

Metagenomic Characterization of The Tracheobronchial Microbiome in Lung Cancer

Stephanie Hogue

Moffitt Cancer Center

Christine M. Pierce

Moffitt Cancer Center

Shirlene Paul

Moffitt Cancer Center

Ram Thapa

Moffitt Cancer Center

Youngchul Kim

Moffitt Cancer Center

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

Moffitt Cancer Center

Research Article

Keywords: Non-small cell lung cancer; Tracheal microbiota, Oral microbiota, Lung microbiome, Microaspiration, Lung inflammation

Posted Date: December 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1162508/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

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.

Methods

In this pilot case-control study, we recruited ten patients with resectable 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 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*, appeared in higher abundance in case versus control lavages. Case lavages also maintained higher relative abundances of 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.

Background

Lung cancer is the most frequent cancer worldwide and the most common cause of cancer deaths with 1.8 million deaths in 2020.(1) In the U.S., 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.(2) An estimated 135,720 Americans died from lung cancer in 2020, exceeding the number of deaths expected from colon, breast, and prostate cancers, combined.(3) Despite enormous research and treatment efforts, the high fatality rate of this malignancy (82%) has changed little over the last few decades.(4) Furthermore, delayed diagnosis of lung cancer continues, with 85% of cases not being recognized until later stages, contributing to the high mortality rate associated with this disease.(5) Screening chest computed tomography offers the opportunity to discover earlier stage disease in high risk individuals, but, despite its ready availability, is underutilized with only 3.9% of eligible people obtaining a scan.(6) 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 microbiota of other anatomic sites. Until recently, the lungs were considered sterile, but evidence indicates this organ is indeed colonized by commensal microbes, including *Acinetobacter*, *Pseudomonas* and *Ralstonia*.⁽⁷⁾ Furthermore, composition and function of the microbiota in lung tissue are distinct from other anatomic sites, including the oral cavity.⁽⁷⁾

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).^(8, 9) Additionally, hypotheses regarding an association between the lung microbiome and lung cancer, potentially mediated by inflammation, has been suggested.⁽⁹⁾ 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 the lung and oral bacteriomes and viromes in early-stage non-small cell lung cancer (NSCLC) patients compared to melanoma controls to discover potentially distinguishable features in the compositions of the oral, tracheal and tumor microbiomes of NSCLC patients. Those results may allow assessment of the potential for minimally invasive samples to act as proxies for the tumor microbiome. The most immediate benefit of finding a microbial “signature” for lung cancer is the possibility of developing a reliable screening technique for detection of high-risk individuals with early cancers.

Methods

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 controls because this cancer type did not show clear evidence of microbial etiology and patients were already undergoing anesthesia with intubation for major resection of extremity melanomas. Lung cancer cases and melanoma controls were matched by age (± 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 one month or prebiotics/probiotics within 3 months of surgery. This study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments and was approved by the Liberty Institutional Review Board, Protocol 14.12.0036 (MCC 17976). Informed consent was obtained from all participants.

Specimen collection

Tracheobronchial Lavages (TBL)

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 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.⁽¹⁰⁾ An intravenous, preoperative, prophylactic antibiotic was started at the time of bronchoscopy so it would not have reached a therapeutic blood level when we obtained lavage samples. Approximately 20 mL of tracheobronchial lavage fluid was collected 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.

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 that was 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 of PBS and aliquoted as two 0.6mL aliquots that were stored at -80°C.

Tissue Samples

Only lung cancer patients provided tumor and adjacent non-neoplastic lung tissue specimens. Immediately after resection, LAR, while wearing a mask, removed 1cm³ from the tumor using sterile instruments in a sterile field in the frozen section room. A similar-sized, non-neoplastic lung specimen was also harvested in the same manner at a distance from the tumor. Tissue specimens were transported to the lab and snap frozen in LN before undergoing macrodissection and long-term storage at -80°C.

DNA extraction

Microbial DNA was extracted 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 10,000xg for 30 seconds at room temperature with resulting supernatant collected. The supernatant was processed to remove PCR inhibitors and eluted with 100µL of buffer. DNA were quantitated using Qubit and quality 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 or about half of the total tissue volume was utilized. Briefly, the tissue was added 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 resulting supernatant further processed and eluted in buffer AE.

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.

Metagenomic Whole Genome Shotgun Sequencing (WGSS)

On DNA 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), and 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 species level and to identify viral signatures.

Bioinformatics and Statistical Analyses

16s rRNA Sequencing Data Analysis

Paired-end sequencing reads were cleaned using Trimmomatic v. 0.39 (11) 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.(12) Next, the cleaned reads were merged with PEAR (0.9.10)(13) and operational taxonomic units (OTUs) were generated by open reference of QIIME1.9.1 pipeline.(14) Only OTUs with the minimum total observation count of 100 were retained. Database used for taxonomic assignment was Silva 128 97_otus_16S.fasta.(15) Alpha- and Beta-diversity were analyzed using QIIME1.9.1. The taxonomy plots were based on 25 most prevalent OTUs. PERMANOVA was used to compare beta diversity estimate.

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 and controls and sample types. Two-sided P values <0.05 were considered statistically significant. Statistical analysis was completed using Phyloseq package in R software (v3.1.1 and v4.1.0, The R Foundation, Vienna, Austria).

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, fungi, virulence markers, and antimicrobial resistance markers curated by CosmosID (CosmosID, Inc., Germantown, MD(16))] with per-sample computation (searches short sequence reads or contigs from draft de novo assemblies against fingerprint sets), detect and classify microbial sequencing reads. To exclude false positives, the platform filtered reads using a filtering threshold derived from internal scores determined by analyzing a large number of diverse metagenomes.

Results

Patient Characteristics

All 20 participants (ten NSCLC cases and ten melanoma controls) were Caucasian (Table 1). Cases had a higher percentage of females compared to 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). The majority of cases (90%) and controls (80%) had not received antibiotics within 2 months prior to their surgery. No significant differences were observed in any of the characteristics measured between cases and controls.

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)	
	IIIC	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)	

Abbreviations: 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.

Microbial Profiling

Cases vs. Controls:

Tracheobronchial Lavages

16S rRNA gene sequencing: The usual lower airway genera *Streptococcus* and *Prevotella*(17) were identified in all lavages (Figures 1A and 2A), while the oral commensals *Granulicatella* (100% versus 30%), *Leptotrichia* (100% versus 50%), *Moryella* (70% versus 20%), and *Neisseria* (80% versus

50%) appeared more prevalent among cases compared to controls, respectively. *Neisseria* was nearly eight-fold more abundant in cases versus controls (Table 2 and Figure 2B). (17) (18)

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 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</i> KLE1280 (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 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 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 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 in Lung tumor	<ol style="list-style-type: none"> 1. <i>Burkholderia</i> (1.23x) 	N/A	N/A
	Lower in Lung tumor	N/A	N/A	N/A

Lung cancer lavage compared to lung cancer gargle	Higher in lung cancer lavage	<ol style="list-style-type: none"> 1. Leptotrichia (5.46x) 2. Campylobacter (2.50x) 3. Fusobacterium (1.13x) 4. Neisseria (1.04x) 	<ol style="list-style-type: none"> 1. Rothia dentocariosa (2.71x) 2. Porphyromonas KLE1280 (2.53x) 3. Gemella hemolysans (1.79x) 4. Granulicatella adiacens (1.71x) 5. 	<ol style="list-style-type: none"> 1. Human parainfluenzavirus 3 (45.8x)
	Lower in lung cancer lavage	<ol style="list-style-type: none"> 1. Rothia (0.35x) 2. Atopobium (0.52x) 3. Porphyromonas (0.54x) 4. Streptococcus (0.57x) 5. Prevotella (0.61x) 6. Granulicatella (0.77x) 7. Actinomyces (0.78x) 8. Veillonella (0.93x) 9. [Prevotella] (0.97x) 	<ol style="list-style-type: none"> 1. Veillonella dispar (0.27x) 2. Prevotella pallens (0.47x) 3. Rothia mucilaginosa (0.053x) 4. Neisseria subflava (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 in lung lavage	<ol style="list-style-type: none"> 1. Streptococcus (651.67x) 2. Prevotella (130.63x) 	N/A	N/A
	Lower in lung lavage	<ol style="list-style-type: none"> 1. Burkholderia (0.08x) 	N/A	N/A
Lung cancer gargle compared to lung tumor	Higher in lung gargle	<ol style="list-style-type: none"> 1. Streptococcus (1133.67x) 2. Prevotella (211.00x) 	N/A	N/A
	Lower in lung gargle	<ul style="list-style-type: none"> • 1. N/A 	N/A	N/A

Abbreviations: 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.

Lung cancer lavages tended to be slightly more diverse compared to controls by Chao1 index, though not significant (Table 3 and Figure 2C).^(18, 19) Beta diversity measured by Bray-Curtis dissimilarity was significantly different, though clear separation between cases versus controls was not observed (Table 3 and Figure 2D).

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)	1. S 2. S 3. S	N/A	N/A	N/A	N/A

Notes: 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. Abbreviations: S= significant ($p \leq 0.05$), MS= marginally significant ($p = 0.08$ or below) and 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).

Whole genome shotgun sequencing: Bacterial species appeared similarly prevalent in cases and controls (Figure S1A). 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 and Figure S1B). Several species of *Prevotella* were more abundant in controls compared to cases. Alpha diversity estimates revealed no consistent pattern and, along with beta diversity, was not significantly different between cases and controls (Table 3 and Figure S1C and S1D).

The virome of the tracheal lavages was assessed through WGSS sequencing too, identifying largely similar prevalence and relative abundance between cases and controls (Table 2 and Figure S2A and S2B). Human betaherpesvirus 7 was more abundant in case versus control lavages but was rare. Though not statistically significant, case lavages consistently showed higher viral alpha diversity compared to control lavages (Table 3 and Figure S2C). Beta diversity was not significantly different.

Cases vs. Controls: Oral gargles

16S rRNA gene sequencing: Oral gargles from lung cancer and melanoma patients showed very little difference in terms of prevalence (Figures 1B and 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). In terms of relative abundance, *Streptococcus* and *Prevotella*, (17) were the most abundant genera in oral gargles from both cases and controls (Figure 3B). *Neisseria* was much more abundant in controls compared to cases, while the opposite trend was observed for *Fusobacterium* (almost 2.5-fold higher in cases versus controls). Alpha diversity and beta diversity across all indices showed no significant differences between gargles (Table 3).

Whole genome shotgun sequencing: Streptococcal species, such as *S. infantis* and *S. pseudopneumoniae*, appeared heavily prevalent in all oral gargles (Figure S3A). While *Neisseria subflava* was 2 times more abundant in gargles from lung cancer cases compared to controls and *Rothia dentocariosa* was more abundant in controls versus cases (Table 2 and Figure S3B), overall, bacterial abundance appeared similar among oral gargle samples. Diversity within samples were not significant. Beta diversity was not different by case or control status (Table 3 and Figure S3D). Bacteriophages appeared more abundant in cases, with *Haemophilus* phages HP1 and HP2 being identified in over 14-fold and 7-fold higher abundance compared to controls (Table 2 and Figure S4B). Human-tropic viruses were similarly prevalent between cases and controls. Neither alpha nor beta diversity indices demonstrated any significant differences between cases and controls (Table 3 and Figure S4C and S4D).

Cases: Tumor versus Normal (Non-neoplastic) Lung Tissue

16S rRNA gene sequencing: In terms of prevalence, *Propionibacterium*, *Atopobium*, and *Granulicatella* were identified in at least one tumor specimen but not in normal tissues (Figure 1C and Figure S5A). Conversely, *Actinomyces* was identified in 20% of normal tissue samples but not in tumor tissues. The most abundant genus, albeit rare, in both the tumor and normal tissue was *Burkholderia*, not considering unclassified, though it is slightly more abundant in tumors (Table 2 and Figure S5B). Alpha and beta diversity were not significantly different between tissue types (Table 3 and Figure S5C and S5D), though, normal tissue generally had lower alpha diversity compared to tumor tissue.(20)

Cases: Lavage versus Gargle

16S rRNA gene sequencing: Several genera appeared more prevalent in lavages compared to gargles (Figure 1D and S6A), including *Leptotrichia*, while genera like *Capnocytophaga* were more prevalent in gargles. Despite an overall similarity in relative abundances, the genera *Streptococcus*, *Prevotella*, and *Rothia* were more abundant in gargles compared to lavages, whereas *Leptotrichia* (>5-fold higher abundance) showed the opposite trend. By the Chao1 index, lavages maintained higher diversity compared to gargles (Table 3 and Figure S6C). Bacterial community structures were significantly different between gargles and lavages by both Bray Curtis dissimilarity and unweighted UniFrac distance (p=0.001).

Whole genome shotgun sequencing: *Rothia dentocariosa* was 2.7x more abundant in lavages versus gargles (Table 2 and Figure S7B). *Veillonella dispar* was much more abundant in gargles though all 10 gargles had *G. adiaciens*. Interestingly, *E. coli* was not identified in the top 25 most prevalent species of lavages but was observed in all gargle samples. The relative abundances of bacterial species maintained some notable differences, despite *R. mucilaginosa* and *N. subflava* comprising the two most abundant species in both gargles and lavages. 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 top 25 most abundant species) and *G. adiaciens* (5.1% versus 2.5%) are more abundant in lavages compared to gargles, respectively. 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 and Figure S7).

In terms of viral signatures, prevalence appears different between these two sample types (Figures S8 and S12C). For example, Human parainfluenza virus 3 and respiratory syncytial virus were identified in more lavages (10 and 6, respectively) compared to 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%). Gargles maintained higher relative abundances of *Haemophilus* phages HP1 and HP2 as well as human betaherpesvirus 7 compared to lavages, though Human parainfluenzavirus 3 was more abundant in the latter (Table 2). Alpha diversity by Chao1 was higher in gargles versus lavages (Table 3). Beta diversity was significant across Bray Curtis dissimilarity ($p=0.001$), weighted UniFrac distance ($p=0.004$), and unweighted UniFrac distance.

Cases: Lavage versus Tumor

16S rRNA gene sequencing: The genus *Burkholderia* is more prevalent in tumor tissue compared to lavages (90% versus 40%, respectively), while *Granulicatella*, *Prevotella*, *Atopobium*, and *Rothia* are just a few of the more prevalent genera found in lavages (Figure S9A and S11A). 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). A large number of bacterial taxa significantly differentiated lavages from tumors, including *Granulicatella*, *Leptotrichia*, *Neisseria*, *Prevotella*, and *Rothia*. 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 and Figure S9C). Similarly, all three Beta diversity measures showed that bacterial community structure is significantly different between the sample types ($p=0.001$ across all three indices) (Figure S9D).

Cases: Gargle versus Tumor

16S rRNA gene sequencing: Most genera were more abundant in gargles versus tumor tissue (Figure S10 and S11B). 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 was

significantly different across all three metrics, showing that these sample types maintain largely disparate community structures.

Discussion

By a conservative estimate, there are at least a billion species of bacteria, but only 30,000 are formally named.(21) Of those, less than 2% can be cultured and identified in the laboratory.(22) 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 identification of unculturable organisms.(23) 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.(24)

Although in preclinical models, the bacterial composition of the gut microbiome appears to determine whether there is a response to immune checkpoint inhibitors (ICI), numerous human studies have failed to identify specific species or phyla that are clearly associated with immunotherapy efficacy in any cancer.(24) Clearly other factors such as bacterial-dependent gut metabolite production that modify blood metabolites and immune competence may hold the key to ICI efficacy.(25) Although the taxonomy of gut microbiota has been under intense investigation, only a few studies have focused on the respiratory microbiome and its relationship to lung cancer.

Microbial Biomarkers

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 towards biomarker studies. Therefore, we compared tracheobronchial lavages and oral gargles that were collected from both lung cancer cases and controls. The results demonstrate that there are few significant differences in overall microbial composition using prevalence, abundance and diversity measures in the oral gargles between lung cancer and melanoma patients, indicating the readily available, noninvasively sampled oral gargle microbiome would not likely serve as a lung cancer biomarker.

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 greater 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 of lung cancer versus melanoma lavages were distinct, although 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 these results were not replicated by WGSS.

Abundance denotes the percentage that a specific bacterium contributes to a sample's overall composition. Whereas prevalence refers to the number (percentage) of cases in a specific group in which a bacterium are detected. Interesting trends were observed in abundance and prevalence of the lavages. Most noteworthy was *Granulicatella adiacens* was more prevalent and abundant in cases. This is a well-recognized oral commensal bacterium that has been etiologically linked to endocarditis.(26) We found this bacterium as one of the top 25 most abundant genera in lung cancer lavages,

and it was much higher prevalence appearing in virtually *all* lung cancer tracheal lavages (100%) 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 biomarker of lung cancer.(27) Replication of this finding in our study suggests that this microbe may actually be important to further investigate as a potential diagnostic biomarker, (27) 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* 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 from one another 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*.

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 potentially 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 as a representation of the lavage microbiota.

Alpha diversity refers to the variation (how diverse it is) of bacteria within a single sample, with a higher alpha diversity 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 that preclude 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.(7)

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,(28) and another identified *Enterobacter* in this same sample type associated with worse survival,(19) emphasizing the importance of continued investigation of the lung microbiome in lung cancer. It has already been hypothesized that Enterobacteriaceae, a bacterial family which express 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.(19) Other studies have suggested that some microbiota may opportunistically invade damaged lung epithelium, caused by smoking, and drive tumorigenesis through production of free radicals like ROS/RNS that can damage the TP53 gene.(20) 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.(29) These cytokines appeared to co-occur with tumor progression in the mice.(29) Additional studies are needed to provide substantiated evidence of the mechanistic relationships between the microbiome, the immune system, and lung cancer.

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 somewhat similar microbial signatures between tumor and normal tissue, it may indicate certain microbes from the normal 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, alpha or beta diversity between tumor and normal tissue samples. Interestingly, normal tissue had lower alpha diversity compared to tumor tissue, contrary to that observed between tumor and healthy tissue controls previously.(30) Finally, slight variations in bacterial prevalence were identified: higher prevalence of the genera *Granulicatella* and *Burkholderia* in tumors was observed, as well as higher prevalence of *Neisseria* and *Fusobacterium* in normal tissues.

The genus *Granulicatella* in particular has been found in a previous study to inhabit the tumor microenvironment, and as it becomes increasingly anaerobic there is production of useful metabolites for this genus.(9) In the current study, *Granulicatella* was also identified in higher prevalence not only in tumor and normal tissue but also was more prevalent 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 the oral and sputum samples of lung cancer patients compared to controls.(31) This provides some intriguing preliminary data suggesting a possible carcinogenic role for some bacteria or at least opportunistic inhabitants of the tumor microenvironment, but testing in larger cohort studies is needed.

Overall, tracheal lavages and gargles do not appear to provide a consistent microbial signature for the tumor microbiome either. In fact, 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 a large number of bacterial genera more abundant in lavages, like *Granulicatella* and *Neisseria*, but one genus was more abundant in tumor tissue, namely *Burkholderia*, an important Gram-negative pathogen of lung infections in cystic fibrosis patients(32) and is the causative agent in the life-threatening respiratory illness meliodosis.(33) Similarly, 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.(7) Indices of alpha diversity also showed gargles to be significantly more diverse than tumor tissues, and LEfSe revealed the genus *Burkholderia* to once again be more abundant in tumor tissue versus oral gargles.

Overall, lavages and gargles cannot accurately stand in as proxies of the tumor microbiome given the substantial differences between them. However, the genus *Burkholderia*, in particular, appeared more abundant and prevalent in tumor tissue versus both lavages and gargles, suggesting a potential role in tumorigenesis or at least opportunistic inhabitation of the tumor microenvironment.

Microaspiration

Microaspiration is a common event, occurring in as many as 50% of healthy people,(34) although it is unknown how many have persistent colonization of the tracheobronchial tree with oral commensals. Previous studies

by Segal and associates(35) demonstrated enrichment of oral commensals in the lower airways of normal individuals is associated with increased host inflammatory tone and increase in checkpoint inhibitor markers. This lower airway dysbiotic signature was found by Tsay and colleagues to distinguish between patients with lung cancer and benign lung nodules.(36)

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 occurs to a larger extent in lung cancer patients compared to 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 the recurrence of lung cancer after resection.(28) Tsay and colleagues, using RNA-seq analysis of lower airway samples, found that the supraglottic predominant taxa 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.(36)

Microvirome

While the characterization of the bacterial (and fungal) members of the human microbiome including the respiratory tree has blossomed in the last decade, studies of the viral component are incomplete. Many challenges exist with exploration of the microvirome due to its low biomass. Even the use of high throughput sequencing technologies is 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.(37) Additionally, there is the difficulty with the contaminating human and bacterial DNA and RNA in samples. Finally, the vast majority of reads with WGSS are commonly called “viral dark matter” since the results can’t be annotated into taxonomic categories due to lack of species available in databases.(37) As a result, only a relatively small number of viruses were identified in our study likely because of the low abundance and inability to classify common viral organisms.

The virome of the tracheal lavages on our study was assessed through WGSS sequencing identifying largely similar prevalence and relative abundance between cases and controls in terms of both lavages and gargles. There was a higher abundance of human respiratory syncytial virus in the melanoma versus lung cancer patient lavages, and conversely human betaherpesvirus 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 possibly enter the respiratory tree by cigarette smoking.

Viral signatures in the oral gargles demonstrated bacteriophages targeting *Haemophilus* bacteria were more prevalent in melanoma controls versus lung cancer cases. Human-tropic viruses, such as endogenous retrovirus K and betaherpesvirus 7, were similarly prevalent between cases and controls. Although LEfSe identified several unclassified viral signatures as being 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 highly abundant in cases and controls.

Comparison of viral signatures between the oral gargle and lavages in the cases demonstrated the prevalence appears different between these two sample types. The most prevalent viral signature in lavages was human parainfluenza virus 3 and the human respiratory syncytial virus compared to gargles. However, human gammaherpesvirus 4 was identified more commonly in gargles. LEfSe revealed several viral taxa 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 quite significant differences in viral community structure between gargles and lavages.

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

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. In addition, the small subsample 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. Additionally, a more thorough evaluation of contamination should be implemented in future studies, though our case and control samples were processed in the same way such that contamination should theoretically not result in 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. This will 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 major 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 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,⁽³⁸⁾ and hepatitis B virus,⁽³⁹⁾ demonstrated in our prior pan-microbial array study of biobanked frozen lung cancers.⁽⁴⁰⁾

Ultimately the question arises as to why does a dysbiotic, inflammatory tracheobronchial microbiome appear 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.⁽⁴¹⁾ If a woman is found to have high risk HPV on her pelvic exam, 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.(42) 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. An eubiotic, low diversity, low pH vaginal microbiome, particularly dominated by lactobacillus species, likely help clear HPV infections and are also cytotoxic by secreting bacteriocins that modulate the immune system to inhibit viral activity. However, the dysbiotic, proinflammatory microbiome induces oxidative DNA damage and promotes viral transformation of the cervix by the resident HPV.(43)

Therefore, we might postulate a similar scenario for the consistent finding of a dysbiotic tracheobronchial microbiome in lung cancer patients. If some or all lung cancer is “caused’ by one or more oncogenic viruses such as HPV, (44) bovine leukemia virus,(40) and HTLV-1(40, 45), then 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 a number of future studies to substantiate.

Conclusions

The primary focus of this study was to evaluate the oral and lower airway microbiome compositions of lung cancer cases compared melanoma controls to reveal any differences that may have potential applications towards biomarker studies. Indeed, in this case-control study, we found that bacterial communities of lung cancer versus melanoma lavages were significantly different although 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 potential biomarkers of lung cancer.

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, and the control lavages have a markedly reduced abundance of oral taxa, perhaps suggesting far more microaspiration and inflammation occurs in lung cancer patients. The tumor microbiome shows substantial difference between the lavages and gargles, such that they cannot accurately stand in as proxies of the tumor. 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 as an opportunistic inhabitant of the tumor microenvironment. Finally, 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 meaningful evaluation of the microvirome. 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.

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.

Declarations

FUNDING: This study was funded by a Moffitt Cancer Center Team Science Award granted to Lary A. Robinson, M.D., and Christine M. Pierce, PhD, MPH, Co-Principal Investigators.

COMPETING INTERESTS: The authors declare that they have no competing or potential conflicts of interests.

DATA AVAILABILITY: The data generated or analysed during the current study are available from the corresponding author on reasonable request.

CODE AVAILABILITY: Not applicable.

AUTHORS' CONTRIBUTIONS: SH was involved in analyzing and interpreting the data as well as being a major contributor in writing the manuscript. CP was involved in creating and carrying out the protocol, analyzing and interpreting the data and was a major contributor in writing the manuscript. SP, RT and YK performed the data analysis, bioinformatics and biostatistical evaluation and creation of the tables and figures. LR created the study protocol, performed all the tissue collections and bronchoscopy, was involved in data analysis and interpretation and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE: 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).

CONSENT FOR PUBLICATION: No identifiable patient data is included in this manuscript.

References

1. World Health Organization. Cancer Fact Sheet Geneva, Switzerland 2021 [October 31, 2021]. Available from: <https://www.who.int/news-room/fact-sheets/detail/cancer>.
2. Henley SJ, Ward EM, Scott S, Ma J, Anderson RN, Firth AU, et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. *Cancer*. 2020;126(10):2225-49.
3. American Cancer Society. Cancer Facts and Figures 2020 [Available from: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/estimated-number-deaths-by-sex-and-age-group-2020.pdf>].

4. American Lung Association. Lung Cancer Fact Sheet 2019 [Available from: <https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/resource-library/lung-cancer-fact-sheet>].
5. Mao Y, Yang D, He J, Krasna MJ. Epidemiology of lung cancer. *Surgical Oncology Clinics of North America*. 2016;25(3):439–45.
6. Richards TB, Doria-Rose VP, Soman A, et al. Lung cancer screening inconsistent with U.S. Preventive Services Task Force recommendations. *Am J Prev Med*. 2019;56(1):66–73.
7. Yu G, Gail MH, Consonni D, Carugno M, Humphrys M, Pesatori AC, et al. Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome Biol*. 2016;17(1):163.
8. Moffatt MF, Cookson WO. The lung microbiome in health and disease. *Clinical medicine (London, England)*. 2017;17(6):525–9.
9. Mur LA, Huws SA, Cameron SJ, Lewis PD, Lewis KE. Lung cancer: a new frontier for microbiome research and clinical translation. *Ecancermedicalscience*. 2018;12:866.
10. Centers for Disease Control and Disease Prevention. Guidelines for Disinfection and Sterilization in Healthcare Facilities 2008 [Available from: <https://www.cdc.gov/infectioncontrol/guidelines/disinfection/healthcare-equipment.html>].
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data.. *Bioinformatics*. 2014:btu170.
12. Edgar RC. UCHIME2: improved chimera prediction for amplicon sequencing bioRxiv [Internet]. 2016.
13. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR *Bioinformatics*. 2014;30(5):614–20.
14. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335–6.
15. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database):D590-6.
16. CosmosID Inc. CosmosID Metagenomics Cloud, app.cosmosid.com 2021 [November 30, 2021]. Available from: <http://www.cosmosid.com>.
17. Huffnagle GB, Dickson RP, Lukas NW. The respiratory tract microbiome and lung inflammation: a two-way street. *Mucosal Immunology*. 2016;10:299–306.
18. Lee SH, Sung JY, Yong D, Chun J, Kim SY, Song JH, et al. Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. *Lung cancer (Amsterdam, Netherlands)*. 2016;102:89–95.
19. Gomes S, Cavadas B, Ferreira JC, Marques PI, Monteiro C, Sucena M, et al. Profiling of lung microbiota discloses differences in adenocarcinoma and squamous cell carcinoma. *Scientific Reports*. 2019;9(1):12838.
20. Greathouse KL, White JR, Vargas AJ, Bliskovsky VV, Beck JA, von Muhlinen N, et al. Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol*. 2018;19(1):123.
21. Dykhuizen D. Species numbers in bacteria. *Proc Calif Acad Sci*. 2005;56:62–71.
22. Wade W. Unculturable bacteria—the uncharacterized organism that cause oral infections. *J Royal Soc Med*. 2001;95:81–3.
23. Wilson MJ, Weightman AJ, Wade WG. Applications of molecular ecology in the characterization of uncultured microorganisms associated with human disease. *Rev Medical Microbiology*. 1997;8:91–101.
24. Pierrard J, Seront E. Impact of the gut microbiome on immune checkpoint inhibitor efficacy—a systematic review. *Current Oncology*. 2019;26:395–403.

25. Tang J. Microbial metabolomics. *Curr Genomics*. 2011;12:391–403.
26. Cincotta MC, Coffey KC, Moonah SN, Uppal D, Hughes MA. Case Report of *Granulicatella adiacens* as a Cause of Bacterascites. *Case Reports in Infectious Diseases*. 2015;2015:5.
27. Cameron SJS, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, et al. A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. *PloS one*. 2017;12(5):e0177062.
28. Patnaik SK, Cortes EG, Kannisto ED, Punnanitnont A, Dhillon SS, Liu S, et al. Lower airway bacterial microbiome may influence recurrence after resection of early-stage non-small cell lung cancer. *The Journal of thoracic and cardiovascular surgery*. 2020.
29. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, et al. Commensal Microbiota Promote Lung Cancer Development via $\gamma\delta$ T Cells. *Cell*. 2019;176(5):998-1013.e16.
30. Mao Q, Jiang F, Yin R, Wang J, Wenjie X, Dong G, et al. Interplay between the lung microbiome and lung cancer. *Cancer Letters*. 2018;415:40–8.
31. Hosgood HD, Sapkota AR, Rothman N, Rohan T, Hu W, Xu H, et al. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. *Environmental and Molecular Mutagenesis*. 2014;55:643–51.
32. Fauroux B, Hart N, Belfar s, Boule M, Tillous-Borde I, Bonnet D, et al. *Burkholderia cepacia* is associated with pulmonary hypertension and increased mortality among cystic fibrosis patients. *J Clin Microbiology*. 2004;42:5537–41.
33. Wiersinga W, Virk H, A. T, et.al. Melioidosis. *Nat Rev Dis Primers*. 2018;4:17107.
34. Gleeson K, Eggli DF, Maxwell SL. Quantitative aspiration during sleep in normal subjects. *Chest*. 1997;111:1266–72.
35. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nature Microbiol*. 2016;1:16031
36. Tsay JC, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P. Airway microbiota is associated with upregulation of the PI3K pathway in lung cancer. *Am J Respiratory Critical Care Med*. 2018;198:1188–98.
37. Abbas A. The human lung viral microbiome in health and disease. Philadelphia, PA.: University of Pennsylvania; 2019.
38. Srinivasan M, Taioli E, Ragin CC. Human papillomavirus type 16 and 18 in primary lung cancers—a meta-analysis. *Carcinogenesis*. 2009;30:1722–8.
39. Sundquist K, Sundquist J, Ji L. Risk of hepatocellular carcinoma and cancers at other sites among patients diagnosed with chronic hepatitis B virus infection in Sweden. *J Med Virol*. 2014;86:18–22.
40. Robinson LA, Jaing C, Campbell CP, Magliocco A, Xiong Y, Magliocco G, et al. Molecular evidence of viral DNA in non-small cell lung cancer (NSCLC) and non-neoplastic lung. *Brit J Cancer*. 2016;115:497–504.
41. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*. 1999;189:12–9.
42. Rodriguez AC, Schiffman M, Herrero R, Hildesheim A, Bratti C, Sherman ME, et al. Longitudinal study of human papillomavirus persistence and cervical intraepithelial neoplasia grade 2/3: Critical role of duration of infection.. *J Natl Cancer Inst*. 2010;102:315–24.
43. Mitra A, MacIntyre DA, Marchesi JR, Lee yS, Bennett PR, Kyrgiou M. The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next? *Microbiome*. 2016;4:58.

44. Xiong W, Xu Q, Li X, Xiao R, Cai L, He F. The association between human papillomavirus infection and lung cancer: a system review and meta-analysis. *Oncotarget*. 2017;8:96419–32.
45. Nomori H, Mori T, Iyama K, Okamoto T, Kamakura M. Risk of bronchioloalveolar carcinoma in patients with human T-cell lymphotropic virus type 1 (HTLV-I): case-control study results. *Ann Thorac Cardiovasc Surg*. 2011;17:19–23.

Figures



Figure 1

Venn diagram comparison of bacterial genera prevalence between cases and controls and between different sample types among controls, as determined by the top 25 most prevalent bacterial genera identified in 16S rRNA sequencing analyses. Bacterial genera in the center of the Venn diagram are found in both groups. **A**. Comparison of bacterial prevalence between lung cancer and melanoma tracheobronchial lavages (TBL). **B**. Comparison of bacterial prevalence between lung cancer and melanoma oral gargles. **C**. Comparison of bacterial prevalence between tumor and normal tissue from lung cancer cases. **D**. Comparison of bacterial prevalence between tracheobronchial lavages and oral gargles from lung cancer cases.

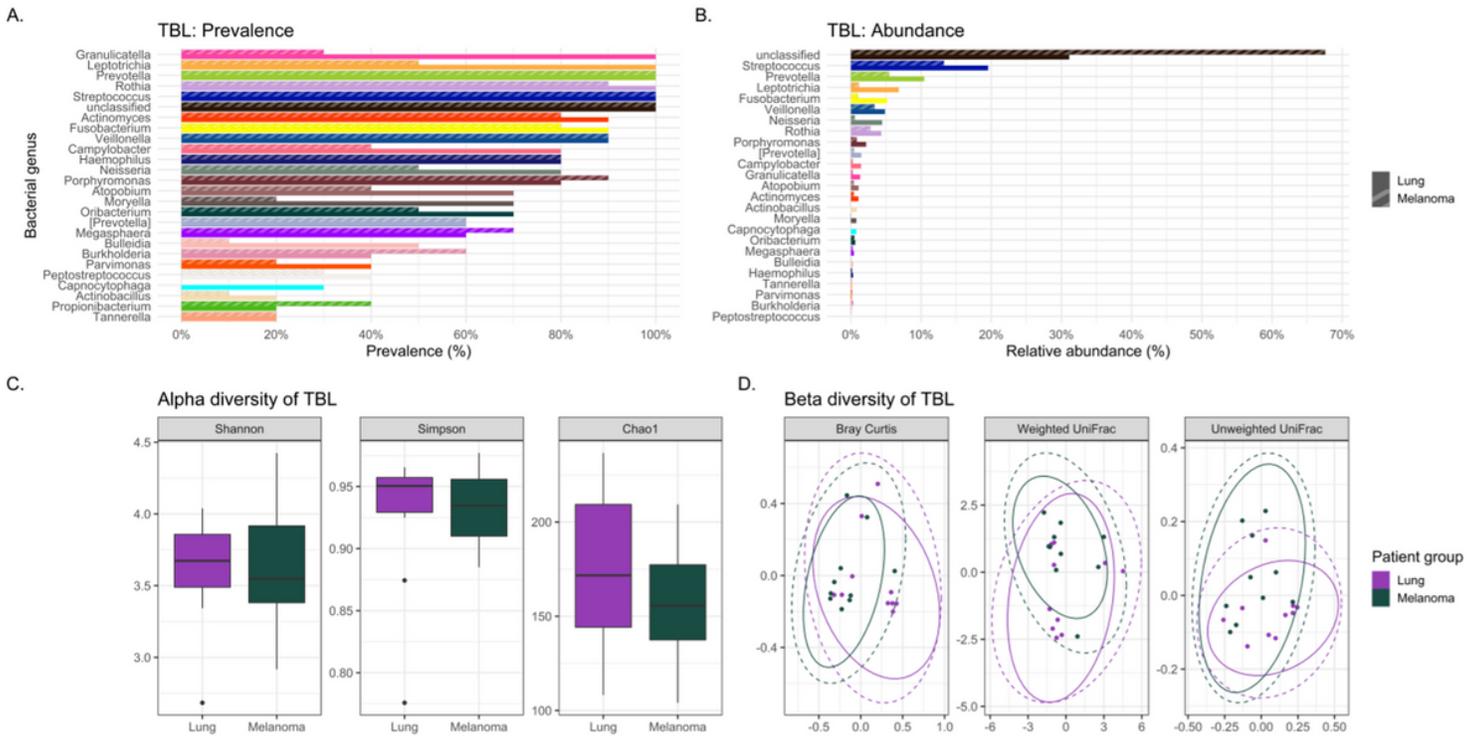


Figure 2

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

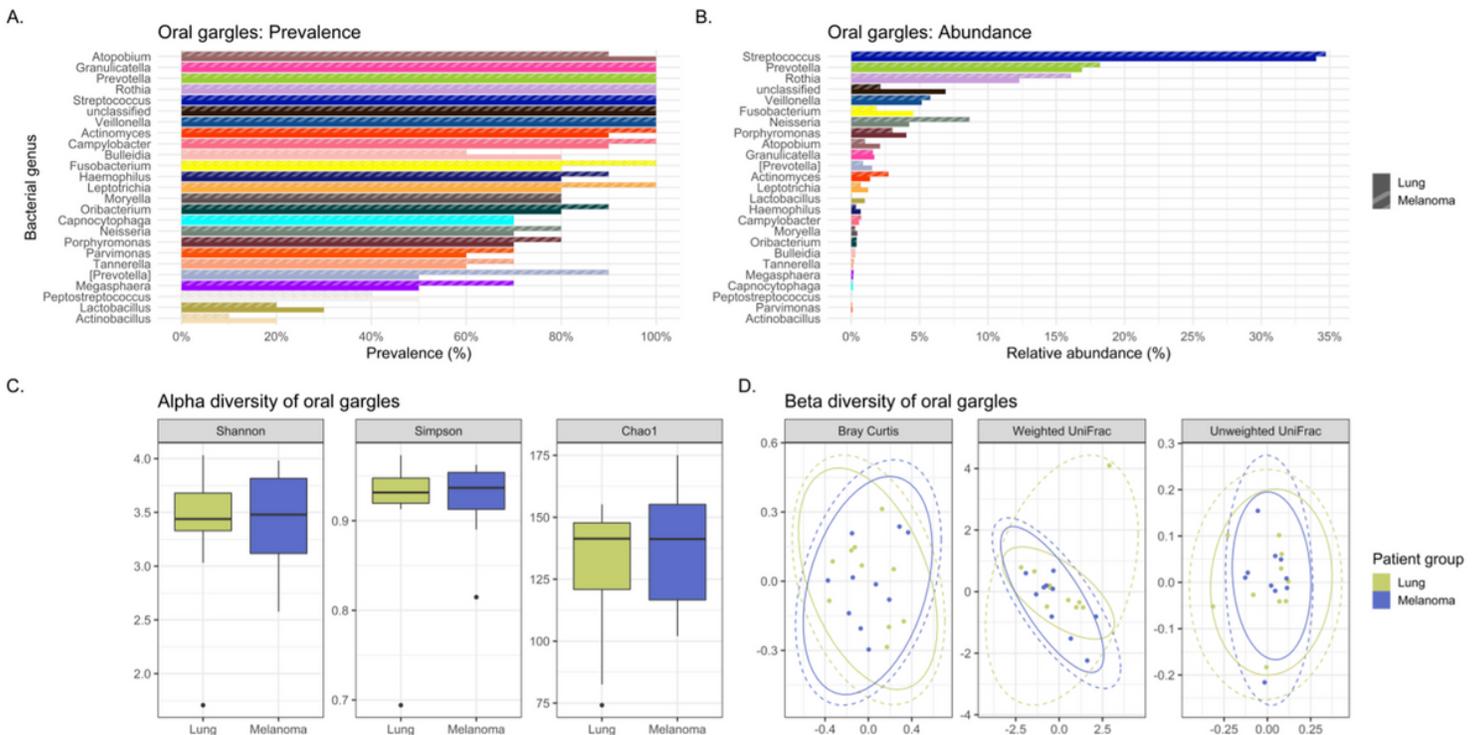


Figure 3

Comparing the bacteriomes, assessed by 16S rRNA gene sequencing, of oral gargles of lung cancer cases and melanoma controls. **A.** Comparison of the prevalence of bacterial genera in lung cancer versus melanoma control oral gargles. **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 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 Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.png](#)
- [FigureS2.png](#)
- [FigureS3.png](#)
- [FigureS4.png](#)
- [FigureS5.png](#)
- [FigureS6.png](#)
- [FigureS7.png](#)
- [FigureS8.png](#)
- [FigureS9.png](#)
- [FigureS10.png](#)
- [FigureS11.png](#)
- [FigureS12.png](#)