Open Access

OBM Genetics

Original Research

Comparison of Sputum and Oropharyngeal Microbiome Compositions in Patients with Non-Small Cell Lung Cancer

Elizaveta Baranova [1], Vladimir Druzhinin [1, 6, *], Ludmila Matskova [2, 3], Pavel Demenkov [4], Valentin Volobaev [5], Alexey Larionov [1]

1. Kemerovo State University, 650000, Krasnaya Str. 6, Kemerovo, Russia; E-Mails: baranovaelisabeth@gmail.com; druzhinin_vladim@mail.ru; alekseylarionov09@gmail.com
2. Institute of Living Systems Immanuel Kant Baltic Federal University, 236016, Alexandra Nevsky Str. 14, Kaliningrad, Russia; E-Mail: liudmila.matskova@ki.se
3. Department of Microbiology, Tumor Biology and Cell Biology (MTC), Stockholm, 171 65, Sweden
4. Federal Research Center Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, 630090, Pirogov Str. 1, Novosibirsk, Russia; E-Mail: demps@bionet.nsc.ru
5. Sirius University of Science and Technology, 3543401, Olympic Ave 1, Sochi, Russia; E-Mail: volobaev.vp@gmail.com
6. Kemerovo State Medical University, 650036, Voroshilov Str. 22A, Kemerovo, Russia

* **Correspondence:** Vladimir Druzhinin; E-Mail: druzhinin_vladim@mail.ru

Academic Editor: Yan Sanders

OBM Genetics
2022, volume 6, issue 4
doi:10.21926/obm.genet.2204169

Received: August 01, 2022
Accepted: October 24, 2022
Published: November 07, 2022

Abstract

Recent findings indicate that the microbiota is involved in the development of lung cancer by inducing inflammatory responses and generating genome damage. This study aimed to compare sputum microbiomes from the mouth and oropharynx in non-small cell lung carcinoma (NSCLC) patients. A second goal was to search for bacterial taxonomic units that behave differently in the microbiome of NSCLC patients and healthy subjects. In the study, the taxonomic composition of the sputum and oropharyngeal microbiomes of 23 male patients with untreated NSCLC and 20 healthy subjects were compared. Next-generation sequencing of bacterial 16S rRNA genes was used to determine the taxonomic composition

of the respiratory microbiome. Using the Kruskal-Wallis test, increased alpha diversity was observed in the sputum microbiome compared to that of the oropharynx, but this was evident only in NSCLC patients and not in healthy subjects. Using the Robust Aitchison PCA test, differences in beta diversity were found between sputum and oropharynx samples, and these differences were significant both for NSCLC patients ($p = 0.045$) and healthy controls ($p = 0.009$). However, no significant statistical differences were detected using the Robust Aitchison PCA when only comparing oropharyngeal samples from NSCLC patients and controls, nor when comparing sputum samples alone. Analysis of differences in the relative percentage of individual bacterial taxa using the Mann-Whitney U-test, and taking into account the FDR correction, showed an increase in the genus *Rothia* in oropharyngeal samples of NSCLC patients, as compared to control subjects (4.98 ± 6.33 vs 2.21 ± 6.28; $p = 0.0008$). However, linear discriminant analysis using LefSe did not show *Rothia* as a differentially regulated feature between NSCLC and controls in the oropharynx. Thus, more research is needed to identify possible bacterial NSCLC biomarkers in the oropharynx.

Keywords
Non-small cell lung cancer; 16S ribosomal RNA gene analysis; oropharyngeal microbiome; sputum microbiome; taxonomic composition; *Rothia*

1. Introduction

Lung cancer (LC), is one of the most common cancers in the world and the leading cause of cancer deaths among men [1]. Early detection is critical for reducing morbidity and mortality from LC. The use of modern approaches is crucial for detecting the early stages of the disease. Approximately 80% of LC cases are associated with exposure to tobacco, but LC develops only in 15% of smokers during their lifetime. This highlights the role of genetic and environmental susceptibility factors in modulating the risk of LC developing [2]. Recently, growing attention has been paid to the possible role of the bacterial microbiome in the development and progression of LC. With the development of metagenomic studies, it was recently shown that the taxonomic composition of the bacterial microbiome of the respiratory tract can differ significantly in LC patients and healthy donors [3-6]. These studies show that certain bacterial taxa as well as dysbiosis of the respiratory microbiota, correlate with the development of LC and suggest that alpha diversity (richness and uniform distribution of taxa in samples) is significantly lower in samples from cancerous tissues compared to healthy tissues, while community similarity (beta diversity) varies widely [7]. Thus, the respiratory microbiome can be considered an important diagnostic and prognostic indicator of LC. However, the search for operational taxonomic units significantly associated with the risk of LC development and progression is far from complete. Several previous studies have shown that changes in the number of specific taxa in the microbiota of the lung tissue, bronchoalveolar lavage fluid (BALF), sputum, and saliva samples may be associated with LC, but these studies have been largely inconsistent with respect to specific taxa [8-14]. In addition, since the microbiota in different sites of the respiratory tract are known to

have different taxonomic compositions, it can be assumed that their bacterial biomarkers in LC will be different.

In this study, we tested for the first time the composition of the oropharyngeal microbiome in patients with non-small cell lung carcinoma (NSCLC) and compared it to the sputum microbiome of the same patients, and to that of healthy subjects using 16S metagenomic sequencing.

Thus, the aim of this study was firstly to compare the microbiome of the mouth and oropharynx sputum from NSCLC patients. The second goal was to search for operational taxonomic units that have a different representation in the microbiome of NSCLC patients and healthy subjects. Understanding the relationship between the microbiota of the mouth, larynx and oropharynx should help in establishing reliable testing procedures, thus facilitating early diagnosis of NSCLC. This type of study will contribute to the creation of a complete picture of bacterial colonization in all possible niches in the human body and a better understanding of the interaction of different microbiota and their effects on the human body.

2. Materials and Methods

2.1 Cohort Information

The composition of the sputum and oropharyngeal bacterial microbiome was studied in 23 newly-diagnosed NSCLC male patients (average age 59.3 ± 7.4 years) who were admitted to the Kemerovo Regional Oncology Center (Kemerovo, Russian Federation) and 20 healthy male donors (average age 51.5 ± 6.8 years) who were residents of Kemerovo. There were differences in the mean age between patients and controls ($p < 0.05$). Among NSCLC patients, 61% were active smokers, compared to 50% in the control subjects. For NSCLC patients, the results of clinical and histological analyses were additionally considered to determine the stage of the disease following the TNM classification [15]. Accordingly, 7 patients (30.4%) had stage I-II, and 16 patients (69.6%) had stage III-IV. Information regarding NSCLC and control subjects is summarized in Table 1. A questionnaire was filled out for each survey participant, containing information on place and date of birth, living environment, occupation, exposure to occupational hazards, health status, dietary habits, and intake of medications (use of antibiotics at least four weeks before sampling), X-ray procedures, smoking and drinking status. Inclusion criteria were adult male aged ≥40 years, willing to participate in the study, donate sputum and sign a written informed consent. Exclusion criteria were any acute or chronic condition that would limit the ability of the patient to participate in the study, use of antibiotics within 4 weeks before collection, failure to obtain a sputum sample, or refusal to give informed consent.

Table 1 Characteristics of the study cohorts.

Variables	NSCLC patients, n = 23	Healthy Kemerovo residents (controls), n = 20
Age (years) (mean ± SD)	59.3 ± 7.4*	51.5 ± 6.8
Living environment (%):		
City	78.0	85.0
Village	22.0	15.0

Smoking status (%):		
Smokers	61.0	50.0
Non-smokers	39.0	50.0
Alcohol status (%):		
Non-drinker	13.0	15.0
Rare drinker (once/2 weeks)	39.0	15.0
Medium drinker (once/week)	48.0	70.0
Diet (%):		
Vegetarian	0.0	0.0
Non-vegetarian	100.0	100.0
Histopathological status (%):		
Squamous cell carcinoma	39.0	-
Adenocarcinoma	22.0	
Large cell carcinoma	22.0	
Small cell carcinoma	13.0	
Undifferentiated	4.0	
TNM[#] (%):		
I, II	30.4	-
III, IV	69.6	
Distant metastasis (%):		
Yes	17.4	-
No	82.6	
Tumor localization (%):		
Central lung cancer	30.4	
Peripheral lung cancer	43.5	-
Mixed lung cancer	8.7	
Bronchial cancer	17.4	

* $p < 0.05$ in comparison with controls; [#] tumor-node-metastasis.

2.2 Ethics Approval and Consent to Participate

All procedures followed the ethical standards of the Helsinki Declaration (1964, amended 2008) of the World Medical Association. All participants (NSCLC and controls) were informed about the aim, methodology and possible risks of the study, informed consent was signed by each donor. The design of this study was approved by the Ethics Committee of the Kemerovo State University (PROTOCOL CODE № 17/2021; 05.04.2021).

2.3 Sample Collection, Processing and Storage

To analyze the composition of the microbiome of the respiratory tract, sputum and oropharyngeal samples obtained from NSCLC patients and control subjects were used. The oropharyngeal and sputum samples were taken in parallel from each enrolled participant before clinical diagnosis and before treatment. Samples were collected on the first day of hospitalization. Before sputum and oropharyngeal sample collection, patients were asked to rinse their mouth. Sputum samples were collected non-invasively through participant-induced coughing (i.e., without

induction). Giemsa-stained cytological slide microscopy was used to test random sputum samples to confirm the presence of columnar airway epithelial cells. Oropharyngeal samples were collected by a physician after applying disposable sterile cotton sampling swabs to the posterior pharynx, sidewalls, and crypts of the tonsil and wiping three times in a rotating manner. Then, the cotton swab was placed into a swab preservation tube. The obtained samples were immediately placed in sterile plastic vials and frozen at -20°C. Frozen samples were transported to the laboratory and stored at -80°C.

2.4 DNA Extraction, 16S rRNA Gene Amplification and Sequencing

DNA extraction, 16S rRNA gene amplification and sequencing were performed as described early [16].

2.5 Taxonomy Quantification Using 16S rRNA Gene Sequences and Statistical Methods

The resulting sequence data was processed using the program QIIME2 [17, 18]. A quality check was carried out and a sequence library was generated. The sequences were combined into operational taxonomic units (OTUs) based on a 99% nucleotide similarity threshold using the Greengenes reference sequence library (versions 13–8) and SILVA (version 132), followed by the removal of singletons (OTUs containing only one sequence). The total diversity of prokaryotic communities (alpha diversity) of sputum and oropharyngeal was estimated by the number of allocated OTU (analog of species richness) and Shannon indices (H = Σpi ln pi, pi – part of i-sh species in a community). When calculating sample diversity indices, 328 sequences were normalized (the minimum number of received sequences per sample). The variation in the structure of the bacterial community in different samples (beta diversity) was analyzed using the Jackard index, and the difference between communities was estimated by UniFrac method [19], a method common in microbial ecology, based on the phylogenetic relationships of the presented taxa, and Robust Aitchison principal components analysis (PCA). The Log-ratio PCA compositional biplot constructed by the DEICODE was used [20]. Additionally, linear discriminant analysis (LDA) effect size (LEfSe) as well as ANCOMBC analyses were used to normalize the observed microbial abundance data [21, 22].

In addition, to assess the significance of differences in the relative percentage of individual bacterial taxa in sputum and oropharyngeal samples, the Mann-Whitney U test was used. Spearman's correlation coefficient was used to calculate correlations. The False Discovery Rate (FDR) correction was used to assess the significance of differences in the relative percentages of individual bacterial taxa considering multiple comparisons. Logistic regression was performed using the GLM function and a binomial family generalized linear model in R. For categorical data, dummy variables were created and tested for each individual factor level in a univariate GLM analysis. Models were adjusted for age, smoking status, drinking status, and living environment. Calculations were performed using the software package STATISTICA.10.

3. Results

In the present study, we compared the composition of the bacterial microbiomes in the sputum and oropharynx of 23 patients with NSCLC and 20 healthy male donors. We used a large-scale

approach for sequencing the V3 – V4 region of the 16S rRNA of bacterial genomes purified from sputum and oropharyngeal samples from the compared groups in the study.

In the NSCLC group, the average number of analyzed sequences in the sputum was 8655 (467, 26154) and 22245 (908, 324170) in the oropharynx. In the healthy control group, the average number of analyzed sputum sequences was 8698 (712, 22653) and 18602 (591, 52101) from the oropharynx.

3.1 Comparison of Diversity and Taxonomy in Oropharyngeal and Sputum Specimens

We identified a total of 9 bacterial phyla with relative frequencies above 0.1%. In our dataset, the prevailing phyla both in the sputum and oropharynx microbiomes were *Firmicutes* and *Bacteroidetes*, which together accounted for more than 70% of the total microbiota (Figure 1). In the NSCLC patient group, the Kruskal-Wallis test showed a statistically significant difference in sputum and oropharynx samples. In the Shannon diversity index, from 1.66 to 7.25, H = 6.777, p = 0.009 and in the Pielou's evenness index, from 0.43 to 0.95, J = 6.663, p = 0.009. For the control group, statistical significance was shown only for the Pielou evenness index from 0.7 to 0.96, J = 13, p = 0.0003 (Figure 2).

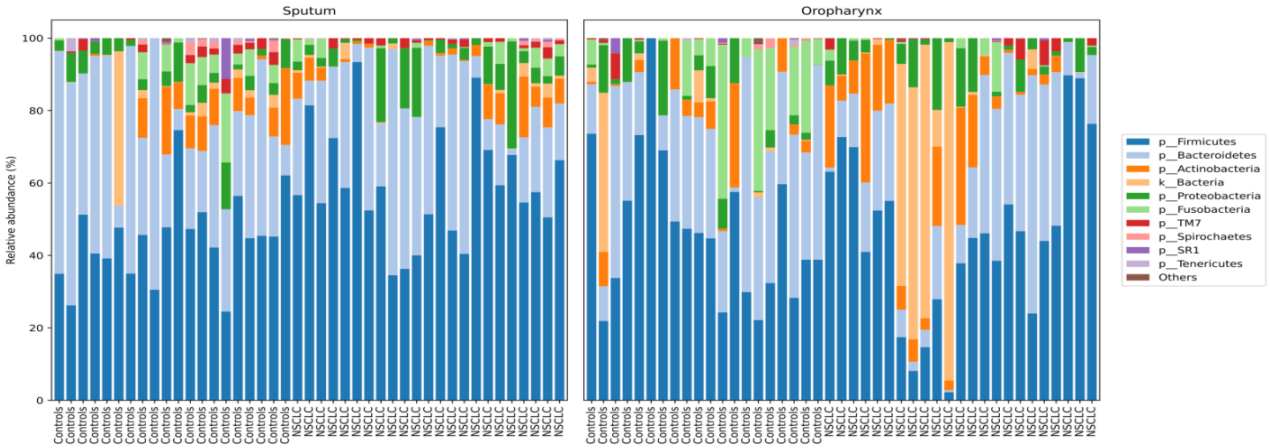

Figure 1 Taxonomical composition (at the phyla level) of sputum and oropharyngeal microbiomes.

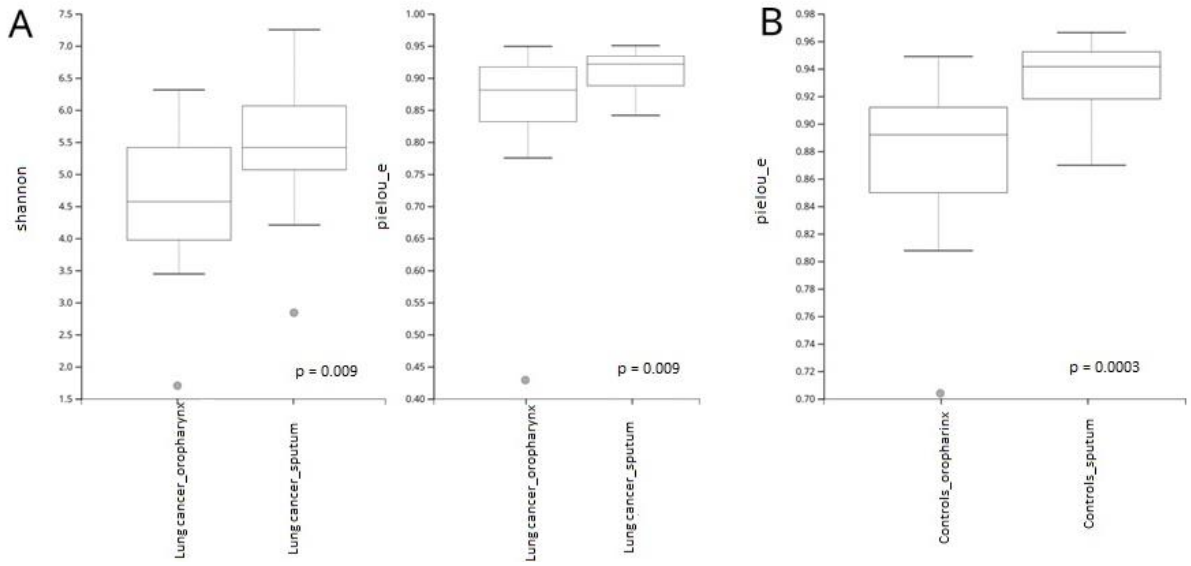

Figure 2 A) Shannon diversity index and *Pielou's evenness* index for the sputum and oropharynx bacterial microbiome in the NSCLC patients. B) *Pielou's evenness* index for the sputum and oropharynx bacterial microbiome in the healthy controls.

Differences in the structure of bacterial communities in sputum and oropharyngeal samples of NSCLC patients are shown in Figure 3. Differences in beta diversity revealed using the Robust Aitchison PCA analysis were found between sputum and oropharynx samples from both NSCLC patients (p = 0.045) and healthy controls (p = 0.009).

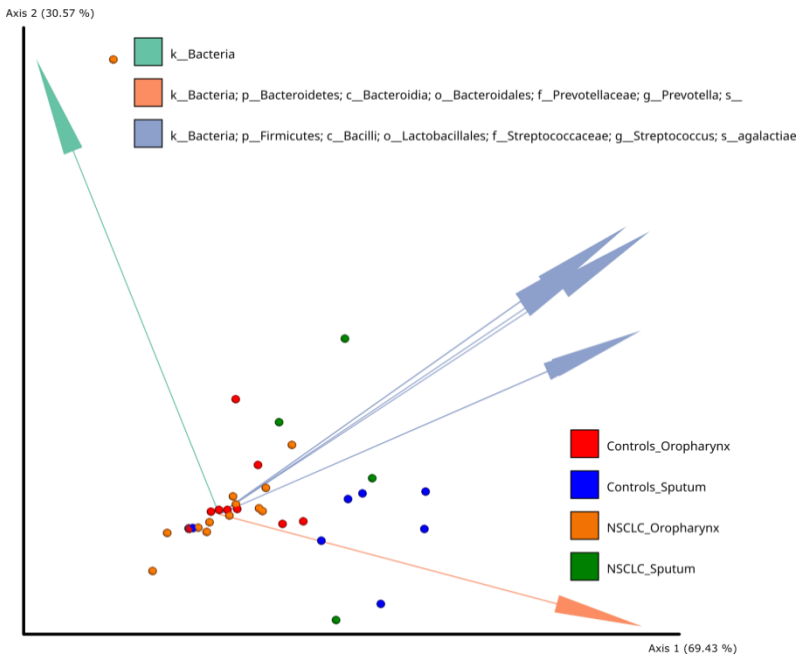

Figure 3 Log-ratio principal components analysis compositional biplot constructed by the DEICODE.

Table 2 summarizes the results of a comparison of the relative abundance of bacterial phyla in sputum and oropharyngeal samples from NSCLC patients and the control group. In general, the

taxonomic composition of sputum and oropharynx microbiomes (at the phylum level) did not show statistically significant differences, either in healthy donors or in patients. The representation of *Bacteroidetes*, *Spirochaetes* and *Tenericutes* in the sputum microbiome from NSCLC patients was significantly increased, but not statistically significant, taking into account the FDR correction (Table 2).

Table 2 Comparison of the occurrence of the main types and genera of bacteria in the sputum and oropharynx of NSCLC patients and healthy donors.

Types/Genus	NSCLC patients		P	Controls		P
	Sputum, n = 23	Oropharynx, n = 23		Sputum, n = 20	Oropharynx, n = 20	
Firmicutes	53.03 ± 13.16	43.33 ± 21.22	0.14	44.99 ± 12.20	48.66 ± 17.74	0.49
Bacteroidetes	23.83 ± 11.54	16.50 ± 11.18	0.03	28.59 ± 11.34	24.15 ± 14.34	0.19
Actinobacteria	8.15 ± 5.44	14.47 ± 10.78	0.05	6.67 ± 5.22	7.18 ± 7.57	0.87
Proteobacteria	7.77 ± 7.60	7.25 ± 6.68	0.85	9.08 ± 12.08	5.98 ± 5.30	0.46
Fusobacteria	5.19 ± 5.44	2.99 ± 3.91	0.11	6.98 ± 3.99	7.70 ± 9.78	0.42
TM7	1.57 ± 1.79	1.19 ± 1.9	0.12	1.84 ± 1.59	1.51 ± 2.42	0.17
Spirochaetes	0.26 ± 0.45	0.03 ± 0.11	0.03	0.79 ± 1.22	0.55 ± 1.17	0.09
SR1	0.08 ± 0.20	0.00	0.08	0.87 ± 2.46	0.20 ± 0.62	0.09
Tenericutes	0.07 ± 0.16	0.01 ± 0.05	0.05	0.34 ± 0.97	0.08 ± 0.23	0.11
Streptococcus	31.99 ± 18.26	21.12 ± 17.22	0.04	19.51 ± 9.20	15.19 ± 12.56	0.08
Alloprevotella	3.83 ± 3.03	2.12 ± 2.59	0.03	2.93 ± 4.38	1.69 ± 1.99	0.45
Fusobacterium	1.49 ± 1.42	0.82 ± 1.57	0.03	2.37 ± 2.87	5.78 ± 9.67	0.85
Gemella	2.69 ± 1.98	1.60 ± 1.59	0.05	1.97 ± 2.46	1.07 ± 1.99	0.03
Rothia	1.89 ± 1.87	4.78 ± 6.49	0.04	1.87 ± 2.42	2.34 ± 6.63	0.07
Bacillus	2.70 ± 1.93	1.61 ± 1.60	0.04	1.83 ± 2.12	1.10 ± 2.08	0.02
Prevotella (f.Paraprevotellacacea)	1.58 ± 1.41	0.63 ± 1.13	0.008	1.40 ± 2.29	0.97 ± 1.17	0.97
Campylobacter	0.72 ± 0.76	0.58 ± 1.02	0.04	0.76 ± 0.79	1.18 ± 1.28	0.53
Capnocytophaga	2.40 ± 5.84	0.56 ± 1.5	0.02	0.68 ± 0.77	0.31 ± 0.99	0.0006*
Treponema	0.24 ± 0.38	0.02 ± 0.06	0.01	0.64 ± 0.94	0.53 ± 1.07	0.31
Macellibacteroides	0.64 ± 1.99	1.36 ± 2.22	0.35	0.45 ± 0.91	2.74 ± 3.37	0.003*
Mycoplasma	0.09 ± 0.16	0.01 ± 0.05	0.02	0.39 ± 0.98	0.10 ± 0.23	0.22
Bergeyella	0.25 ± 0.33	0.07 ± 0.17	0.006	0.29 ± 0.87	0.35 ± 0.90	0.53

Note: * P value is lesser than FDR corrected P value.

At the taxonomic levels of genus and species, 84 genera and 33 species of bacteria were identified. A comparison of the taxonomic composition of the sputum and oropharynx microbiome at the genus and species levels (only those genus and species with an average

abundance of more than 0.1% are presented) for NSCLC patients and controls is presented in Table 2 and Table 3. The data in Table 2 showed that there were no significant differences in the relative abundance of the 44 bacterial genera analyzed when the microbiomes from sputum and oropharyngeal samples from NSCLC patients were compared.

Table 3 Comparison of the occurrence of the main species of bacteria in the sputum and oropharynx of NSCLC patients and healthy donors.

Bacterial species	NSCLC patients		P	Controls		P
	Sputum, n = 23	Oropharynx, n = 23		Sputum, n = 20	Oropharynx, n = 20	
Leptotrichia sp. oral clone GT018	1.39 ± 1.43	0.54 ± 1.12	0.006*	1.22 ± 1.34	0.89 ± 1.33	0.48
Bergeyella sp. AF14	0.23 ± 0.72	0.09 ± 0.33	0.002*	0.31 ± 0.88	0.27 ± 0.51	0.34
Bergeyella zoohelcum	0.27 ± 0.32	0.06 ± 0.16	0.001*	0.29 ± 0.87	0.28 ± 0.72	0.53
Prevotella Tannarae	0.50 ± 0.84	0.13 ± 0.44	0.01*	0.68 ± 1.29	0.44 ± 0.87	0.48
Rothia dentocariosa ATCC 17931	0.40 ± 0.67	0.46 ± 1.02	0.83	0.32 ± 0.49	0.00	0.002*
Filifactor alocis ATCC 35896	0.07 ± 0.16	0.30 ± 1.07	0.86	0.33 ± 0.46	0.02 ± 0.1	0.003*

Note: * P value is lesser than FDR corrected P value.

Interestingly, the data in Table 2 showed a significant increase in the number of representatives of the genus *Capnocytophaga* in the sputum microbiome of healthy subjects when compared to the microbiome of oropharynx samples (0.68 ± 0.77 vs 0.31 ± 0.99; P = 0.0006), and a significant decrease in *Macellibacteroides* (0.45 ± 0.91 vs 2.74 ± 3.37; P = 0.003). In addition, the data in Table 3 showed that in the sputum microbiome of healthy subjects, the contents of *Rothia dentocariosa ATCC 17931* (0.32 ± 0.49 vs 0; P = 0.002) and *Filifactor alocis ATCC 35896* (0.33 ± 0.46 vs 0.02 ± 0.1; P = 0.003) were increased, as compared to the microbiome of the oropharynx.

Although we did not observe any differences in the sputum and oropharyngeal microbiomes of patients at the genus level, differences were noted at the species level (Table 3). *Leptotrichia sp. oral clone GT018* (1.39 ± 1.43 vs 0.54 ± 1.12; P = 0.006); *Bergeyella sp. AF14* (0.23 ± 0.72 vs 0.09 ± 0.33; P = 0.002) and *Bergeyella zoohelcum* (0.5 ± 0.84 vs 0.13 ± 0.44; P = 0.01), were all significantly overrepresented in the sputum microbiome of NSCLC patients as compared to their oropharyngeal microbiome. In this study, we found no specific association between any bacterial taxon in the sputum or oropharynx and the age of the NSCLC patients. However, in the sputum of healthy controls, the representation of *Streptococcus* (P = 0.04) and *Treponema* (P = 0.04) was positively associated with age, while the presence of *Campylobacter* (P = 0.007) was inversely correlated with increasing age. While no such associations were found in oropharyngeal samples from the control donors.

In addition, the influence of smoking status on the microbiota composition in patients with NSCLC and control subjects was studied. In the oropharyngeal microbiome of NSCLC patients, we found no significant difference in the bacterial genera or species between smokers and nonsmokers. In the sputum of NSCLC smokers, a significant decrease in the representation of *Streptococcus agalactiae* (22.66 ± 14.14 vs 41.41 ± 18.72; P = 0.03) and a significant increase in the number of *Dialister* (0.49 ± 0.56 vs 0.001 ± 0.002; P = 0.01) were found in comparison with nonsmokers. However, the representation of *Porphyromonas* (0.87 ± 1.15 vs 4.42 ± 3.48; P = 0.01), *Streptobacillus* (0.9 ± 1.43 vs 2.33 ± 1.66; P = 0.03) and *Bergeyella* (0.05 ± 0.15 vs 0.6 ± 1.24; P = 0.03) was decreased in the oropharynx of smokers in the control group.

A comparison of the microbiome composition between patients with different histopathological status, NSCLC stages (I-II and III-IV), as well as between subgroups with different localization of the primary tumor, revealed no differences.

3.2 Comparison of Microbiome Diversity and Taxonomy in NSCLC Patients and Controls

Next, we compared the microbiota biodiversity in sputum and oropharyngeal samples between patients and healthy donors. There was no significant difference in alpha diversity indexes in sputum and oropharynx between NSCLC patients and controls.

Figure 1 shows the difference in abundance of the microflora in oropharyngeal samples of control and NSCLC patients. However, the Robust Aitchison PCA test results indicated no significant statistical differences in beta diversity in oropharyngeal samples as well as in sputum samples from NSCLC patients and controls (Figure 3).

Next, we applied differential abundance testing methods to find the most significant features that distinguish data from each group. Using the LEfSe method, we detected an increased abundance of some taxa in oropharyngeal samples from the control group as compared to the NSCLC patients (Figure 4). In sputum samples, the LEfSe method revealed an increased abundance of some taxa both in the NSCLC group (green) and in the control group (red) (Figure 5). While analysis of data from the same sputum samples by the ANCOMBC method did not reveal significant differences between the NSCLC and control groups. In the oropharynx, there was an increase in the absolute abundance of *Streptococcus agalactiae* (beta = 2.1, W = 4.08, SE = 0.5, q-value = 0.004) and of *Prevotella* (beta = 1.5, W = 3.58, SE = 0.42, q-value = 0.03) in the NSCLC group relative to the control.

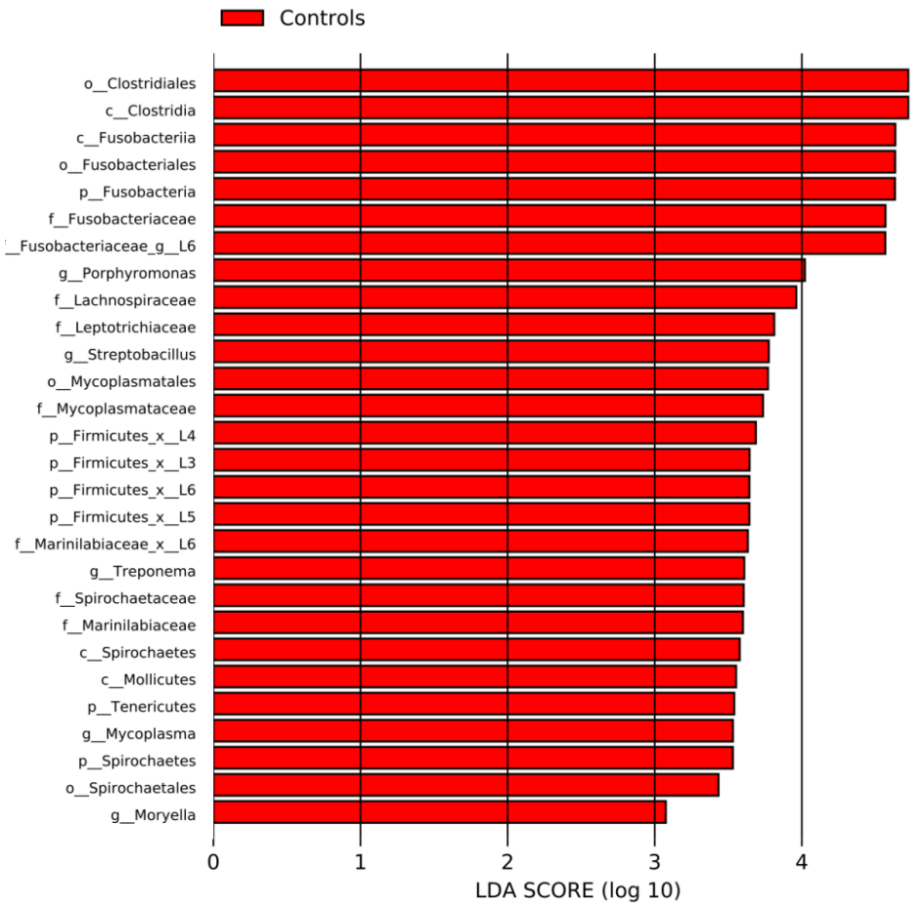

Figure 4 Linear discriminant analysis (LDA) of the effect size for particular taxa in the oropharyngeal samples of the control group. LDA scores >2.0 with p-value <0.05.

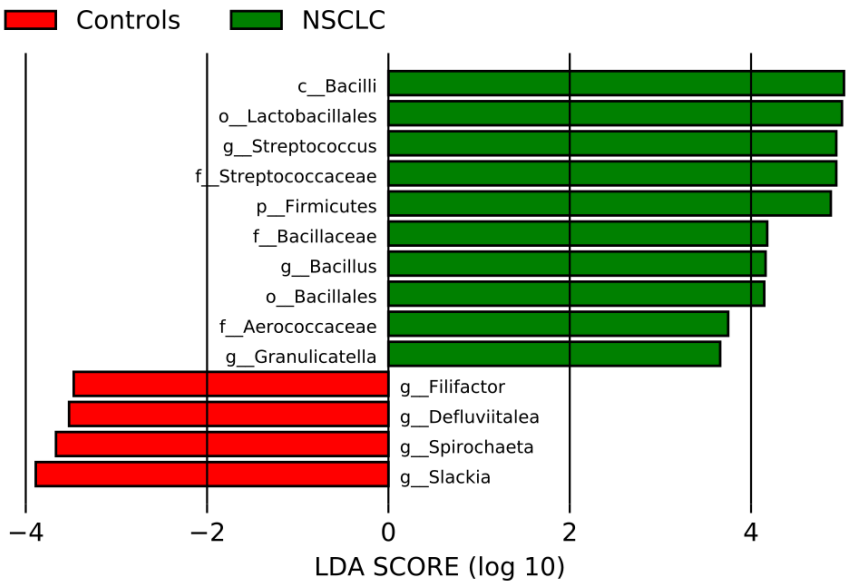

Figure 5 Linear discriminant analysis (LDA) of the effect size for particular taxa in sputum samples in the NSCLC and Control groups. LDA scores >2.0 with p-value <0.05.

The relative abundance of bacterial types, genera and species in the microbiome of sputum and oropharynx in NSCLC patient and control groups are presented in Table 4. None of the bacterial

taxa at the type, genus or species level differed in their relative abundance in the microbiome of sputum in the NSCLC patient and control groups, since the initially determined P values were not less than FDR-corrected P values. At the same time, we found significant differences in the microbiome of the oropharynx in patient and control groups at the level of genera and species. Specifically, in the microbiome of the oropharynx from NSCLC patients, the abundance of the genus *Rothia* was more than two times higher than in the microbiome of the control group (4.98 ± 6.33 vs 2.21 ± 6.28; P = 0.0008). Another two bacterial genera, *Parvimonas* and *Catonella* were more represented in the microbiome of the control group than in NSCLC patients (Table 4).

Table 4 Comparison of the relative abundance of bacterial taxa in the sputum and oropharynx of NSCLC patients and healthy controls.

Types/Genus/Species	Sputum			Oropharynx		
	NSCLC patients, n = 23	Controls, n = 20	P	NSCLC patients, n = 23	Controls, n = 20	P
Types:						
Firmicutes	53.03 ± 13.16	44.99 ± 12.2	0.04	43.33 ± 21.22	48.66 ± 17.74	0.66
Bacteroidetes	23.83 ± 11.54	28.59 ± 11.34	0.15	16.50 ± 11.18	24.15 ± 14.34	0.04
Actinobacteria	8.15 ± 5.44	6.67 ± 5.22	0.37	14.47 ± 10.78	7.18 ± 7.57	0.01
Proteobacteria	7.77 ± 7.60	9.08 ± 12.08	0.64	7.25 ± 6.68	5.98 ± 5.30	0.51
Fusobacteria	5.19 ± 5.44	6.98 ± 3.99	0.03	2.99 ± 3.91	7.70 ± 9.78	0.009
TM7	1.57 ± 1.79	1.84 ± 1.59	0.31	1.19 ± 1.90	1.51 ± 2.42	0.62
Spirochaetes	0.26 ± 0.45	0.79 ± 1.22	0.08	0.03 ± 0.11	0.55 ± 1.17	0.11
SR1	0.08 ± 0.20	0.87 ± 2.46	0.09	0	0.20 ± 0.62	0.13
Tenericutes	0.07 ± 0.16	0.34 ± 0.97	0.29	0.01 ± 0.05	0.08 ± 0.23	0.24
Genus:						
Streptococcus	31.65 ± 17.96	18.64 ± 9.90	0.01	21.18 ± 17.22	16.79 ± 16.16	0.31
Rothia	1.89 ± 1.87	1.84 ± 2.31	0.64	5.02 ± 6.40	2.21 ± 6.28	0.001*
Actinomyces	3.62 ± 2.66	2.78 ± 2.71	0.79	6.18 ± 6.67	3.21 ± 4.97	0.02
Porphyromonas	2.43 ± 2.39	4.18 ± 3.52	0.08	1.64 ± 2.67	2.65 ± 3.11	0.009
Granulicatella	1.67 ± 1.75	1.01 ± 0.99	0.30	2.10 ± 1.81	0.99 ± 1.41	0.03
Fusobacterium	1.49 ± 1.42	2.37 ± 2.87	0.31	0.82 ± 1.57	5.78 ± 9.67	0.03
Haemophilus	0.39 ± 0.90	0.12 ± 0.35	0.09	0.65 ± 1.83	0.00	0.02
Parvimonas	0.42 ± 0.86	0.95 ± 1.22	0.03	0.08 ± 0.19	0.78 ± 1.26	0.01*
Filifactor	0.07 ± 0.16	0.34 ± 0.46	0.04	0.02 ± 0.10	0.33 ± 0.69	0.04
Johnsonella	0.03 ± 0.12	0.44 ± 0.72	0.03	0.14 ± 0.34	0.42 ± 0.68	0.07
Catonella	0.05 ± 0.14	0.14 ± 0.24	0.10	0.03 ± 0.12	0.52 ± 0.82	0.005
Species:						
Streptococcus agalactiae	29.99 ± 18.25	18.80 ± 8.95	0.04	21.32 ± 16.87	16.36 ± 16.31	0.24
Granulicatella balaenopterae	2.04 ± 1.60	1.80 ± 1.14	0.76	1.97 ± 1.66	0.99 ± 1.41	0.03
Rothia terrae	1.50 ± 1.82	1.57 ± 2.44	0.79	4.51 ± 6.50	0.87 ± 2.10	0.002*

Rothia dentocariosa ATCC 17931	0.40 ± 0.67	0.32 ± 0.49	0.90	0.46 ± 1.02	0.00	0.008*
Porphyromonas endodontalis	0.18 ± 0.35	0.65 ± 0.74	0.02	0.10 ± 0.36	1.54 ± 2.84	0.002*
Oribacterium Sinus	0.12 ± 0.34	0.38 ± 0.56	0.04	0.05 ± 0.17	0.18 ± 0.58	0.88
Prevotella nigrescens	0.15 ± 0.29	0.67 ± 1.50	0.07	0.01 ± 0.06	0.77 ± 2.50	0.02
Prevotella oris F0302	0.29 ± 0.49	1.15 ± 1.68	0.07	0.14 ± 0.48	2.56 ± 8.61	0.02

Note: * P value is lesser than FDR corrected P value.

In the oropharyngeal microbiome of NSCLC patients, an increase at the species level was noted in the abundance of *Rothia* representatives: *Rothia terrae* (4.51 ± 6.5 vs 0.87 ± 2.1; P = 0.002) and *Rothia dentocariosa ATCC 17931* (0.46 ± 1.02 vs 0; P = 0.01). On the contrary, another bacterial species, *Porphyromonas endodontalis* was more represented in the microbiome of the oropharynx from healthy donors than in the microbiome of NSCLC patients (1.54 ± 2.84 vs 0.1 ± 0.36; P = 0.002).

Moreover, conditional logistic regression models adjusted for age, smoking status, alcohol consumption status, and living environment, and the phyla (*Rothia*, *Rothia terrae*, *Rothia dentocariosa*, *Porphyromonas endodontalis*) were constructed. In these models, age (OR, 1.17 [95% CI, 1.05 to 1.35], P = 0.011), and presence of *Porphyromonas endodontalis* (0.107 [0.011-0.55], P = 0.021) were more strongly associated with NSCLC as compared to healthy subjects.

4. Discussion

In the current study, 16S rRNA sequencing was used to assess and compare the composition and diversity of sputum and oropharyngeal microbiota associated with NSCLC and a healthy control group in a Russian population from Western Siberia. To the best of our knowledge, this report is the first to use 16S rRNA approaches to profile and compare the microflora composition in oropharyngeal swabs and sputum samples and to determine microbiota characteristics and biomarkers for the early diagnosis of NSCLC.

Previous studies have revealed topographical differences in the composition of the bacterial microbiome in different parts of the human respiratory tract [23, 24]. While the microbiome of sputum and saliva in LC has been the subject of several previous studies [6, 10, 25-29], no such study has been performed on oropharyngeal samples. According to our results, the overall taxonomic composition of the oropharyngeal microbiome was similar to the composition of the sputum microbiome. However, in NSCLC patients, alpha diversity was decreased in the oropharynx compared to the sputum microbiome, as indicated by the Kruskal-Wallis test. And this was evident in NSCLC patients only and not in healthy subjects. The PERMANOVA test (Adonis), which uses a matrix of differences constructed according to the Jacquard method also showed a significant difference in prokaryotic communities (beta diversity) in the sputum and oropharynx of both patients and controls at the different phylogenetic levels.

At the phylogenetic levels of type and genus, no statistically significant differences were found between the oropharynx and sputum of NSCLC samples (Table 2). In the oropharynx and the sputum of healthy donors, the relative abundance of most bacterial genera also represented no significant differences. The exceptions were representatives of *Capnocytophaga*, which were

significantly lower in the oropharynx than in the sputum, as well as *Macellibacteroides*, which, on the contrary, were higher in the oropharynx than in the sputum of controls.

Capnocytophaga is a genus of Gram-negative bacteria and is part of the oral commensal flora of immunocompetent patients [30]. Thus, its predominance in the sputum microbiome of healthy individuals in this study may reflect a satisfactory form of their immune system. *Macellibacteroides* is a genus from the family of *Porphyromonadaceae*. The significant decrease of *Macellibacteroides* in the sputum microbiome of healthy subjects in this study may indicate a lack of available iron in the sputum, which may appear during an acute immune response. Thus, this further indicates the steady state of the immune system in healthy individuals in this study.

Significant differences in the relative abundance of bacteria were observed only in the sputum samples from NSCLC patients, and only for rare bacteria such as *Leptotrichia sp. oral clone GT018*, *Bergeyella sp. AF14*, *Bergeyella zoohelcum*, *Prevotella Tannarae* and others (Table 3). The similarity in the taxonomic composition of the sputum and oropharyngeal microbiome is not surprising, given the proximity of the origin of these samples in the respiratory tract. The composition of the microbiota tends to show greater homogeneity even in more distant parts of the respiratory tract, as, for example, by BALF and saliva samples from LC patients [14]. Nevertheless, there were certain differences between these samples, including statistically significant ones.

Comparison of the taxonomic composition of bacteria from sputum and oropharyngeal samples between NSCLC patients and healthy controls was the next main objective of our study. There was no significant difference in alpha diversity indexes between NSCLC According to the journal layout rule, references should be numbered in numerical order. Please kindly edit the number and reference list. patients and controls in the sputum and oropharynx, as shown by Robust Aitchison analysis. In this respect, our current results are inconsistent with previous findings that loss of bacterial diversity is common in LC [9].

At the bacterial taxa level, there was a tendency for an increase in *Firmicutes* in the sputum of patients compared to controls, but the difference was not statistically significant when FDR adjustment was applied. In the oropharynx of patients, there were other tendencies for differences in the content of bacterial types between patients and controls (Table 4). In particular, the samples from the oropharynx of NSCLC patients contained more representatives of *Actinobacteria* and less of representatives of *Bacteroidetes*, but all these differences were also insignificant considering the FDR correction.

The *Streptococcus* genus has been shown to be a shared indicator of LC, as demonstrated in saliva samples, bronchial biopsy specimens, bronchoalveolar lavages [31, 32] and sputum [6, 10, 25, 28]. In the current study, we also observed an increased abundance of *Streptococcus* in sputum samples from patients compared to controls, however, this increase was not statistically significant when adjusted for FDR correction. This may be because the relatively small sample size used for the comparisons in this study (n = 23 for NSCLC patients and n = 20 for controls) affected the statistical power. Therefore, to assess the significance of increased *Streptococcus* in the sputum of NSCLC patients, we believe it is necessary to use a larger set of samples, as in our previous study [16].

Nevertheless, the relatively small set of samples available in the current study was sufficient to detect a statistically significant increase in the representation of the genus *Rothia* and specifically the species *Rothia terrae* and *Rothia dentocariosa* when the oropharyngeal samples from patients

were compared to controls. In general, we found that the genus *Rothia* was more abundant in the oropharynx than in the sputum (Table 3), while the content of *Rothia* in the sputum of NSCLC patients and controls was practically the same. In addition to the significant enrichment of *Rothia* in the oropharynx samples from NSCLC patients, there was a significant decrease in representatives of the genera *Parvimonas* and *Catonella*, as well as in the species *Porphyromonas endodontalis*. *Rothia* is a gram-positive, aerobic, rod-shaped bacterial genus from the family of Micrococcaceae. The genus *Rothia* predominates in saliva and can be detected in the oropharyngeal flora as well. *Rothia* is a conditionally pathogenic opportunistic pulmonary agent. *Rothia* bacteria can cause disease in healthy and immunosuppressed subjects [33]. An increased relative abundance of the genus *Rothia* was previously found in the sputum of patients with the chronic obstructive pulmonary disease [34, 35] and cystic fibrosis [36]. Increased representation of *Rothia* and *Actinomyces* in saliva samples from both lung adenocarcinoma and lung squamous cell carcinoma patients has been described previously. And *Rothia* also showed a significant difference between a lung adenocarcinoma group and a healthy control group (4.77% vs 2.06%, P = 0.04) [14]. Bacteria in the genus *Rothia* produce enterobactin, a potent siderophore, a secreted iron-binding bacterial compound [37].

Porphyromonas endodontalis is known to actively transport free *iron.* It is an obligate anaerobic rod-shaped bacterium implicated as a major pathogen in endodontic infections [38].

The significantly altered representation of *Porphyromonas endodontalis and Rothia* in patient samples may indicate a role of iron metabolism in the pathogenesis of LC. It is known that LC cells are tolerant to high concentrations of Fe(II) ions [39].

In our study, the overrepresentation of *Rothia* in the oropharynx of NSCLC patients and *Porphyromonas endodontalis* in the sputum of healthy participants indicated that these niches offer favorable conditions for these bacteria, quite possibly due to excess iron in the oropharynx of patients with NSCLC (due to massive cell death there) and a lack of free iron in their saliva. Thus, the study of the taxonomic composition of the microbiome may play an important role in understanding the mechanisms of tumor-associated iron metabolism, which in the long term may contribute to more effective treatments for LC.

In our study, the content of *Actinomyces* in the oropharynx of patients was also higher than in control samples (6.18 ± 6.67 vs 3.21 ± 4.97; P = 0.02), however, these differences were not significant after taking into account the FDR correction. S. Bello et al. recently observed the relative abundance of *Rothia* (among other bacterial genera such as *Streptococcus*, *Gemella* and *Lactobacillus*) in the saliva of patients with central LC [30], but to date, *Rothia* has not been discussed as a biomarker for this type of cancer.

5. Conclusions

The object of this study was to better understand the microbial communities in sputum and oropharyngeal samples from healthy controls and NSCLC patients and to identify lung-specific taxa associated with cancer or specific to these sites. Despite the generally observed similarity taxonomic composition of the oropharyngeal microbiome and sputum, there were undoubtedly common and/or species differences between these two sample sites in the healthy controls and the NSCLC group. In general, microbiota biodiversity was found to be higher in sputum samples compared to the oropharynx. Our results indicated that individual bacterial taxa in the oropharynx

could be associated with NSCLC, but their unequivocal identification and association with NSCLC must await further studies using a larger cohort. Specifically, differences in the relative percentage of members of the genus *Rothia* identified in oropharyngeal samples from NSCLC patients compared to controls using the Mann-Whitney U-test were not consistent with the results of LDA using LefSe. Nevertheless, we believe that further research in this direction, provided a significant increase in the number of samples, as well as the use of quantitative PCR, will advance the search for NSCLC metagenomic biomarkers.

Acknowledgments

The authors wish to thank the physicians and staff of the Kemerovo Regional Oncology Center, all of the surveyed individuals who voluntarily participated in this study and the employees of the Kemerovo State University and the Institute of Human Ecology who participated in the organization and conducting of this research.

Author Contributions

V.D. and L.M. conceived the study; V. D. wrote the manuscript; E.B., V.V., A.L performed laboratorial work; P.D. carried out bioinformatics and statistical analyses; L.M. critically revised the manuscript.

Funding

This work was supported by Russian Science Foundation Grant No. 18-14-00022p.

Competing Interests

The authors have declared that no competing interests exist.

References

1. Cheng TYD, Cramb SM, Baade PD, Youlden DR, Nwogu C, Reid ME. The international epidemiology of lung cancer: Latest trends, disparities, and tumor characteristics. J Thorac Oncol. 2016; 11: 1653-1671.
2. Pallis AG, Syrigos KN. Lung cancer in never smokers: Disease characteristics and risk factors. Crit Rev Oncol Hematol. 2013; 88: 494-503.
3. Mao Q, Jiang F, Yin R, Wang J, Xia W, Dong G, et al. Interplay between the lung microbiome and lung cancer. Cancer Lett. 2018; 415: 40-48.
4. Goto T. Airway microbiota as a modulator of lung cancer. Int J Mol Sci. 2020; 21: 3044.
5. Tsay JCJ, Wu BG, Sulaiman I, Gershner K, Schluger R, Li Y, et al. Lower airway dysbiosis affects lung cancer progression. Cancer Disc. 2021; 11: 293-307.
6. Leng Q, Holden VK, Deepak J, Todd NW, Jiang F. Microbiota biomarkers for lung cancer. Diagnostics (Basel). 2021; 11: 407.
7. Jin J, Gan Y, Liu H, Wang Z, Yuan J, Deng T, et al. Diminishing microbiome richness and distinction in the lower respiratory tract of lung cancer patients: A multiple comparative study design with independent validation. Lung Cancer. 2019; 136: 129-135.

8. Hasegawa A, Sato T, Hoshikawa Y, Ishida N, Tanda N, Kawamura Y, et al. Detection and identification of oral anaerobes in intraoperative bronchial fluids of patients with pulmonary carcinoma. Microbiol Immunol. 2014; 58: 375-381.
9. 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. 2016; 102: 89-95.
10. 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: e0177062.
11. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, et al. Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. Int J Cancer. 2018; 142: 769-778.
12. Peters BA, Hayes RB, Goparaju C, Reid C, Pass HI, Ahn J. The microbiome in lung cancer tissue and recurrence-free survival. Cancer Epidemiol Biomarkers Prev. 2019; 28: 731-740.
13. Zhang W, Luo J, Dong X, Zhao S, Hao Y, Peng C, et al. Salivary microbial dysbiosis is associated with systemic inflammatory markers and predicted oral metabolites in non-small cell lung cancer patients. J Cancer. 2019; 10: 1651-1662.
14. Wang K, Huang Y, Zhang Z, Liao J, Ding Y, Fang X, et al. A preliminary study of microbiota diversity in saliva and bronchoalveolar lavage fluid from patients with primary bronchogenic carcinoma. Med Sci Monit. 2019; 25: 2819-2834.
15. Goldstraw P. New staging system: How does it affect our practice? J Clin Oncol. 2013; 31: 984-991.
16. Druzhinin VG, Matskova LV, Demenkov PS, Baranova ED, Volobaev VP, Minina VI, et al. Genetic damage in lymphocytes of lung cancer patients is correlated to the composition of the respiratory tract microbiome. Mutagenesis. 2021; 36: 143-153.
17. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010; 7: 335-336.
18. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019; 37: 852-857.
19. Lozupone C, Knight R. UniFrac: A new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005; 71: 8228-8235.
20. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A novel sparse compositional technique reveals microbial perturbations. mSystems. 2019; 4: e00016-19.
21. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011; 12: R60.
22. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020; 11: 3514.
23. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011; 184: 957-963.
24. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, et al. Bacterial topography of the healthy human lower respiratory tract. mBio. 2017; 8: e02287-16.

25. Hosgood 3rd HD, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, et al. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. Environ Mol Mutagen. 2014; 55: 643-651.
26. Yan X, Yang M, Liu J, Gao R, Hu J, Li J, et al. Discovery and validation of potential bacterial biomarkers for lung cancer. Am J Cancer Res. 2015; 5: 3111-3122.
27. Yang J, Mu X, Wang Y, Zhu D, Zhang J, Liang C, et al. Dysbiosis of the salivary microbiome is associated with non-smoking female lung cancer and correlated with immunocytochemistry markers. Front Oncol. 2018; 8: 520.
28. Ran Z, Liu J, Wang F, Xin C, Shen X, Zeng S, et al. [Analysis of pulmonary microbial diversity in patients with advanced lung cancer based on high-throughput sequencing technology]. Zhongguo Fei Ai Za Zhi. 2020; 23: 1031-1038.
29. Shi J, Yang Y, Xie H, Wang X, Wu J, Long J, et al. Association of oral microbiota with lung cancer risk in a low-income population in the southeastern USA. Cancer Causes Control. 2021; 32: 1423-1432.
30. Jolivet-Gougeon A, Sixou JL, Tamanai-Shacoori Z, Bonnaure-Mallet M. Antimicrobial treatment of Capnocytophaga infections. Int J Antimicrob Agents. 2007; 29: 367-373.
31. Bello S, Vengoechea JJ, Ponce-Alonso M, Figueredo AL, Mincholé E, Rezusta A, et al. Core microbiota in central lung cancer with streptococcal enrichment as a possible diagnostic marker. Arch Bronconeumol. 2021; 57: 681-689.
32. Seixas S, Kolbe AR, Gomes S, Sucena M, Sousa C, Vaz Rodrigues L, et al. Comparative analysis of the bronchoalveolar microbiome in Portuguese patients with different chronic lung disorders. Sci Rep. 2021; 11: 15042.
33. Baeza Martínez C, Zamora Molina L, García Sevila R, Gil Carbonell J, Ramos Rincon JM, Martín Serrano C. Rothia mucilaginosa pneumonia in an immunocompetent patient. Arch Bronconeumol. 2014; 50: 493-495.
34. Garcia-Nuñez M, Millares L, Pomares X, Ferrari R, Pérez-Brocal V, Gallego M, et al. Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease. J Clin Microbiol. 2014; 52: 4217-4223.
35. Leitao Filho FS, Alotaibi NM, Ngan D, Tam S, Yang J, Hollander Z, et al. Sputum microbiome is associated with 1-year mortality after chronic obstructive pulmonary disease hospitalizations. Am J Respir Crit Care Med. 2019; 199: 1205-1213.
36. Carmody LA, Zhao J, Schloss PD, Petrosino JF, Murray S, Young VB, et al. Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. Ann Am Thorac Soc. 2013; 10: 179-187.
37. Uranga CC, Arroyo Jr P, Duggan BM, Gerwick WH, Edlund A. Commensal oral rothia mucilaginosa produces enterobactin, a metal-chelating siderophore. mSystems. 2020; 5: e00161-20.
38. Zerr M, Drake D, Johnson W, Cox CD. Porphyromonas endodontalis binds, reduces and grows on human hemoglobin. Oral Microbiol Immunol. 2001; 16: 229-234.
39. Fonseca-Nunes A, Jakszyn P, Agudo A. Iron and cancer risk--a systematic review and meta-analysis of the epidemiological evidence. Cancer Epidemiol Biomarkers Prev. 2014; 23: 12-31.

Enjoy *OBM Genetics* by:

1. Submitting a manuscript
2. Joining in volunteer reviewer bank
3. Joining Editorial Board
4. Guest editing a special issue

For more details, please visit:
http://www.lidsen.com/journals/genetics