotic, inflammatory microbiome is needed. A eubiotic, low diversity, low pH vaginal microbiome, mainly dominated by lactobacillus species, likely helps clear HPV infections and is also cytotoxic by secreting bacteriocins that modulate the immune system to inhibit viral activity. However, the dysbiotic, proinflammatory vaginal microbiome induces oxidative DNA damage and promotes viral transformation of the cervix by the resident HPV (Mitra et al., 2016).

Therefore, we might postulate a similar scenario for the consistent finding of a dysbiotic tracheobronchial microbiome in lung cancer patients. In fact, chronic lung inflammation along with concurrent inflammatory cytokines, growth factors and adhesions molecules provide a favorable environment to promote tumor cell growth and proliferation (Wang et al., 2021). Suppose some or all lung cancer is “caused” by one or more oncogenic viruses suggested by prior studies, such as HPV (Xiong et al., 2017), bovine leukemia virus (Robinson et al., 2016), and HTLV-1 (Nomori et al., 2011; Robinson et al., 2016). In that case, the development of a dysbiotic, inflammatory tracheobronchial microbiome, such as that found in the current study and others, may be the promoting factor that allows existing colonized, oncogenic viruses to cause malignant transformation in the lung. However, this attractive hypothesis will require many future studies to substantiate.

4.7 Limitations and strengths

The small sample size of this study results in significant limitations that may have obscured statistically significant differences in microbiome compositions between lung cancer cases and melanoma controls. The use of melanoma patients undergoing resection of a peripheral extremity tumor under general anesthesia was the best possible proxy for control patients to collect oral gargles and tracheal lavages. Theoretically, the best control would be bronchoscopy and lavages on normal people without cancer or lung disease, but we did not feel ethically comfortable attempting to recruit normal people for a procedure that carries some procedural risk. In addition, the small sample sizes prevent us from appropriate sub-analysis of smokers versus non-smoker results. Future research will require larger cohorts to allow sufficient power to detect clinically meaningful differences that could hold biomarker potential. Although, a more thorough evaluation of specimen contamination could be implemented in future studies, our case and control samples were processed with similar reagents and at the same time, so contamination should not result in a substantial difference in microbiome signatures between our comparison groups.

Due to the study’s case-control design, the effect of changes in the microbiome over time could not be established to identify when microbial alterations may have occurred in lung cancer patients as compared to the controls. Therefore, further research into microbial dysbiosis in lung cancer will ideally require collecting samples at various time points using prospective cohort designs, although this

would be a more challenging study to accomplish. However, such a study might enable a better understanding of when microbial dysbiosis occurs and how it is associated with clinically important events, such as disease initiation, progression, or treatment response.

Despite these limitations, this study has several significant strengths, including the direct comparison of the oral, tracheal, lung tumor, and non-neoplastic lung microbiome versus the oral and tracheal microbiome of control patients without lung cancer. Also important is the use of WGSS in addition to 16S rRNA gene sequencing of the specimens. WGSS enabled greater taxonomic resolution, specifically to the species level—more so than 16S rRNA gene sequencing would have enabled alone.

WGSS additionally enabled the elucidation of viral, not merely bacterial, signatures to generate a more holistic view of the microbial environments among the different sample types. However, since the vast majority (93%) of viral signatures were unclassified, we are conducting additional research studies focusing on the more specific PCR approaches targeting specific viral taxa suspected to be associated with lung cancer, including human retroviruses, human papillomavirus (Srinivasan et al., 2009; Xiong et al., 2017), and hepatitis B virus (Sundquist et al., 2014), as documented in our prior pan-microbial array study of biobanked frozen lung cancers (Robinson et al., 2016).

4.8 Conclusions

The primary focus of this study was to evaluate the oral and lower airway microbiome compositions of lung cancer cases compared to melanoma controls to reveal any differences that may have potential applications as lung cancer biomarkers. Indeed, in this case-control study, we found that bacterial communities of lung cancer versus melanoma lavages were significantly different although unfortunately no such trend was observed for gargles. Several bacterial taxa, including oral commensals *Neisseria subflava* and *Granulicatella adiacens* were significantly 8-fold differentially abundant in the tracheobronchial lavages of lung versus melanoma patients suggesting these organisms may warrant future study as more-invasive potential confirmatory biomarkers of lung cancer when bronchoscopy results are equivocal. Like other published studies, we found a dysbiotic tracheal microbiome with extensive 2-3 times enrichment of oral microbiota (higher abundances of oral commensals *Granulicatella*, *Capnocytophaga*, *Leptotrichia* and *Neisseria*) in lung cancer patients compared to controls. The control lavages have a markedly reduced abundance of oral taxa. This suggests far more microaspiration and resultant inflammation occurs in lung cancer patients.

The tumor microbiome shows substantial differences between the lavages and gargles, such that they cannot accurately stand in as proxies of the tumor microbiome. However, the genus *Burkholderia* in particular, appeared more abundant and prevalent in tumor tissue versus both lavages and gargles, suggesting a potential role of this organism in tumorigenesis or at least it may be an opportunistic inhabitant of the tumor microenvironment. Finally, for the microbiome, our WGSS and bioinformatics approaches left the

vast majority of the viral taxa unclassified in lung cancer and control lavages (92.6%) and in gargles (74.6%), thus hampering a meaningful evaluation of the role of virus in lung cancer. Overall, this study generated encouraging preliminary results confirming some of the findings in the published literature that can be used in hypothesis generation for basing future studies directed at identifying potential microbial biomarkers of lung cancer.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI SRA, PRJNA1177881.

Ethics statement

All patients provided informed consent to participate in this study and only de-identified patient data is included in this manuscript. This study was approved by the Moffitt Scientific Review Committee and the Liberty Institutional Review Board, Protocol 14.12.0036 (MCC 17976).

Author contributions

AB: Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. SH: Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. CP: Conceptualization, Funding acquisition, Investigation, Project administration, Writing – original draft, Writing – review & editing. SP: Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. NL: Writing – original draft, Writing – review & editing. RT: Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. YK: Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. LR: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by a Moffitt Cancer Center Team Science Award granted to LR, and CP, Co-Principal Investigators.

Acknowledgments

This work has been supported in part by the Tissue Core, Molecular Genomics Core, and Biostatistics and Bioinformatics

Shared Resource at the H. Lee Moffitt Cancer Center & Research Institute, a comprehensive cancer center designated by the National Cancer Institute and funded in part by Moffitt's Cancer Center Support Grant (P30-CA076292). The authors acknowledge Maria F. Gomez, who assisted with study development and initial analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frmbi.2024.1457537/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Comparing the bacteriomes, assessed by whole genome shotgun sequencing, of tracheobronchial lavages from lung cancer cases and melanoma controls. (A) Comparison of the prevalence of bacterial species. (B) Comparison of the relative abundance of bacterial species. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures.

SUPPLEMENTARY FIGURE 2

Comparing the bacteriomes, assessed by whole genome shotgun sequencing, of oral gargles of lung cancer cases and melanoma controls. (A) Comparison of the prevalence of bacterial species in lung cancer versus melanoma control oral gargles. (B) Comparison of the relative abundance of bacterial species in lung cancer versus melanoma control oral gargles. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices, between lung cancer case and melanoma control oral gargles. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures, between lung cancer case and melanoma control oral gargles.

SUPPLEMENTARY FIGURE 3

Comparing the viromes, assessed by whole genome shotgun sequencing, of oral gargles of lung cancer cases and melanoma controls. (A) Comparison of the prevalence of viral species in lung cancer versus melanoma control oral gargles. (B) Comparison of the relative abundance of viral species in lung cancer versus melanoma control oral gargles. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices, between lung cancer case and melanoma control oral gargles. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures, between lung cancer case and melanoma control oral gargles.

SUPPLEMENTARY FIGURE 4

Comparing the bacteriomes, assessed by 16S rRNA gene sequencing, of tumor and normal tissue from lung cancer patients. (A) Comparison of the prevalence of bacterial genera in lung cancer tumor versus normal tissue. (B) Comparison of the relative abundance of bacterial genera in lung cancer tumor versus normal tissue. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices, between tumor and normal tissue from lung cancer patients. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures, between tumor and normal tissue from lung cancer patients.

SUPPLEMENTARY FIGURE 5

Comparing the bacteriomes, assessed by 16S rRNA gene sequencing, of TBLs and oral gargles of lung cancer cases. (A) Comparison of the prevalence of bacterial genera in lung cancer oral gargles versus TBLs. (B) Comparison of the relative abundance of bacterial genera in lung cancer versus melanoma control oral gargles. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices, between lung cancer oral gargles and TBLs. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures, between lung cancer oral gargles and TBLs.

SUPPLEMENTARY FIGURE 6

Comparing the bacteriomes, assessed by whole genome shotgun sequencing, of TBLs and oral gargles of lung cancer cases. (A) Comparison of the prevalence of bacterial species in lung cancer oral gargles versus TBLs. (B) Comparison of

the relative abundance of bacterial species in lung cancer oral gargles versus TBLs. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices, between lung cancer oral gargles and TBLs. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures, between lung cancer oral gargles and TBLs.

SUPPLEMENTARY FIGURE 7

Comparing the viromes, assessed by whole genome shotgun sequencing, of oral gargles and TBLs of lung cancer cases. (A) Comparison of the prevalence of viral species in lung cancer oral gargles versus TBLs. (B) Comparison of the relative abundance of viral species in lung cancer oral gargles versus TBLs. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices, between lung cancer oral gargles and TBLs. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures, between lung cancer oral gargles and TBLs.

SUPPLEMENTARY FIGURE 8

Comparing the bacteriomes, assessed by 16S rRNA gene sequencing, of TBLs and tumor tissue of lung cancer cases. (A) Comparison of the prevalence of bacterial genera in lung cancer TBLs versus tumor tissue. (B) Comparison of the relative abundance of bacterial genera in lung cancer TBLs versus tumor tissue. (C) Comparison of the alpha diversity, as measured by Shannon, Simpson, and Chao1 indices, between lung cancer TBLs versus tumor tissue. (D) Comparison of the beta diversity, measured by Bray Curtis, Weighted and Unweighted UniFrac distance measures, between lung cancer TBLs versus tumor tissue.

References

- Abbas, A. (2019). *The Human Lung Viral Microbiome In Health And Disease*. U. o (Philadelphia, PA: Pennsylvania, Ed.).
- American Cancer Society (2024) (Atlanta, GA).
- American Lung Association (2024) (Chicago, IL).
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Cameron, S. J. S., Lewis, K. E., Huws, S. A., Hegarty, M. J., Lewis, P. D., Pachebat, J. A., et al. (2017). A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. *PLoS One* 12, e0177062. doi: 10.1371/journal.pone.0177062
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Cincotta, M. C., Coffey, K. C., Moonah, S. N., Uppal, D., and Hughes, M. A. (2015). Case report of *Granulicatella adiacens* as a cause of bacterascites. *Case Rep. Infect. Dis.* 2015, 132317. doi: 10.1155/2015/132317
- CosmosID Inc (2021) (CosmosID Inc.).
- Dykhuizen, D. (2005). Species numbers in bacteria. *Proc. Calif Acad. Sci.* 56, 62–71.
- Edgar, R. C. (2016). UCHIME2: improved chimera prediction for amplicon sequencing. *bioRxiv*, 074252. doi: 10.1101/074252
- Elinav, E., Nowarski, R., Thaiss, C. A., Hu, B., Jin, C., and Flavell, R. A. (2013). Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat. Rev. Cancer* 13, 759–771. doi: 10.1038/nrc3611
- Fauroux, B., Hart, N., Belfar, S., Boulé, M., Tillous-Borde, I., Bonnet, D., et al. (2004). *Burkholderia cepacia* is associated with pulmonary hypertension and increased mortality among cystic fibrosis patients. *J. Clin. Microbiol.* 42, 5537–5541. doi: 10.1128/JCM.42.12.5537-5541.2004
- Gleeson, K., Eggle, D. F., and Maxwell, S. L. (1997). Quantitative aspiration during sleep in normal subjects. *Chest* 111, 1266–1272. doi: 10.1378/chest.111.5.1266
- Gomes, S., Cavadas, B., Ferreira, J. C., Marques, P. I., Monteiro, C., Sucena, M., et al. (2019). Profiling of lung microbiota discloses differences in adenocarcinoma and squamous cell carcinoma. *Sci. Rep.* 9, 12838. doi: 10.1038/s41598-019-49195-w
- Greathouse, K. L., White, J. R., Vargas, A. J., Bliskovsky, V. V., Beck, J. A., von Muhlen, N., et al. (2018). Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol.* 19, 123. doi: 10.1186/s13059-018-1501-6
- Henley, S. J., Ward, E. M., Scott, S., Ma, J., Anderson, R. N., Firth, A. U., et al. (2020). Annual report to the nation on the status of cancer, part I: National cancer statistics. *Cancer* 126, 2225–2249. doi: 10.1002/ncr.32802
- Hosgood, H. D. 3rd., Sapkota, A. R., Rothman, N., Rohan, T., Hu, W., Xu, J., et al. (2014). The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. *Environ. Mol. Mutagen* 55, 643–651. doi: 10.1002/em.21878
- Huffnagle, G. B., Dickson, R. P., and Lukacs, N. W. (2017). The respiratory tract microbiome and lung inflammation: a two-way street. *Mucosal Immunol.* 10, 299–306. doi: 10.1038/mi.2016.108
- International Committee on Taxonomy of Viruses (2023) (Jena, Germany).
- Jin, C., Lagoudas, G. K., Zhao, C., Bullman, S., Bhutkar, A., Hu, B., et al. (2019). Commensal microbiota promote lung cancer development via $\gamma\delta$ T cells. *Cell* 176, 998–1013.e1016. doi: 10.1016/j.cell.2018.12.040
- Lee, J. S., Collard, H. R., Raghu, G., Sweet, M. P., Hays, S. R., Campos, G. M., et al. (2010). Does chronic microaspiration cause idiopathic pulmonary fibrosis? *Am. J. Med.* 123, 304–311. doi: 10.1016/j.amjmed.2009.07.033
- Lee, S. H., Sung, J. Y., Yong, D., Chun, J., Kim, S. Y., Song, J. H., et al. (2016). Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. *Lung Cancer* 102, 89–95. doi: 10.1016/j.lungcan.2016.10.016
- Li, R., Li, J., and Zhou, X. (2024). Lung microbiome: new insights into the pathogenesis of respiratory diseases. *Signal Transduction Targeted Ther.* 9, 19. doi: 10.1038/s41392-023-01722-y
- Mao, Q., Jiang, F., Yin, R., Wang, J., Xia, W., Dong, G., et al. (2018). Interplay between the lung microbiome and lung cancer. *Cancer Lett.* 415, 40–48. doi: 10.1016/j.canlet.2017.11.036
- Mao, Y., Yang, D., He, J., and Krasna, M. J. (2016). Epidemiology of lung cancer. *Surg. Oncol. Clin. N Am.* 25, 439–445. doi: 10.1016/j.soc.2016.02.001
- Mitra, A., MacIntyre, D. A., Marchesi, J. R., Lee, Y. S., Bennett, P. R., and Kyrgiou, M. (2016). The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next? *Microbiome* 4, 58. doi: 10.1186/s40168-016-0203-0
- Moffatt, M. F., and Cookson, W. O. (2017). The lung microbiome in health and disease. *Clin. Med. (Lond)* 17, 525–529. doi: 10.7861/clinmedicine.17-6-525
- Mur, L. A., Huws, S. A., Cameron, S. J., Lewis, P. D., and Lewis, K. E. (2018). Lung cancer: a new frontier for microbiome research and clinical translation. *Eccancermediscalsci* 12, 866. doi: 10.3332/ecancer.2018.866
- Nomori, H., Mori, T., Iyama, K., Okamoto, T., and Kamakura, M. (2011). Risk of bronchioloalveolar carcinoma in patients with human T-cell lymphotropic virus type 1 (HTLV-1): case-control study results. *Ann. Thorac. Cardiovasc. Surg.* 17, 19–23. doi: 10.5761/atcs.09.01529
- Nyein, A. F., Bari, S., Hogue, S., Zhao, Y., Maller, B., Sha, S., et al. (2022). Effect of prior antibiotic or chemotherapy treatment on immunotherapy response in non-small cell lung cancer. *BMC Cancer* 22, 101. doi: 10.1186/s12885-022-09210-2
- Patnaik, S. K., Cortes, E. G., Kannisto, E. D., Punnanitont, A., Dhillon, S. S., Liu, S., et al. (2021). Lower airway bacterial microbiome may influence recurrence after resection of early-stage non-small cell lung cancer. *J. Thorac. Cardiovasc. Surg.* 161, 419–429.e416. doi: 10.1016/j.jtcvs.2020.01.104

- Pierrard, J., and Seront, E. (2019). Impact of the gut microbiome on immune checkpoint inhibitor efficacy—a systematic review. *Curr. Oncol.* 26, 395–403. doi: 10.3747/co.26.5177
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Racaniello, V. (2013). *Virology Blog* (New York: Virology Blog, New York).
- Richards, T. B., Doria-Rose, V. P., Soman, A., Klabunde, C. N., Caraballo, R. S., Gray, S. C., et al. (2019). Lung cancer screening inconsistent with U.S. Preventive services task force recommendations. *Am. J. Prev. Med.* 56, 66–73. doi: 10.1016/j.amepre.2018.07.030
- Robinson, L. A., Jaing, C. J., Pierce Campbell, C., Magliocco, A., Xiong, Y., Magliocco, G., et al. (2016). Molecular evidence of viral DNA in non-small cell lung cancer and non-neoplastic lung. *Br. J. Cancer* 115, 497–504. doi: 10.1038/bjc.2016.213
- Rodríguez, A. C., Schiffman, M., Herrero, R., Hildesheim, A., Bratti, C., Sherman, M. E., et al. (2010). Longitudinal study of human papillomavirus persistence and cervical intraepithelial neoplasia grade 2/3: critical role of duration of infection. *J. Natl. Cancer Inst.* 102, 315–324. doi: 10.1093/jnci/djq001
- Rutala, W. A., and Weber, D. J. (2008) (Chapel Hill, NC: Centers for Disease Control and Disease Prevention).
- Segal, L. N., Clemente, J. C., Tsay, J. C., Koralov, S. B., Keller, B. C., Wu, B. G., et al. (2016). Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat. Microbiol.* 1, 16031. doi: 10.1038/nmicrobiol.2016.31
- Srinivasan, M., Taioli, E., and Ragin, C. C. (2009). Human papillomavirus type 16 and 18 in primary lung cancers—a meta-analysis. *Carcinogenesis* 30, 1722–1728. doi: 10.1093/carcin/bgp177
- Sundquist, K., Sundquist, J., and Ji, J. (2014). Risk of hepatocellular carcinoma and cancers at other sites among patients diagnosed with chronic hepatitis B virus infection in Sweden. *J. Med. Virol.* 86, 18–22. doi: 10.1002/jmv.23754
- Tang, J. (2011). Microbial metabolomics. *Curr. Genomics* 12, 391–403. doi: 10.2174/138920211797248619
- Tsay, J. J., Wu, B. G., Badri, M. H., Clemente, J. C., Shen, N., Meyn, P., et al. (2018). Airway microbiota is associated with upregulation of the PI3K pathway in lung cancer. *Am. J. Respir. Crit. Care Med.* 198, 1188–1198. doi: 10.1164/rccm.201710-2118OC
- Wade, W. (2002). Unculturable bacteria—the uncharacterized organisms that cause oral infections. *J. R. Soc. Med.* 95, 81–83. doi: 10.1177/014107680209500207
- Walboomers, J. M., Jacobs, M. V., Manos, M. M., Bosch, F. X., Kummer, J. A., Shah, K. V., et al. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* 189, 12–19. doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F
- Wang, D., Cheng, J., Zhang, J., Zhou, F., He, X., Shi, Y., et al. (2021). The role of respiratory microbiota in lung cancer. *Int. J. Biol. Sci.* 17, 3646–3658. doi: 10.7150/ijbs.51376
- Wiersinga, W. J., Virk, H. S., Torres, A. G., Currie, B. J., Peacock, S. J., Dance, D. A. B., et al. (2018). Melioidosis. *Nat. Rev. Dis. Primers* 4, 17107. doi: 10.1038/nrdp.2017.107
- Wilson, M. J., Weightman, A. J., and Wade, W. G. (1997). Applications of molecular ecology in the characterization of uncultured microorganisms associated with human disease. *Rev. Res. Med. Microbiol.* 8, 91–102. doi: 10.1097/00013542-199704000-00005
- Woo, Y. R., Cho, S. H., Lee, J. D., and Kim, H. S. (2022). The human microbiota and skin cancer. *Int. J. Mol. Sci.* 23 (3), 1813. doi: 10.3390/ijms23031813
- World Health Organization (2022) (Geneva, Switzerland: World Health Organization).
- Xiong, W. M., Xu, Q. P., Li, X., Xiao, R. D., Cai, L., and He, F. (2017). The association between human papillomavirus infection and lung cancer: a system review and meta-analysis. *Oncotarget* 8, 96419–96432. doi: 10.18632/oncotarget.21682
- Yu, G., Gail, M. H., Consonni, D., Carugno, M., Humphrys, M., Pesatori, A. C., et al. (2016). Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome Biol.* 17, 163. doi: 10.1186/s13059-016-1021-1
- Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620. doi: 10.1093/bioinformatics/btt593