

Alterations of lung microbiota in patients with non-small cell lung cancer

Wen Zeng^{a,b}, ChengZhu Zhao^{a,#}, Mengge Yu^a, Hailong Chen^b, Yiyun Pan^b, Yuhuan Wang^a, Hejing Bao^a, Hao Ma, and Shudong Ma^a

^aDepartment of Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China; ^bDepartment of Oncology, Ganzhou Cancer Hospital, Gannan Medical University, Ganzhou, Jiangxi, China

ABSTRACT

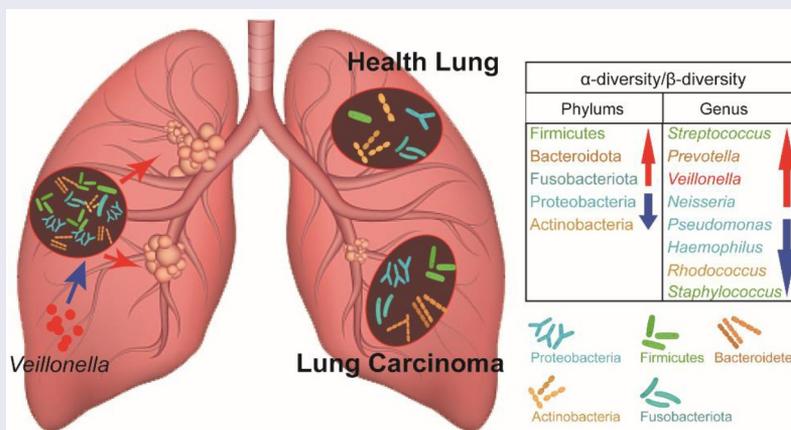
The role of lung microbiota in non-small cell lung cancer remains unclear. We investigated the characteristics and functional roles of lung microbiota in non-small cell lung cancer. Bronchoalveolar lavage fluid samples were obtained from patients with non-small cell lung cancer (n = 46) and with benign lung disease (n = 29). The differences in composition and gene expression in the microbiota between the samples were analyzed using 16S rRNA sequencing. The oncogenic genus (*Veillonella*) was then evaluated in the progression of lung cancer in C57 BL/6 mice. Compared to benign lung disease, the lung microbiota in non-small cell lung cancer was significantly altered, both in terms of α - and β -diversity. In terms of bacterial composition, the non-small cell lung cancer group was enriched with two Phyla (Firmicutes, Bacteroidetes) and three genera (*Streptococcus*, *Prevotella*, *Veillonella*). *Prevotella* and *Veillonella* were most strongly associated with non-small cell lung cancer, and *Veillonella* significantly promoted the progression of lung cancer *in vivo*. Moreover, metabolic prediction revealed that ribosomes, biosynthesis of secondary metabolites, and pyrimidine metabolism were among the enriched pathways that may be involved in the progression of non-small cell lung cancer. Overall, results suggest that the progression of non-small cell lung cancer is followed by significant changes in the composition and function of the lung microbiota. These differing genera may be potential diagnostic markers and therapeutic targets.

ARTICLE HISTORY

Received 12 January 2022
Revised 17 February 2022
Accepted 18 February 2022

KEYWORDS

Non-small cell lung cancer; microbiota; microbial diversity; 16S rRNA gene; *veillonella*



1. Introduction

Lung cancer is one of the most common cancers and causes of cancer-related deaths globally. Each year, over 1.8 million people are diagnosed with lung cancer, and 1.6 million of them succumb

to the disease. Non-small cell lung cancer (NSCLC) is the most common type of lung cancer pathology and is caused by environmental factors and host genetics [1]. Smoking is also recognized as a major known risk factor for lung cancer, but

CONTACT Shudong Ma  hxingwei2020@yeah.net  Department of Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, People's Republic of China

#Co-first author

 Supplemental data for this article can be accessed [here](#)

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

less than 15% of patients with NSCLC smoke [2]. Chronic obstructive pneumonia, bronchitis, and pneumonia are associated with an increased risk of lung cancer [3], and inflammatory diseases have overlapping pathways with those implicated in the pathogenesis of NSCLC [4]. Furthermore, although the genomic profile of NSCLC has been extensively described, the external factors that influence the development of NSCLC remain unclear [5].

Since studies have reported a causal relationship between *Helicobacter pylori* and gastric cancer [6], there is growing evidence that local microorganisms play a key role in cancer pathogenesis and treatment, such as *Fusobacterium nucleatum* promoting colon cancer and *Enterotoxigenic Bacteroides fragilis* promoting breast cancer [7,8]. The lung microbiota is involved in the regulation of the host pulmonary immune system, and the balance between pulmonary immunity and microbiota is crucial for protection against infection [9]. Dysbiosis of the lung microbiota can induce activation of resident immune cells, such as M1 macrophages and $\gamma\delta$ T cells, and induce oxygen free radicals and gene mutations, thus promoting lung carcinogenesis [10]. In mouse models, dysregulation of lung bacteria generates a pro-inflammatory environment and promotes tumor growth. However, the relationship between the function of the lung microbiota and the host in the context of lung cancer remains unknown.

In this study, we hypothesized that there exists a significant change in composition and function of the lung microbiota during the progression of NSCLC. Newly diagnosed NSCLC patients are suitable for profiling the relationship between lung microbiota and cancers. To test this hypothesis, we examined the microbiota composition of bronchoalveolar lavage fluid (BALF) from 46 patients with newly diagnosed NSCLC and 29 patients with benign lung disease using 16s rRNA sequencing and searched for associations between NSCLC and the microbiota through bioinformatic analysis and *in vivo* experiments. We aim for our results to provide a reference for the clinical application of lung microbiota evaluation in the diagnosis and treatment of NSCLC.

2. Materials and methods

2.1 Patients and samples collection

A total of 75 patients who underwent computed tomography examination suggestive of suspicious lung nodules and clinical bronchoscopy at the Ganzhou cancer hospital (Jiangxi, China) from 1 December 2020 to November 31, 2021, were enrolled in this study. The patients included those with severe bronchitis, chronic obstructive pneumonia, bronchiectasis, and pulmonary fibrosis. Forty-six patients diagnosed with NSCLC were included in the newly diagnosed lung cancer group, and 29 patients with a pathological diagnosis of benign lesions were included in the control group. No antibiotics were administered within 4 weeks of enrollment. Population and clinical data were recorded, including age, sex, body mass index (BMI), smoking history, type of pathology, and tumor stage. The study was approved by the Ethics Committee of Ganzhou cancer hospital (2021 Research Ethics Review No. 5), and all enrolled patients signed an informed consent form.

BALF was collected by an experienced clinician following the optical fiber bronchoscopy protocol to avoid oral contamination to the greatest extent possible. In total, 10–15 mL of BALF were collected from each patient, filtered through a single layer of gauze to remove mucus, and centrifuged at 10,000 g for 10 min. The precipitate was then collected and frozen at -80°C for further use.

2.2 DNA extraction and polymerase chain reaction amplification

Total genomic DNA was extracted from BALF using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). DNA purity and concentration were measured using a NanoDrop 2000 (Thermo Scientific, USA), and the integrity of the extracted DNA was determined by 1% agarose gel electrophoresis. Polymerase chain reaction (PCR) amplification was performed using TransStart Fastpfu DNA Polymerase (AP221-02, TransGen, Beijing, China), using the following primers to amplify the 16s rRNA 'V3-V4' sequence: 338 F 'ACTCCTACGGGAGGCAGCAG' and 806 R

'GGACTACHVGGGTWTCTAAT'. The V3-V4 region sequences were sequenced and analyzed.

2.3 Gene sequencing

The 16s rRNA V3-V4 region genes in the samples were sequenced in both directions using the Illumina MiSeq high-throughput platform (Meiji Biological Co., Ltd. Shanghai, China). The paired-end reads obtained from MiSeq sequencing were first spliced based on overlap relationships, while the sequence quality was quality controlled and filtered, then analyzed using QIIME2, and the reads were clustered into operational taxonomic units (OTUs) using UCLUST with 97% similarity [11]. Taxonomic analysis of OTUs was performed according to Greengene (Release 13.5 <http://greengenes.secondgenome.com/>) [12].

2.4 Sequencing analysis

Shannon index, Simpson diversity index, Sobs index, and Chao1 index diversity were used for alpha (α)-diversity analysis of species, and the Wilcoxon rank-sum test was used to test for differences between index groups. Similarities or differences in community composition between different grouped samples underwent principle coordinate analysis (PCoA), analysis of similarities (ANOSIM), partial least squares discriminant analysis (PLS-DA) based on the principal coordinates of Bray Curtis, and analysis of unweighted and weighted UniFrac distance matrices [13]. The linear discriminant analysis (LDA) effect size (LEfSe) algorithm (huttenhower.sph.harvard.edu/galaxy/) was used for the non-parametric Kruskal–Wallis sum-rank test to detect differences in species abundance between groups and to obtain significantly different species and metabolic pathways [14]. LDA was used to evaluate the magnitude of the effect of different species or metabolic pathways on the difference between groups, with taxa with an LDA score >4 and a p-value <0.05 being considered significantly enriched. Species correlation networks were constructed from one-way network analysis to analyze species interactions in the environment, and correlation coefficients were described using

a heat map. Reconstruction of Unobserved States2 (PICRUST2) was used to predict the functional composition of the taxa in the samples from amplicon sequencing results, and the predicted Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway level 3 was analyzed descriptively.

2.5 Bacterial strains, cell line, and culture conditions

Veillonella parvula ATCC 10790 (0867, ATCC, USA) was purchased from ATCC and maintained in Wilkins-Chalgren anaerobe broth (Thermo Fisher Scientific, USA) in an anaerobic jar (80% N₂, 10% H₂, 10% CO₂) at 37°C. The Lewis lung cancer (LLC) cell line was purchased from Procell Life Science & Technology Co., Ltd (Wuhan, China). Cells were maintained in Dulbecco's Modified Eagle Medium (Gibico, USA) supplemented with 10% fetal bovine serum (Gibico, USA).

2.6 Syngeneic tumor models

Six- to 8-week-old C57 bl/6 j male mice were obtained from Southern Medical University (Guangzhou, China). LLC cells ($5 \times 10^5/100 \mu\text{l}$) were injected either subcutaneously on the left flank region or intratracheal instillation, and the animals were kept in a specific pathogen-free environment. Mice were executed at the fourth week after inoculation with LLC cells, and tumor weights and volumes were recorded. Volume (mm^3) was calculated as $L \times W \times W/2$, with L for the length of the tumor and W for the width of the tumor (in centimeters).

2.7 Statistical analysis.

The student's t-test was performed to compare the variables between the two sample groups. Data are expressed as the mean \pm standard deviation or mean \pm standard error of the mean, as indicated in the figure legends. Statistical tests were two-tailed, and statistical significance was set at $P < 0.05$.

3. Results

3.1 Overview

We conducted a cross-sectional study to compare the differences in composition between the lung microbiota of NSCLC and benign lung disease by 16S rRNA gene sequencing. The diversity and composition of the lung microbiota differed significantly between the NSCLC group and the control group. The lung microbiota influenced NSCLC through metabolic pathways, and the *Veillonella* genus was identified as having an oncogenic role in promoting lung adenocarcinoma progression in mice.

3.2 Characteristics of the participants

To assess the variation and role of lower respiratory microbiota in lung cancer, the composition of BALF microbiota was examined using 16s rRNA gene sequencing in 29 control subjects and 46 patients with lung cancer. The median age was 63.8 ± 11.2 in the lung cancer group and 64.2 ± 8.4 in the control group. These patients did not have a serious infection, nor were they treated with antibiotics in the last 4 weeks before enrollment. Patients in both groups had similar age, BMI, male/female ratio, and comparable baseline information. Patients in the lung cancer group were all diagnosed with NSCLC (25 with adenocarcinomas and 21 with squamous carcinomas), and the control group was dominated by benign lung nodules (inflammatory pseudotumors, $n = 23$; malignant hemangiomas, $n = 4$; and sclerosing hemangiomas, $n = 2$). The stage of patients in the lung cancer group was predominantly intermediate and advanced, including 3 patients with stage I, 7 with stage II, 22 with stage III, 14 with stage IV, 13 with distant metastasis, and 1 with non-metastasis. Overall, the trends observed in the enrolled patients were consistent with the epidemiological trends of lung cancer (Table 1).

3.3 Lung microbiome profiles analyzed by 16s rRNA sequencing

During the analysis of 16s rRNA sequencing data from BALF from 75 samples, we screened 4,272,518 sequences with an average length of 421, which were then annotated in the rRNA library database

Table 1. Baseline characteristics of the patients.

	Non-small cell Lung cancer(n = 46)	Control (n = 29)	P value*
Demographics/ anthropometric			
Age/year	63.8 ± 11.2	64.2 ± 8.4	0.8693
Male/Female	34/12	21/8	0.8871
BMI(kg/m ²)	23.24 ± 2.12	23.54 ± 2.32	0.5668
Tumor stage (%)			
I	3	N/A	
II	7	N/A	
III	22	N/A	
IV	14	N/A	
Tumor Type(%)			
ADC	25	N/A	
LCC	21	N/A	
Tumor metastasis			
Metastasis	13	N/A	
Non-metastasis	1	N/A	
Smoking status (%)			
Never smoker	9	7	
Ever smoker	37	22	

N/A: Not applicable

(Greengenes) and analyzed for OTUs (Table S1). Taxonomic analysis of OTU representative sequences at 97% similarity level, sequence sampling, and exclusion of species with sequence sum <100 resulted in a total of 791 OTUs. Twenty-four and 25 OTUs were identified in the control and lung cancer groups, respectively, at the phylum, class, order, family, genus, and species levels. A detailed analysis of the unique microbial taxa in the different groups including a total of 66 OTUs, 47 species, 26 genera, and 1 phylum were detectable only in the cancer group, whereas another 13 OTUs, 8 species, and 7 genera were only detectable in the control group (Figure 1a, Table S2).

We found that the process of lung carcinogenesis was accompanied by significant changes in the composition of the lower respiratory microbiota. At the phylum level, the top five in abundance were *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, and *Fusobacteriota*. *Proteobacteria* and *Actinobacteria* showed high abundance in the control group, while *Firmicutes*, *Bacteroidota*, and *Fusobacteriota* were highly abundant in the lung cancer group (Figure 1b, d). At the genus level, the top 10 genera, *Prevotella*, *Streptococcus*, *Pseudomonas*, *Haemophilus*, *Veillonella*, *Rhodococcus*, *Neisseria*, *Porphyromonas*, *Staphylococcus*, and *Actinomyces*, were most common between the control and lung cancer groups, with *Pseudomonas*, *Haemophilus*,

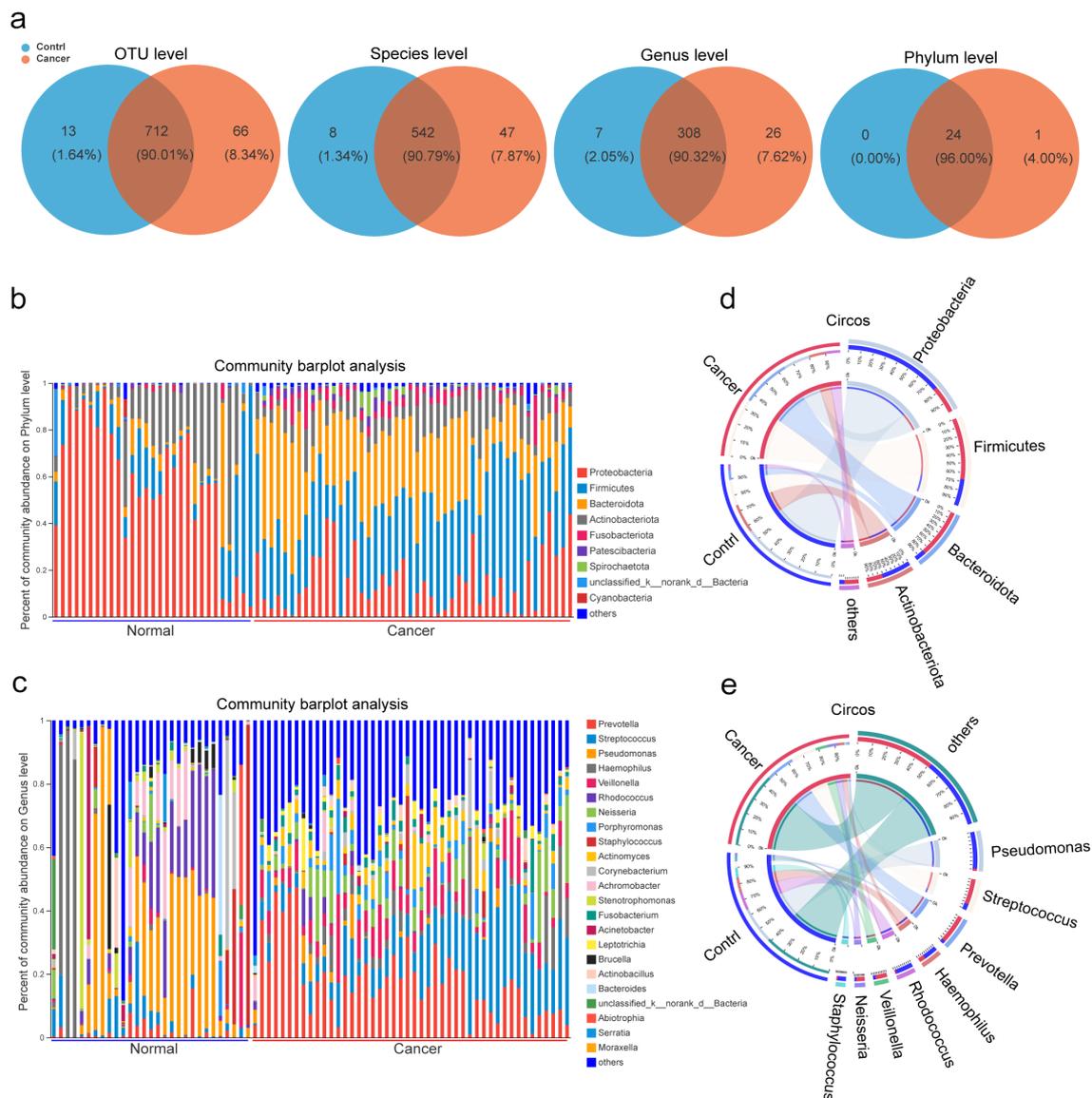


Figure 1. The profiles of lung microbiome composition in the control and lung cancer groups. (a) Venn diagrams show the numbers of microbiota at OTUs, species, genera, phyla commonly shared between both control and lung cancer groups. (b, c) Heatmaps show the relative frequency of lung microbiota at the phylum (b) and genus levels (c) in each sample. (d, e). Circos show the composition of lung microbiota between the control and lung cancer groups at the phylum (d) and genera (e) levels. The red and blue dots represent the cancer and control samples, respectively.

Rhodococcus, and *Staphylococcus* being highly abundant in the control group, while *Streptococcus*, *Prevotella*, *Veillonella*, and *Neisseria* were highly abundant in the lung cancer group (Figure 1c, e).

3.4 Biodiversity in the lung microbiota in NSCLC

The α -diversity in the bacterial community in the lower respiratory tract differed between patients in the lung cancer group and patients in the control group. The Sobs and Chao indices reflected the

difference in community richness between the controls and NSCLC patients ($p < 0.001$). Shannon and Simpson's indices reflected the difference in two groups ($p < 0.001$). These results showed lower respiratory microbiota may change considerably during lung carcinogenesis (Figure 2a, B, Figure S1A, B).

The composition of the lower respiratory microbiota of lung cancer samples and control samples was further assessed for a clear trend of segregation based on analysis of OTU levels.

Various β diversity indices (PCoA analysis of Bray Curtis, weighted UniFrac and unweighted UniFrac analysis) revealed the bacterial community composition was determined to be significantly different between the NSCLC and control samples ($p = 0.001$) (Figure 2c, Figure S1C, D). In addition, PLS-DA also showed that the bacterial community composition was significantly different between two groups (Figure 2d). Therefore, the composition of the lower respiratory microbiota was significantly altered in progression of NSCLC.

3.5 Distinct taxa in BALF samples between NSCLC and controls

The LEfSe method was implemented, and this analysis (LDA>4) showed significant differences in the microbiota of BALF samples between the control

and lung cancer groups (Figure 3a). Compared to the control group, Desulfobacterota, Firmicutes, Bacteroidetes, Fusobacteriota, Synergistota, Patescibacteria, and Campilobacterota were significantly enriched in the NSCLC group at the phylum level, while Proteobacteria and Calditrichota were enriched in the control samples (Figure 3b). At the genus level, *Prevotella*, *Streptococcus*, *Veillonella*, *Neisseria*, *Actinomyces*, *Alloprevotella*, and *Porphyromonas* were enriched in the NSCLC samples, while *Pseudomonas*, *Rhodococcus*, *Stenotrophomonas*, *Haemophilus*, *Achromobacter*, and *Brucella* were enriched in the control samples (Figure 3c). In addition, we also analyzed the differences in lung microbiota at the class, family, and species levels, contributing to a broader understanding of the important microbiota that distinguishes NSCLC from normal controls (Figure S2A-C).

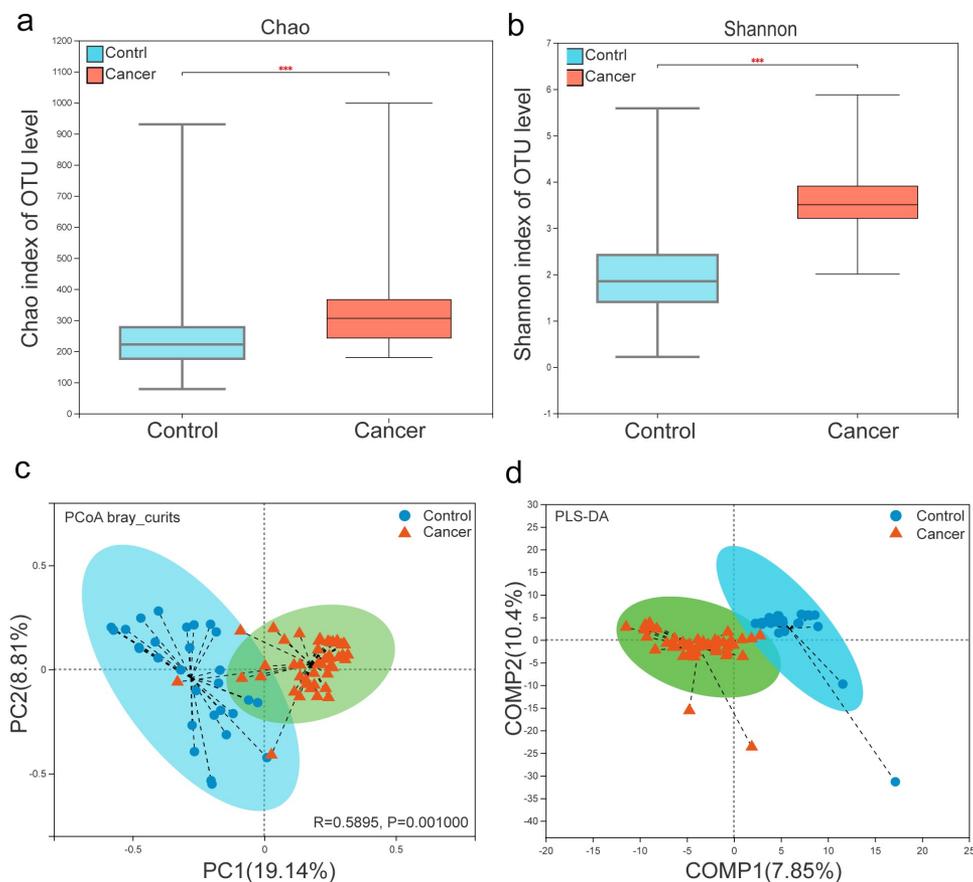


Figure 2. Analysis of α - and β -diversity in the microbial community of BALF samples from the NSCLC and control group. (a, b) Parameters Chao1 (a) and Shannon (b) indices were used for α -diversity analysis. (c) PCoA plots of Bray Curtis distance matrix. (d) Partial least squares discriminant analysis (PLS-DA) of β -diversity. The red and blue dots represent the cancer and control samples, respectively.

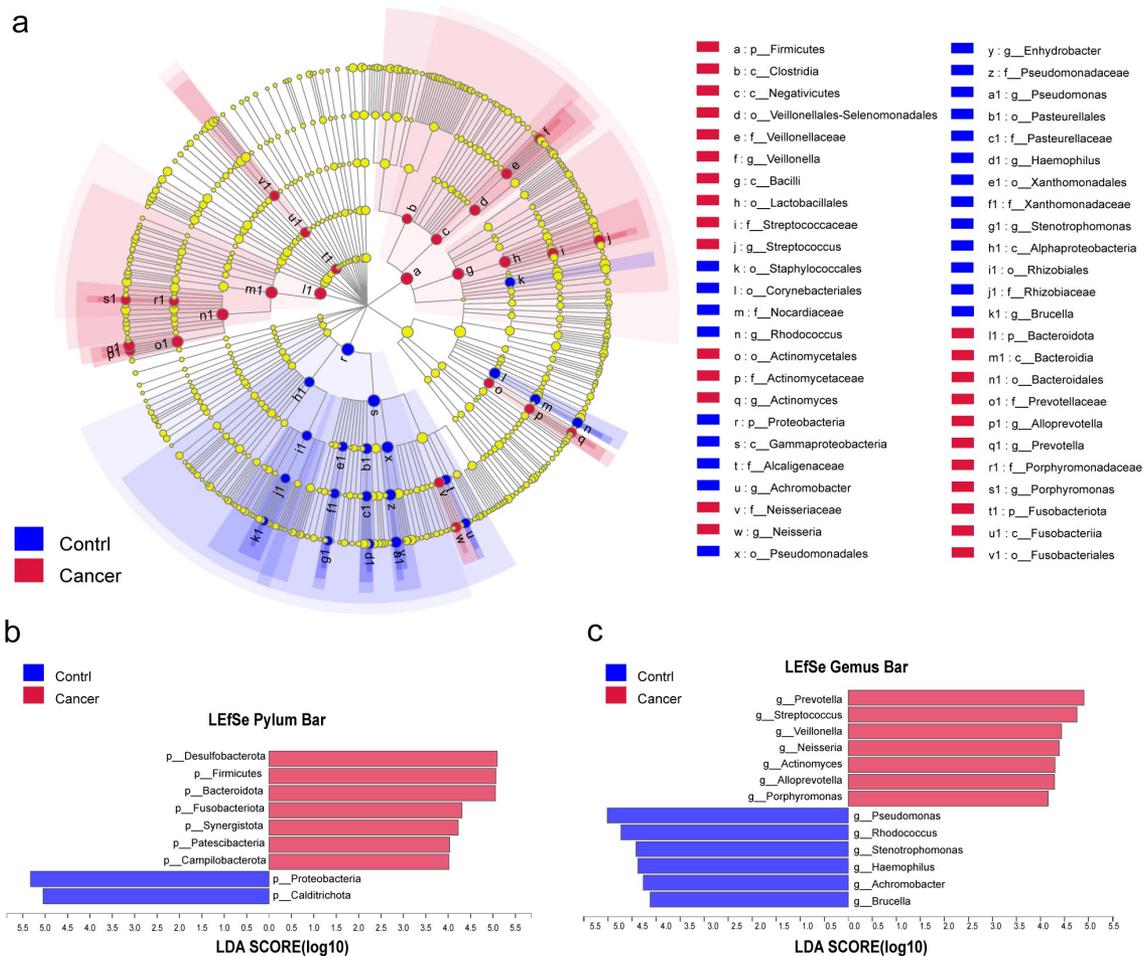


Figure 3. Distinct taxa were characterized in BALF samples by applying LEfSe analysis. (a) A Cladogram was constructed using the LEfSe method to represent the phylogenetic distribution of bacteria. (b,c) Linear discriminant analysis (LDA) scores show bacterial within two groups at the phylum level (b) and genus level (c). Blue color represents taxa enriched in the control patients, and red color represents taxa enriched in NSCLC patients.

3.6 Co-occurrence network analysis of microbiota in NSCLC

The genus-level correlation network plots revealed significant interactions between different genera, with *Prevotella*, *Alloprevotella*, *Veillonella*, *Pseudomonas*, and *Rhodococcus* being the five most significantly associated genera in the network. The first three genera were enriched in the lung cancer group and the latter two were enriched in the control group (Figure 4a). In the lung cancer group, *Prevotella* and *Prevotella* were positively correlated with *Megasphaera* ($\rho = 0.57595$, $\rho = 0.5275$), and *Alloprevotella* was positively correlated with *Actinomyces* ($\rho = 0.60058$), whereas in the control group, *Pseudomonas* was positively correlated with

Rhodococcus ($\rho = 0.59812$). Genus-to-genus and genus-to-sample interactions are important factors influencing lung cancer progression (Figure 4b).

3.7 Analysis of metabolic pathways of lung microbiota in NSCLC

The differential functions of bacterial communities were analyzed using by PICRUSt2, followed by LEfSe analysis KEGG functional predictions in cellular processes, human diseases, metabolism, environmental information processing, genetic information processing, and organismal systems. The differences between the NSCLC and control groups are shown in Figure 5a. Overall, 16 and 20 different pathways were enriched in the control

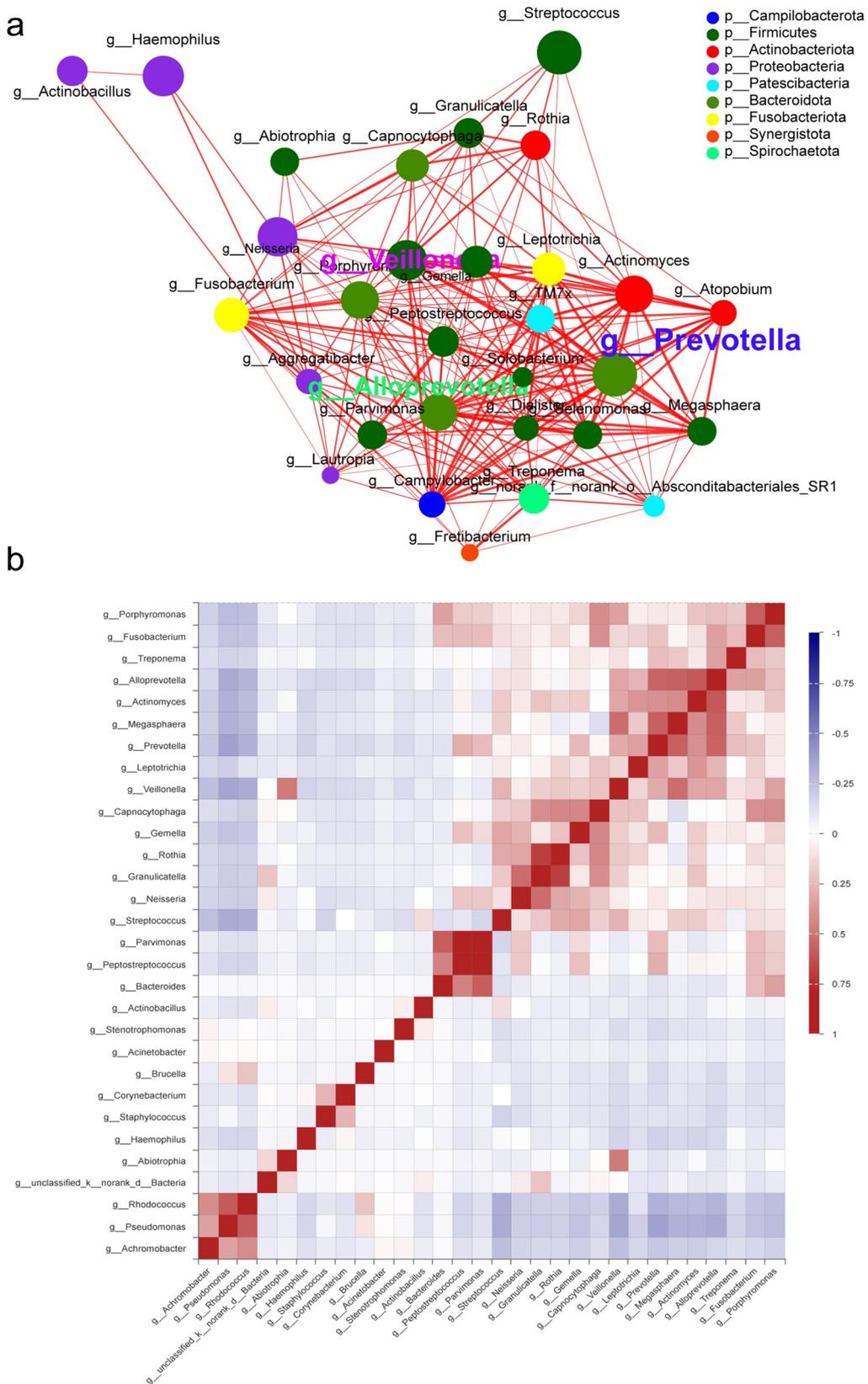


Figure 4. Co-occurrence network of the microbiome at the genus level. (a) Each node represents a genus, colored according to its phylum level, and each edge indicates an important symbiosis, which is colored according to its relevance (red: positive, green: negative). (b) Pearson correlations were calculated and analyzed between the top 50 most abundant bacterial genus. Correlation values range from -1.00 (blue) to 1.00 (red).

and NSCLC groups, respectively, with metabolic pathways, ribosome, biosynthesis of secondary metabolites, and pyrimidine metabolism being the distinct metabolic processes identified in the lung cancer group, while two-component system, microbial metabolism in diverse environments, and ABC transporters were significantly present in the control group.

Further classification of KEGG functional predictions into cellular processes, human diseases, metabolism, organismal systems, and environmental information processing suggested that the different metabolic functions of the microbiota may be an important factor influencing the progression of lung cancer (see Table 1) (Figure S3).

3.8 *Veillonella* promotes the growth of LLC tumors *in vivo*

To further observe the effect of the key genus in the lung on the growth of NSCLC, we used C57BL/6 mice and LLC cell lines for subcutaneous transplantation of tumors and *in situ* cancer models to conduct *in vivo* studies on the interaction of

bacteria with lung cancer. Peri-tumor injection with *Veillonella parvula* in LLC subcutaneous tumors significantly promoted the growth of LLC transplanted tumors in a relatively short period (Figure 6a). In contrast, intratracheal instillation of *Veillonella parvula* in mice with LLC *in situ* tumors revealed different dynamics in tumor growth. Specifically, *Veillonella parvula* instillation caused disruption of the lung microbiota and did not affect the growth of individual tumors (Figure 6b), but this difference was not statistically significant.

4. Discussion

The main causative factors of lung cancer are exposure to carcinogenic substances such as cigarette smoke, chronic airway inflammation, lung fibrosis, etc [15,16]. These pathogenic factors may alter the lung microbiota and are associated with lung cancer progression, phenotype, and severity [17,18]. However, little research has been conducted on the direct relationship between lung cancer and the lung microbiota. Bronchoalveolar

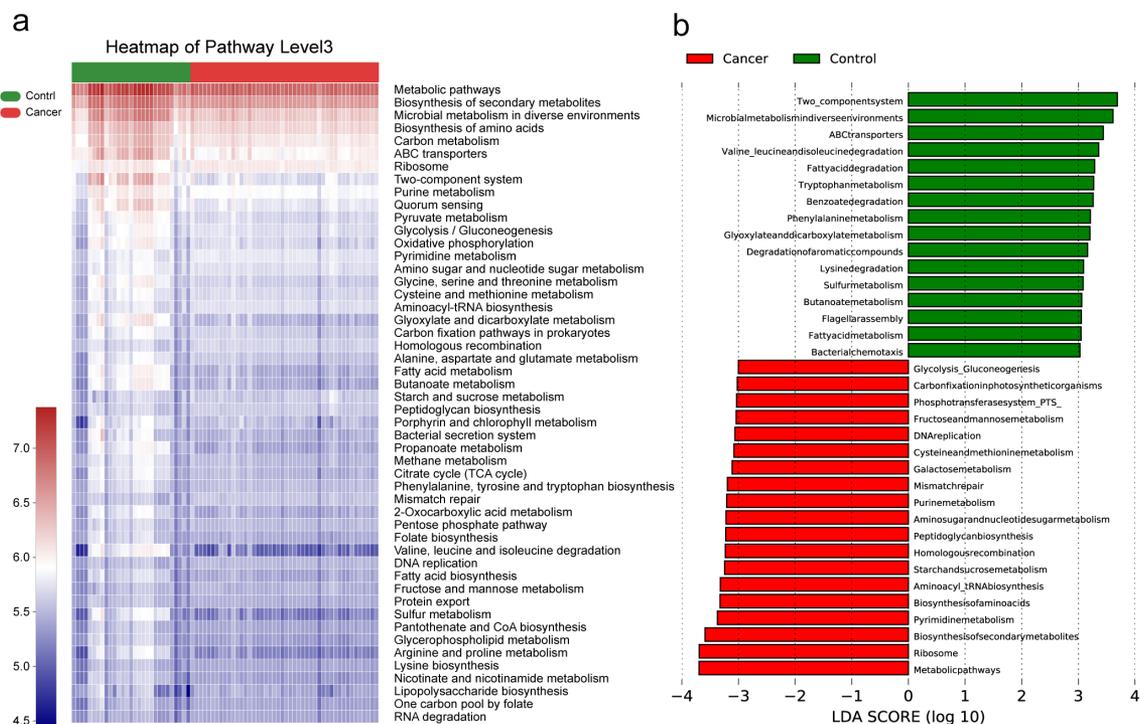


Figure 5. Predicting gene function in the lung microbiota using PICRUSt2. The impact of differentially enriched KEGG pathways between lung cancer and the control group is shown using a heat map (a) and evaluated through the LDA score. Only KEGG pathways meeting an LDA score >3.0 are shown (b). The columns in red and blue represent the cancer and control groups, respectively.

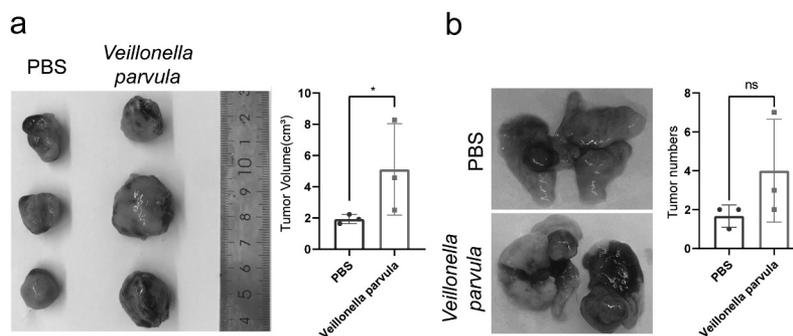


Figure 6. *Veillonella* promotes the growth of Lewis lung cancer (LLC) tumors *in vivo*. (a) C57BL/6 mice after subcutaneous transplantation with LLC cells, *Veillonella parvula* injections around the tumor twice a week after tumor cell engraftment; PBS was used as control. Data were analyzed using student's test ($n = 3$ per group). (b) C57BL/6 mice after orthotopic transplantation with LLC cells, and intratracheal infusion of *Veillonella parvula* twice a week after tumor cell engraftment; PBS was used as control. Data were analyzed using student's test ($n = 3$ per group). * $P < 0.05$; ns, not significant.

lavage, bronchial brushing tissue, buccal samples, surgical resection tissue, or exhaled breath condensate are used for the common specimens characterizing the lung microbiota, and although the microbiota of the lower and upper respiratory tracts are highly similar [19,20], the dysregulated characteristics of the lower respiratory microbiota are more suitable for distinguishing lung cancer from benign lung disease [21].

Comparing the microbiota of BALF samples between patients with NSCLC and control patients, the lung microbiota of NSCLC showed a more complex diversity with higher abundance and α -diversity. Laroumagne et al. [22] showed that the abundance and diversity of the microbiota in the lower lung in patients with lung cancer were significantly reduced compared to healthy individuals. Peters et al. showed that the α -diversity of the microbiota was higher in primary or recurrent lung cancer than in normal lung tissue [23,24]. In contrast, Sang [25] found no difference in α - and β -diversity between the two groups. This difference in findings may be due to differences related to geographical differences, type, and the number of samples collected, as well as the analysis of sequencing data. In the current study, ANOSIM further confirmed the significant differences in the microbiota between lung cancer lesions and controls. Thus, dysregulation in the lung microbiota occurs during NSCLC development.

Among the lung microbiota, Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria were the four most common core phyla, while

Prevotella, *Streptococcus*, *Veillonella*, *Neisseria*, *Haemophilus*, and *Fusobacterium* were the most abundant genera [26,27]. In a study on postoperative lung tissue microbiota characterization, lung cancer tissue lower airway brush samples ($n = 39$) showed a stronger enrichment in *Veillonella* and *Streptococcus* than did the benign disease ($n = 36$) and healthy control samples ($n = 10$) [28]. In a Korean study, the abundance of *Veillonella* and *Megasphaera* was greater in BALF samples from patients with lung cancer than in those with benign lung lesions ($n = 8$) [25]. Similar to the previous study, in our study, the lung microbiota in the NSCLC group was enriched at the phylum level for Firmicutes and Bacteroidetes, and the top three enriched genera included *Streptococcus*, *Prevotella*, and *Veillonella*. Furthermore, in the NSCLC group samples compared to those of the control group, *Prevotella* abundance was increased from 1.93% to 17.73%, that of *Veillonella* from 1.76% to 7.62%, and that of *Streptococcus* from 4.01% to 15.76%. The conclusions differ from the Markus Hilty study, in which *Veillonella* and *Prevotella* were expressed in higher abundance in healthy lungs controls in those with non-lung cancer disease [29]. Although some differences exist, these studies suggest that changes in the dynamics of the bacterial phyla Firmicutes and Bacteroidetes and the genera *Streptococcus*, *Veillonella*, and *Prevotella* are closely associated with the development of diseases, including lung cancer.

LEfSe analysis was applied to identify species that were different between groups, and LDA was

further used to evaluate the effect of these differences in species on disease progression. Previous studies have also been controversial, showing that Firmicutes was enriched and Proteobacteria reduced in lung cancer tissues compared to emphysema samples [30], whereas Proteobacteria enrichment was present in most lung tumors in smokers [31]. Unlike studies on the relationship between *Clostridium perfringens* and colon cancer, there is a lack of studies on the influence of a single genus of bacteria on lung cancer progression. Tsay et al [21] administered commensal bacteria (*Veillonella*) in a KP mouse model and induced upregulation of Erk/MAPK and inflammatory signaling pathways in respiratory epithelial cells and an increase in Th17 cells through the constructed dysbiosis environment, which in turn [10] significantly promoted the progression of lung cancer when bacteria (*Herbaspirillum* and *Sphingomonadaceae*) isolated from advanced tumors in SPF mice were re-transplanted back into the intestine.

Prevotella is an important pathogen of periodontitis and secretes peptides through proteinase-activated receptors (PARs), which are involved in the regulation of proliferation, apoptosis, immunity, cytokine production, and microenvironmental inflammation in oral squamous cell carcinoma [32,33]. Gram-positive (*Streptococcus*) and gram-negative (*Escherichia coli*) bacteria can promote distant metastasis of NSCLC through activation of Toll-like receptors 2/4 and interleukin-6 secretion [34,35]. *Veillonella parvula*, the most commonly known species of the *Veillonella* genus, was found to significantly promote the progression of LLC in mice, consistent with *Veillonella parvula* promoting the formation of lung adenocarcinoma in KPA mice. Thus, this suggests the role and potential mechanisms by which changes in the abundance of core genera affect lung cancer progression. In addition, the role of a smaller proportion of groups in lung cancer should not be ignored as well; further research is needed to explore more causative organisms and their mechanisms of action.

Bacteria coexist in complex networks of interactions, and interactions within these networks affect the species involved, which in turn leads to disease. As shown in our co-expression network

relationship diagram, the microbiota in NSCLC exhibit complex ecological relationships, with *Prevotella*, *Alloprevotella*, and *Veillonella* clustering together to form the densest network of interactions. Similar results to those of LEfSe analysis of key genera further suggest a key role for *Prevotella* and *Veillonella* in the progression of NSCLC, which could distinguish lung cancer from normal samples. Based on 16s rRNA sequencing information, the PICRUSt 2 method was used to make inferential predictions about the function of the bacterial community, focusing mostly on metabolic pathways, ribosome, biosynthesis of secondary metabolites, and microbial metabolism in LEfSe's analysis of certain gene functions in the community; among the metabolic pathways enriched in NSCLC, gluconeogenesis, amino acid metabolism, and inflammatory signaling pathways were similarly enriched in chronic obstructive pulmonary disease, a risk factor for lung carcinogenesis [36]. In contrast, cell growth, death, reduced transport, and catabolism are associated with reduced transport and catabolic pathways of bacterial carbohydrate metabolism [37]. If more detailed information on the functional changes of the bacteriophage genes is needed, macrogenome sequencing should be performed for further analysis.

Limitations

This research is a cross-sectional study and does not establish a causal relationship between microbiota and lung cancer, which must be further confirmed by more mechanistic studies on animal and cellular aspects. Furthermore, we hypothesize that *Prevotella* and *Veillonella* are closely associated with NSCLC, but their relevance to clinical and prognostic aspects requires more clinical samples and information for analysis.

Conclusion

Significant changes in bacterial composition and bacterial gene function were found in NSCLC. In particular, Firmicutes, Bacteroidetes, Fusobacteriota at the phylum level, and *Prevotella*, *Alloprevotella*, and *Veillonella* at the genus level were significantly enriched in NSCLC samples. We also confirmed that

Veillonella (Parvula) plays the role of a carcinogenic bacteria for lung cancer *in vivo*. These important genera may be potential diagnostic markers and therapeutic targets.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

Data will be provided on reasonable request

Contributions

(I) Conception and design: W Zeng and SD Ma; (II) Administrative support: CZ Zhao; (III) Provision of study materials or patients: HL Chen and YY Pan; (IV) Collection and assembly of data: W Zeng and CZ Zhao; (V) Data analysis and interpretation: W Zeng, CZ Zhao, and MG Yu; (VI) Manuscript writing: All authors; (VII) Final approval of the manuscript: All authors.

Funding

This study was supported by the Natural Science Foundation of Guangdong Province[#1] under Grant (Number. 2021A151011719); Ganzhou Municipal Health Care Committee Municipal Research Program Project[#2] under Grant (Number. Ganzhou Science and Technology Plan Project 2020-2-47).

References

- [1] Salloum RG, Braithwaite D. Expansion of guideline-recommended lung cancer screening eligibility: implications for health equity of joint screening and cessation interventions. *J Thorac Oncol.* **2022**;17(1):13–15.
- [2] Zhang T, Joubert P, Ansari-Pour N, et al. Genomic and evolutionary classification of lung cancer in never smokers. *Nat Genet.* **2021**;53(9):1348–1359.
- [3] Terzikhan N, Verhamme KMC, Hofman A, et al. Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study. *Eur J Epidemiol.* **2016**;31(8):785–792.
- [4] Durham AL, Adcock IM. The relationship between COPD and lung cancer. *Lung Cancer.* **2015**;90(2):121–127.
- [5] Collisson EA, Campbell JD, Brooks AN *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* **2014**;511(7511):543–550.
- [6] Khatoon J. Role of *Helicobacter pylori* in gastric cancer: updates. *World J Gastrointest Oncol.* **2016**;8(2):147–158.
- [7] Garrett WS. The gut microbiota and colon cancer. *Science.* **2019**;364(6446):1133–1135.
- [8] Parida S, Wu S, Siddharth S, et al. A procarcinogenic colon microbe promotes breast tumorigenesis and metastatic progression and concomitantly activates notch and β -catenin axes. *Cancer Discov.* **2021**;11(5):1138–1157.
- [9] Lloyd CM, Marsland BJ. Lung homeostasis: influence of age, microbes, and the immune system. *Immunity.* **2017**;46(4):549–561.
- [10] Jin C, Lagoudas GK, Zhao C, et al. Commensal microbiota promote lung cancer development via $\gamma\delta$ T cells. *Cell.* **2019**;176(5):998–1013.e16.
- [11] Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods.* **2013**;10(10):996–998.
- [12] Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* **2019**;37(8):852–857.
- [13] Chu DM, Ma J, Prince AL, et al. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat Med.* **2017**;23(3):314–326.
- [14] Jalili V, Afgan E, Gu Q, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* **2020**;48(W1):W395–w402.
- [15] Eapen MS, Hansbro PM, Larsson-Callertfelt A-K, et al. Chronic Obstructive Pulmonary Disease and Lung Cancer: underlying Pathophysiology and New Therapeutic Modalities. *Drugs.* **2018**;78(16):1717–1740.
- [16] Shiels MS, Albanes D, Virtamo J, et al. Increased risk of lung cancer in men with tuberculosis in the alpha-tocopherol, beta-carotene cancer prevention study. *Cancer Epidemiol Biomarkers Prev.* **2011**;20(4):672–678.
- [17] Carbone C, Piro G, Di Noia V, et al. Lung and gut microbiota as potential hidden driver of immunotherapy efficacy in lung cancer. *Mediators Inflamm.* **2019**;2019:7652014.
- [18] Xiong K, Sun W, He Y, et al. Advances in molecular mechanisms of interaction between Mycobacterium tuberculosis and lung cancer: a narrative review. *Transl Lung Cancer Res.* **2021**;10(10):4012–4026.
- [19] Chau J, Yadav M, Liu B, et al. Prospective correlation between the patient microbiome with response to and development of immune-mediated adverse effects to immunotherapy in lung cancer. *BMC Cancer.* **2021**;21(1):808.
- [20] Reinhold L, Möllering A, and Wallis S, et al. Dissimilarity of airway and lung tissue microbiota in

- smokers undergoing surgery for lung cancer. *Microorganisms*. 2020;8(6):794. doi:10.3390/microorganisms8060794.
- [21] Tsay J-CJ, Wu BG, Sulaiman I, et al. Lower airway dysbiosis affects lung cancer progression. *Cancer Discov*. 2021;11(2):293–307.
- [22] Laroumagne S, Lepage B, Hermant C, et al. Bronchial colonisation in patients with lung cancer: a prospective study. *Eur Respir J*. 2013;42(1):220–229.
- [23] Peters BA, Hayes RB, Goparaju C, et al. The microbiome in lung cancer tissue and recurrence-free survival. *Cancer Epidemiol Biomarkers Prev*. 2019;28(4):731–740.
- [24] Greathouse KL, White JR, Vargas AJ, et al. Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol*. 2018;19(1):123.
- [25] Lee SH, Sung JY, Yong D, 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.
- [26] Dickson RP, Erb-Downward JR, Martinez FJ, et al. The microbiome and the respiratory tract. *Annu Rev Physiol*. 2016;78(1):481–504.
- [27] Stasiewicz M, Karpiński TM. The oral microbiota and its role in carcinogenesis. *Semin Cancer Biol*. 2021. DOI:10.1016/j.semcancer.2021.11.002
- [28] Tsay J-CJ, Wu BG, Badri MH, et al. Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer. *Am J Respir Crit Care Med*. 2018;198(9):1188–1198.
- [29] Hilty M, Burke C, Pedro H, et al. Disordered microbial communities in asthmatic airways. *PLoS One*. 2010;5(1):e8578.
- [30] Liu Y, O'Brien JL, Ajami NJ, et al. Lung tissue microbial profile in lung cancer is distinct from emphysema. *Am J Cancer Res*. 2018;8(9):1775–1787.
- [31] Nejman D, Livyatan I, Fuks G, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science*. 2020;368(6494):973–980.
- [32] Eftekhari R, De Lima SG, Liu Y, et al. Microenvironment proteinases, proteinase-activated receptor regulation, cancer and inflammation. *Biol Chem*. 2018;399(9):1023–1039.
- [33] Van Spaendonk H, Ceuleers H, Witters L, et al. Regulation of intestinal permeability: the role of proteases. *World J Gastroenterol*. 2017;23(12):2106–2123.
- [34] Gowing SD, Chow SC, Cools-Lartigue JJ, et al. Gram-positive pneumonia augments non-small cell lung cancer metastasis via host toll-like receptor 2 activation. *Int J Cancer*. 2017;141(3):561–571.
- [35] Gowing SD, Chow SC, Cools-Lartigue JJ, et al. Gram-negative pneumonia augments non-small cell lung cancer metastasis through host toll-like receptor 4 activation. *J Thorac Oncol*. 2019;14(12):2097–2108.
- [36] Millares L, Pérez-Brocal V, Ferrari R, et al. Functional metagenomics of the bronchial microbiome in COPD. *PLoS One*. 2015;10(12):e0144448.
- [37] Zheng X, Liu R, Zhou C, et al. ANGPTL4-Mediated Promotion of Glycolysis Facilitates the Colonization of *Fusobacterium nucleatum* in Colorectal Cancer. *Cancer Res*. 2021;81(24):6157–6170.