



## OPEN ACCESS

## EDITED BY

Eva Maria Weissinger,  
Hannover Medical School, Germany

## REVIEWED BY

José Patané,  
Butantan Institute, Brazil  
Xiaoqin Wei,  
University of Virginia, United States

## \*CORRESPONDENCE

Baosong Xie  
✉ baosongx@fjmu.edu.cn  
Xiujuan Yao  
✉ 953381422@qq.com  
Ruotong Ren  
✉ ruorabbit@163.com  
Wenjie Wu  
✉ jjezi\_1@126.com

†These authors have contributed  
equally to this work and share  
first authorship

RECEIVED 24 October 2023

ACCEPTED 08 October 2024

PUBLISHED 31 October 2024

## CITATION

Hong R, Lin S, Zhang S, Yi Y, Li L, Yang H,  
Du Z, Cao X, Wu W, Ren R, Yao X and Xie B  
(2024) Pathogen spectrum and microbiome  
in lower respiratory tract of patients with  
different pulmonary diseases based on  
metagenomic next-generation sequencing.  
*Front. Cell. Infect. Microbiol.* 14:1320831.  
doi: 10.3389/fcimb.2024.1320831

## COPYRIGHT

© 2024 Hong, Lin, Zhang, Yi, Li, Yang, Du, Cao,  
Wu, Ren, Yao and Xie. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Pathogen spectrum and microbiome in lower respiratory tract of patients with different pulmonary diseases based on metagenomic next-generation sequencing

Rujun Hong<sup>1†</sup>, Sheng Lin<sup>1†</sup>, Siting Zhang<sup>1†</sup>, Yaxing Yi<sup>2†</sup>,  
Lanfeng Li<sup>1</sup>, Haitao Yang<sup>1</sup>, Zhenshan Du<sup>2</sup>, Xuefang Cao<sup>2</sup>,  
Wenjie Wu<sup>2\*</sup>, Ruotong Ren<sup>2,3\*</sup>, Xiujuan Yao<sup>1\*</sup> and Baosong Xie<sup>1\*</sup>

<sup>1</sup>Department of Pulmonary and Critical Care Medicine, Fujian Provincial Hospital, Fujian Provincial Clinical Medical College, Fujian Medical University, Fuzhou, China, <sup>2</sup>Medical Department, Matridx Biotechnology Co., Ltd, Hangzhou, China, <sup>3</sup>Institute of Biophysics, Chinese Academy of Sciences, Beijing, China

**Introduction:** The homeostasis of the microbiome in lower respiratory tract is crucial in sustaining normal physiological functions of the lung. Different pulmonary diseases display varying degrees of microbiome imbalance; however, the specific variability and clinical significance of their microbiomes remain largely unexplored.

**Methods:** In this study, we delineated the pathogen spectrum and commensal microorganisms in the lower respiratory tract of various pulmonary diseases using metagenomic sequencing. We analyzed the disparities and commonalities of the microbial features and examined their correlation with disease characteristics.

**Results:** We observed distinct pathogen profiles and a diversity in lower airway microbiome in patients diagnosed with cancer, interstitial lung disease, bronchiectasis, common pneumonia, Nontuberculous mycobacteria (NTM) pneumonia, and severe pneumonia.

**Discussion:** This study illustrates the utility of Metagenomic Next-generation Sequencing (mNGS) in identifying pathogens and analyzing the lower respiratory microbiome, which is important for understanding the microbiological aspect of pulmonary diseases and essential for their early and precise diagnosis.

## KEYWORDS

pulmonary disease, lower respiratory tract, mNGS, pathogen, microbiome

## 1 Introduction

Serving as a respiratory organ, healthy lungs facilitate the exchange of O<sub>2</sub>/CO<sub>2</sub> between blood and ambient air, which is crucial for sustaining the body's normal physiological functions. Historically, the understanding of the microbiome in healthy lungs was limited, with a prevailing belief that they are sterile (Dickson et al., 2013). This belief stemmed from the lack of invasive sampling of the healthy lungs and methodological constraints, such as the difficulties to culture fastidious organisms. In 2010, Markus Hilty et al. utilized 16s rRNA sequencing to confirm the presence of bacteria in healthy lungs, identifying bacterial genera such as *Prevotella*, *Veronella*, *Streptococcus*, and *Haemophilus* in cytological brushes from the left upper lobe (~2000 bacterial genomes/cm<sup>2</sup> surface area) (Hilty et al., 2010). In recent years, the advances of culture-independent molecular diagnostic methods, particularly high-throughput sequencing technologies, have significantly enhanced the identification of a broad range of pathogens in a variety of biological specimens. Studies utilizing nucleic acid sequencing of bronchoalveolar lavage fluid (BALFs) have confirmed the presence of fungi, such as *Cladosporium* and *Aspergillus* spp. and viruses, including anellovirus, Gardnerella phages and Lactobacillus phages (Wilson et al., 2019; Tian et al., 2022). Therefore, it is evident that the lung harbors a microbiome. Being connected to the external environment, human lungs experience a constant exchange of microorganisms through the upper respiratory tract (oral and nasal cavities, pharynx, and trachea) with each breath, leading to a dynamic microbial cycle and ever-changing microbiome (Natalini et al., 2023). Conversely, the microbiota absorbs nutrients and necessitates dynamic microbiological diversity, self-renewal, and maintenance from both the air phase (airways) and the fluid/cellular phase (*i.e.*, alveoli and their cellular components) within the biological niche of the lungs. The homeostasis of the pulmonary microbiome is critical for the normal physiological function of the lungs. Prior research has indicated that various pulmonary diseases, including lung cancer (Tsay et al., 2018), infectious pneumonia, interstitial pneumonia, and non-infectious diseases such as bronchiectasis (Cox et al., 2017), display alterations in the microbiome, including both pathogenic and commensal organisms of the lower respiratory tract. On the other hand, a close association was observed between pulmonary microbiome and the development and progression of various respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), cystic fibrosis (CF), non-CF bronchiectasis, tuberculosis, COVID-19, and lung cancer (Chung, 2017; Hong et al., 2018; Ding et al., 2021; Llorens-Rico et al., 2021; Ramshah et al., 2021). However, the similarities and differences in the microbiome across diseases and their clinical relevance remain inadequately studied.

In this study, we analyzed data from patients diagnosed with various lung diseases, admitted to the Department of Respiratory Medicine at Fujian Provincial Hospital between 2019-2022. We collected BALF, peripheral blood, and tissue samples (lung biopsies obtained from lesions), along with the conventional microbiological test results, routine clinical test results. Using mNGS on BALF, we examined the spectrum of microorganisms present in the lower

respiratory tract. Meanwhile, we analyzed the microbiome composition and its association with the pulmonary disease. This study indicates that mNGS is effective in identifying pathogens and microbiome of the lower respiratory tract. Furthermore, distinct pulmonary diseases are characterized by unique pathogen and microbiome profiles. It enhances our understanding of the pathology underlying various pulmonary diseases and may benefit diagnosis and therapeutic approaches of these diseases.

## 2 Materials and methods

### 2.1 Patient enrollment

This retrospective study included 305 clinical samples from 208 patients, diagnosed with various lung diseases in the intensive care unit (ICU) of Fujian Provincial Hospital between 12 June 2019 and 19 July 2022. The types of diseases included interstitial pneumonia, lung cancer, common or severe pneumonia, NTM pneumonia, and bronchiectasis. Data including age, sex, exposure history, comorbidities, onset, symptoms, imaging, laboratory tests, diagnostic methods, treatment, and clinical outcomes, were extracted from electronic medical records. The Ethics Committee of Fujian Provincial Hospital approved the study, and all data were anonymized before analysis. The study adhered to the Declaration of Helsinki, with data procured from the General ICU of Fujian Provincial Hospital.

### 2.2 Clinical sample collection and DNA extraction

Bronchoalveolar lavage fluid (BALF), sputum, peripheral blood and other samples were obtained from each patient, following the acquisition of consent from either the patients themselves or their surrogates. Experienced bronchoscopists collected the BALF samples after administration of anesthesia with midazolam. Peripheral blood underwent centrifugation at 1600g for 10 minutes, followed by a further centrifugation of the supernatant at 16000g for 10 minutes to isolate plasma. For other samples, genomic DNA was extracted from 1 mL of specimens. The resulting DNA underwent library preparation (enzymatic fragmentation of genomic DNA, end repairing, terminal adenylation and adaptor ligation) and purification. All steps were performed according to a previous study (Luan et al., 2021). DNA extraction and library preparation from clinical samples were conducted utilizing a point-of-care automation device (Matridx Biotechnology Co., Ltd, Hangzhou, China) (Luan et al., 2021). The quality of extracted DNAs was evaluated with a BioAnalyzer 2100 (Agilent Technologies; Santa Clara, CA, United States), in conjunction with quantitative PCR, to assess the adapters prior to sequencing. The name of the kit used for NGSmaster was Matridx Biotechnology's Next-Generation Sequencing Reagent Kit (Cat. No. CW0531M). For DNA extraction, we used a kit from Matridx, Cat. No. MAR002 and followed SOPs provided by the manufacturer.

## 2.3 Metagenomic next-generation sequencing

Qualified DNA libraries were combined and sequenced using the Illumina NextSeq500 system (50 bp single end; San Diego, CA, United States). Each run included one negative control, consisting of artificial plasma mixed with fragmented human genomic DNA, and one positive control, comprising a mixture of inactivated bacteria, fungi, and pseudoviral particles containing synthesized DNA or RNA fragments of adenovirus and influenza A virus, for quality control. Each sample generated a total of 10 - 20 million reads. Initially, raw sequenced reads underwent quality control processing to eliminate short (length < 35 bp), low-quality, and low complexity reads, along with the adapter sequences. Sequences from the host were excluded by aligning them to the human-specific database in NCBI (GRCh38.p13), utilizing Bowtie2 (version 2.3.5.1). Clean reads were then aligned to a curated in-house microbial database, which incorporated sequences from the NCBI GenBank nucleotide (nt) database and assembly database, as well as sequences assembled from our own pure fungal cultures, with Kraken2 (version 2.1.2; confidence = 0.5) for rapid taxonomic classification. The aligned microbial reads underwent further validation through a secondary alignment to the microbial database, utilizing Bowtie2. When inconsistencies arose between the results of Kraken2 and Bowtie2, the classification of reads was determined using BLAST (version 2.9.0) (Altschul et al., 1990; Langmead and Salzberg, 2012; Wood and Salzberg, 2014). The parameters and thresholds used in BLAST included: E-value threshold: 1e-5; Identity cutoff: 90%; Alignment length cutoff: 100 base pairs. These thresholds were selected to ensure high confidence matches and to minimize false positives. Prior to data analysis, microbes identified in clinical samples were compared to those detected in NTC (no template control). Microorganisms with reads per million (RPM) above 10, or those not detected in NTC, were retained for subsequent analysis. Essentially, all microbial species were searched in PubMed to determine whether the organisms can cause pneumonia. If yes, the microorganisms were classified as pathogens.

## 2.4 Pathogen reporting criteria

Microbial reads identified from a library were reported if: 1) the sequencing data passed quality control filters (library concentration > 50 pM, Q20 > 85%, Q30 > 80%); 2) negative control (NC) in the same sequencing run does not contain the species or the RPM (sample)/RPM (NC)  $\geq$  5, which was determined according to previous studies as a cutoff for discriminating true-positives from background contaminations (Schlaberg et al., 2017; Wilson et al., 2019; Luan et al., 2021).

## 2.5 Statistical analysis

Categorical variables were represented as frequencies and percentages and compared utilizing Fisher's exact test. Continuous data with normal distribution were represented as mean  $\pm$  standard

deviation ( $x \pm s$ ), while data with non-normal distribution were represented by median (range). The Wilcoxon test or Kruskal-Wallis test was employed to calculate differences and significance for non-normal distribution data. Statistical analysis was conducted using SPSS 26.0 (IBM Corporation). R (Version 4.2.1) was utilized for data visualization. Specifically, unsupervised clustering methods, referencing the core steps of limma, voom, fit, eBayes were employed for bivariate or multivariate difference analysis. The limma package's plotMDS illustrated the final similarities (or differences) between samples, and results were output through the topTable method, sorted by P-value. Before analyzing their relative abundance, RPKM values of microbes underwent log<sub>2</sub> transformation. The limma package was utilized to analyze variations in the composition and abundance of microbes between groups. Particularly, the FDR (False Discovery Rate) method was employed to correct the primary P-value for multiple comparisons. Subsequently, corrected P-values  $\leq$  0.05 were deemed statistically significant.

## 3 Results

### 3.1 Microbial detection in lower respiratory tract samples from patients with different lung diseases

In this retrospective study, we collected data from 277 patients diagnosed with various pulmonary diseases at the Department of Respiratory Medicine of Fujian Provincial Hospital between 2019 and 2022. A total of 208 patients met the inclusion criteria. The patients were categorized into six groups based on the diagnosis: lung cancer (CA), interstitial lung disease (IP), bronchiectasis (BRO), common pneumonia (PN), NTM pneumonia (NTM), and severe pneumonia (SP) (Figure 1A). Of the 208 patients, 185 provided BALF samples only and 21 patients had two types of samples including BALF and others such as peripheral blood, tissue, and pleural fluid. The sample distribution of the six groups were shown in Supplementary Figure S1. Additionally, some patients had BALF collected at two different time points, leading to a total of 305 samples (Figure 1B). We also obtained results from conventional microbiological tests, routine tests, and clinical consultations.

The overall positivity rate of mNGS for pathogen detection was 78.36%, with 80.17%, 76.19%, 60.0%, and 63.16% for lower respiratory tract specimens (LRTS), peripheral blood, pleural fluid, and tissue samples, respectively (Figure 1C). Little overlap was observed among the pathogens identified in peripheral blood, pleural fluid, and tissue samples. In contrast, pathogens identified in BALF showed a higher degree of concordance with those in peripheral blood and tissue samples, at 28.7% and 29.6% respectively (Figure 1D).

The mNGS results were compared with the conventional microbiological tests (CMTs). CMTs identified 31 distinct pathogens for a total of 106 times, in contrast to mNGS that reported 110 distinct pathogens for a total of 630 times (Figure 2A). Subsequently, we ranked the top 10 pathogens based on frequency by both mNGS and CMTs and three species were in common: *Candida albicans*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus* (Figure 2B).

### 3.2 The lower respiratory pathogen spectrum of different pulmonary diseases and clinical relevance

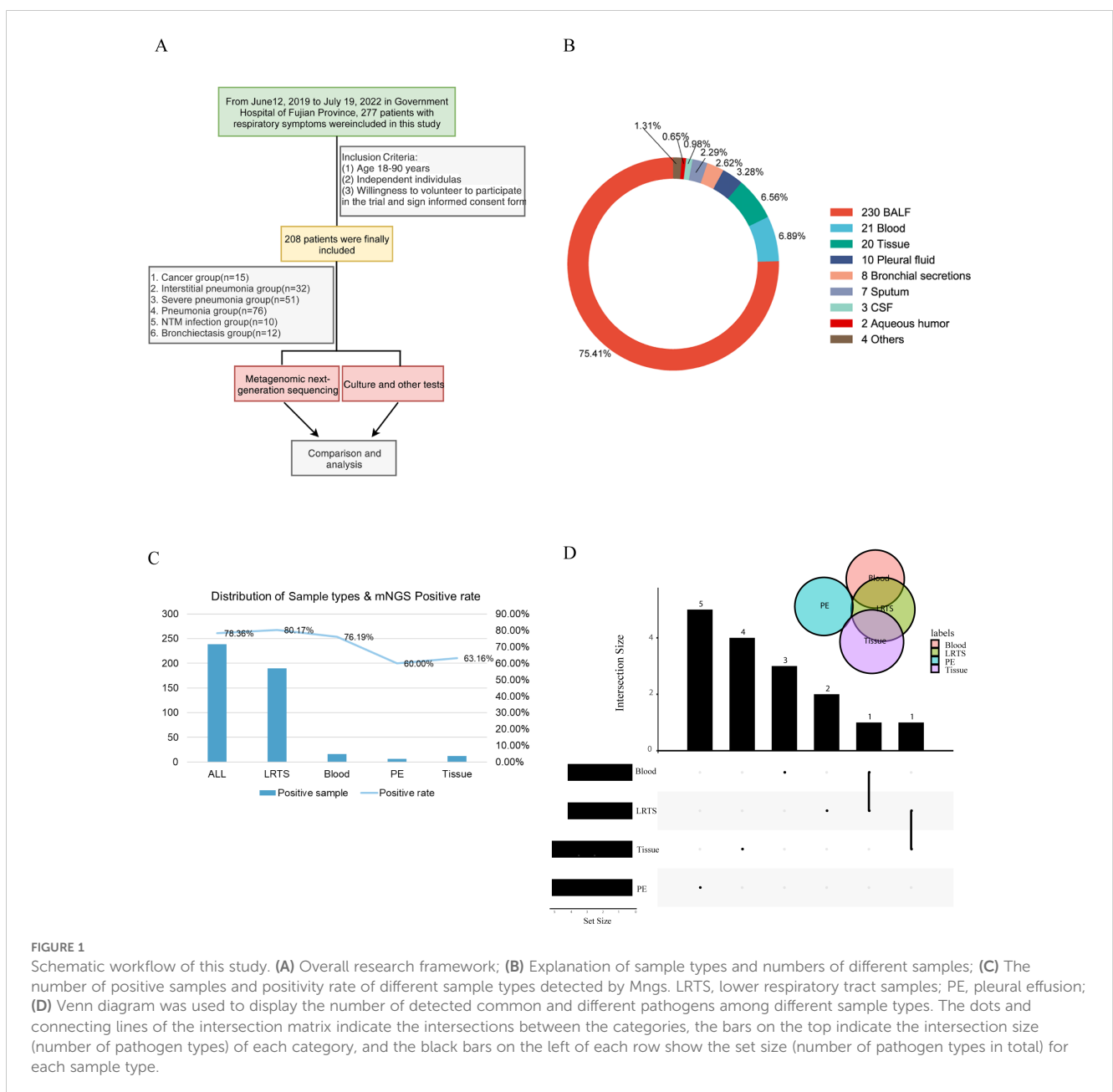
The pathogen profile of different patient groups was analyzed, revealing significant disparities. For instance, the lung cancer, bronchiectasis, and NTM pneumonia groups primarily showed bacterial infections with few viruses detected. The interstitial lung disease group had bacteria, viruses, and fungi (Figures 3A–C). Additionally, it was observed that over 70% of the samples in each group showed mixed infections (Supplementary Figures S2, S3).

Using the Kruskal-Wallis rank sum test, we found enrichments of distinct pathogens in different groups. For instance, the bronchiectasis group showed an enrichment for *Pseudomonas*

aeruginosa, and the detection of *Tropheryma whipplei* was higher in lung cancer than other groups (Figures 3D, E). These findings underscore the varying pathogen profiles associated with different pulmonary diseases and their clinical implications.

### 3.3 Relationship between microbiome and clinical relevance

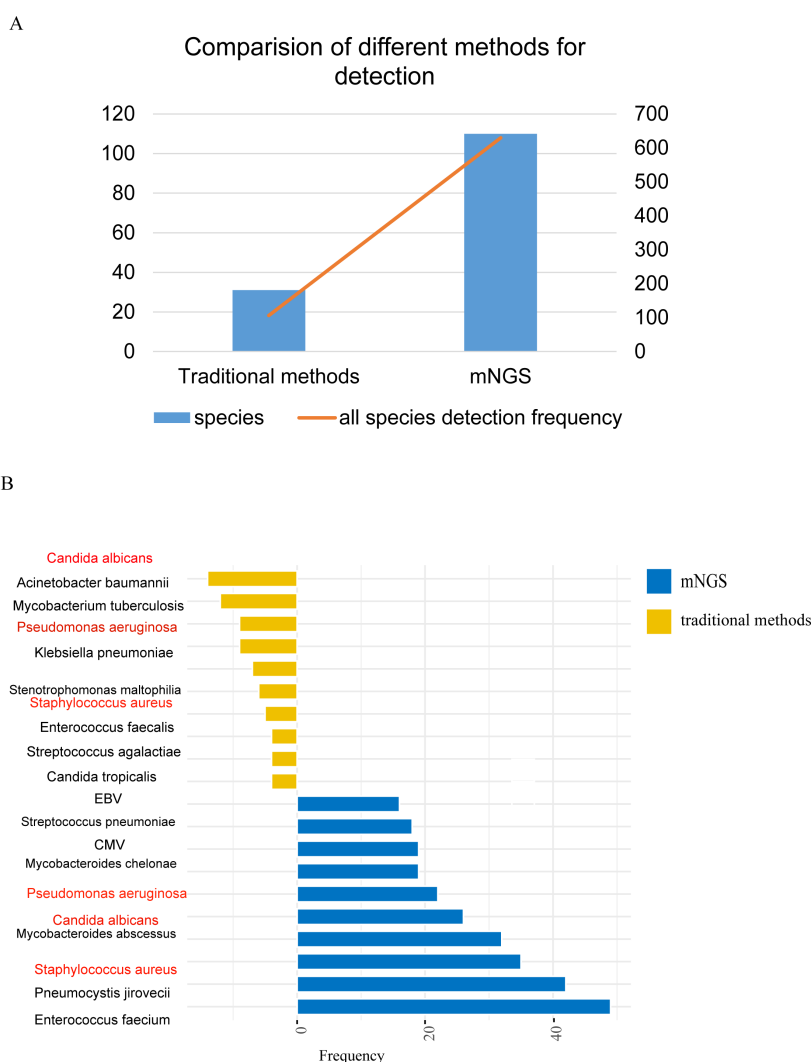
The microbiological findings of mNGS include both pathogenic and commensal microorganisms. Symbiotic microecology can potentially influence the viability of pathogenic entities and may even impact the progression or prognosis of lung diseases. Consequently, we investigated the diversity of organisms in the



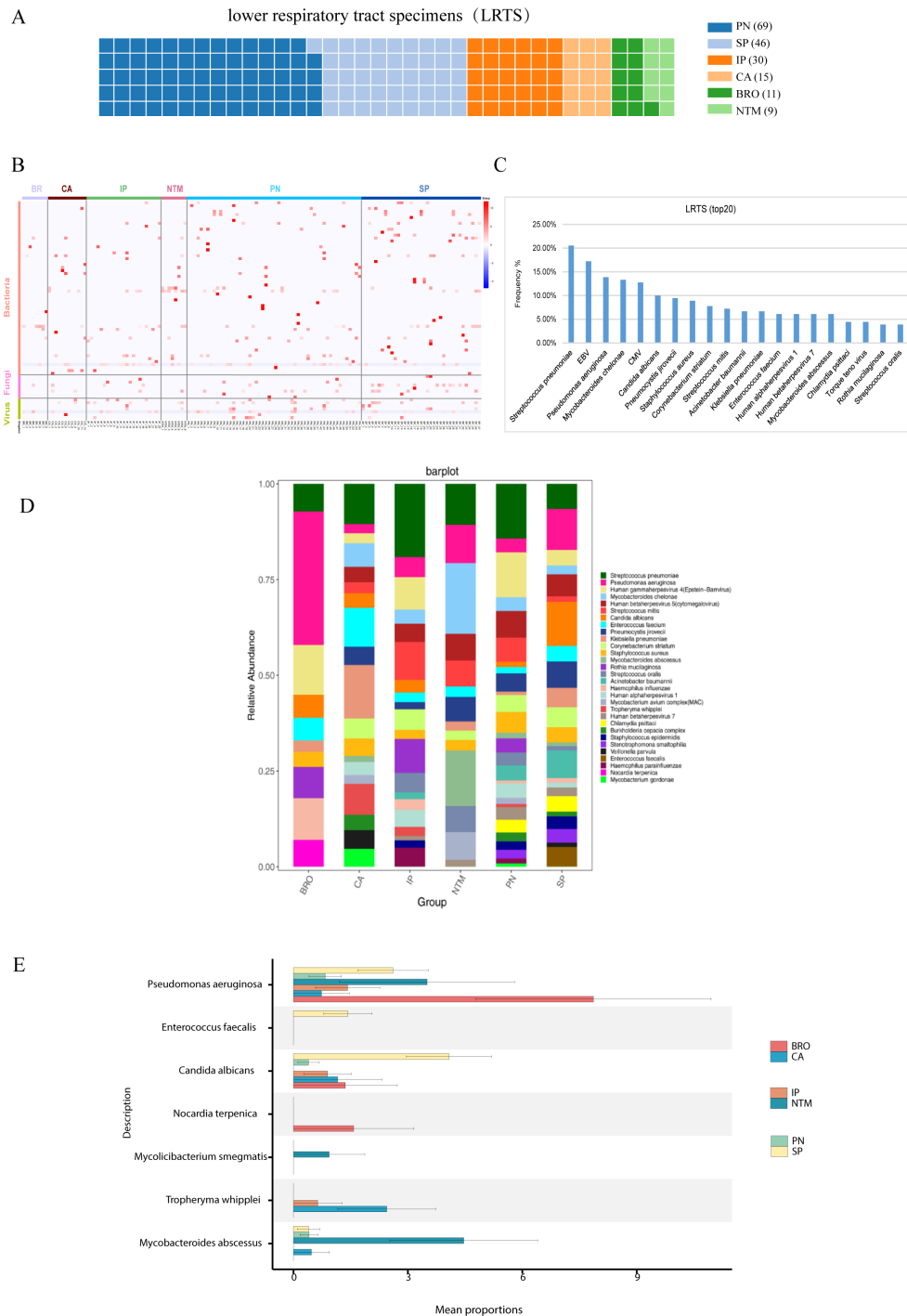
lower respiratory tract across various pulmonary diseases, which revealed distinct compositions and diversities of microorganisms. At the genus level, the lung cancer group was predominantly characterized by *Klebsiella* and *Pseudomonas*, the interstitial lung disease group by *Pseudomonas* and *Candida* spp. In addition, we found enrichment of *Pseudomonas* in the bronchiectasis group, *Klebsiella* and *Pseudomonas* in the common pneumonia group, *Pseudomonas*, *Klebsiella* and *Pseudomonas* in the NTM pneumonia group, *Klebsiella* and *Pseudomonas* in the severe pneumonia group (Figure 4A).

Alpha diversity was evaluated using four indices: Shannon (P=0.00086), Simpson (P=0.00044), Pielou (P=0.0018), and InvSimpson (P=0.00044). The results indicated significant differences in microbial diversity levels among disease groups (Figure 4B). As for beta-diversity, the PERMANOVA analysis indicates that there are statistically significant differences in

community composition among the six patient groups with a p-value of 0.015. The R<sup>2</sup> value of 0.05451 suggests that 5.451% of the total variance is explained by differences between the groups, while 94.549% is attributed to within-group differences. The F-statistic of 1.4297 supports the presence of notable between-group differences. Despite the significant p-value, the relatively low R<sup>2</sup> value indicates that within-group variation plays a major role. This implies that while the groups differ significantly in community composition, individual variability within each group is substantial and warrants further investigation. In addition, to quantify the discriminative power of pathogen profiles among these groups, we performed a bootstrap analysis calculating the area under the curve (AUC) for each disease group. The result revealed distinct discriminatory capabilities among the groups. The 'OTHER' group, which amalgamated NTM pneumonia (NTM), lung cancer (CA), interstitial lung disease (IP), and bronchiectasis (BRO) due to



**FIGURE 2** Comparison of pathogens detected by traditional pathogen detection and mNGS. (A) Comparison of pathogen detection numbers and detection frequencies between traditional pathogen detection methods and mNGS; (B) Top 10 pathogens detected by different detection methods, with red indicating pathogens detected by both methods.



**FIGURE 3** Pathogen profile analysis. **(A)** Analysis of pathogen spectrum in lower respiratory tract samples, and the contribution of lower respiratory tract samples from different groups. Each square refers to one specimen; **(B)** Pathogen distribution diagram, showing the distribution of bacteria, fungi, and viruses in the Lung cancer (CA), interstitial lung disease (IP), bronchiectasis (BRO), common pneumonia (PN), NTM pneumonia (NTM), and severe pneumonia (SP) groups; **(C)** Top 20 detected pathogens and their frequency in lower respiratory tract samples; **(D)** Pathogen stacked bar chart for different groups; **(E)** Comparison of differentially detected pathogens among groups.

limited sample size, exhibited an AUC of 0.8696, suggesting a high accuracy in distinguishing this composite group from others. The AUC of PN and SP groups were 0.7955 and 0.7470, respectively (Supplementary Figure S4). To assess the similarity and disparity of the microbiome, we enumerated the microorganisms found in

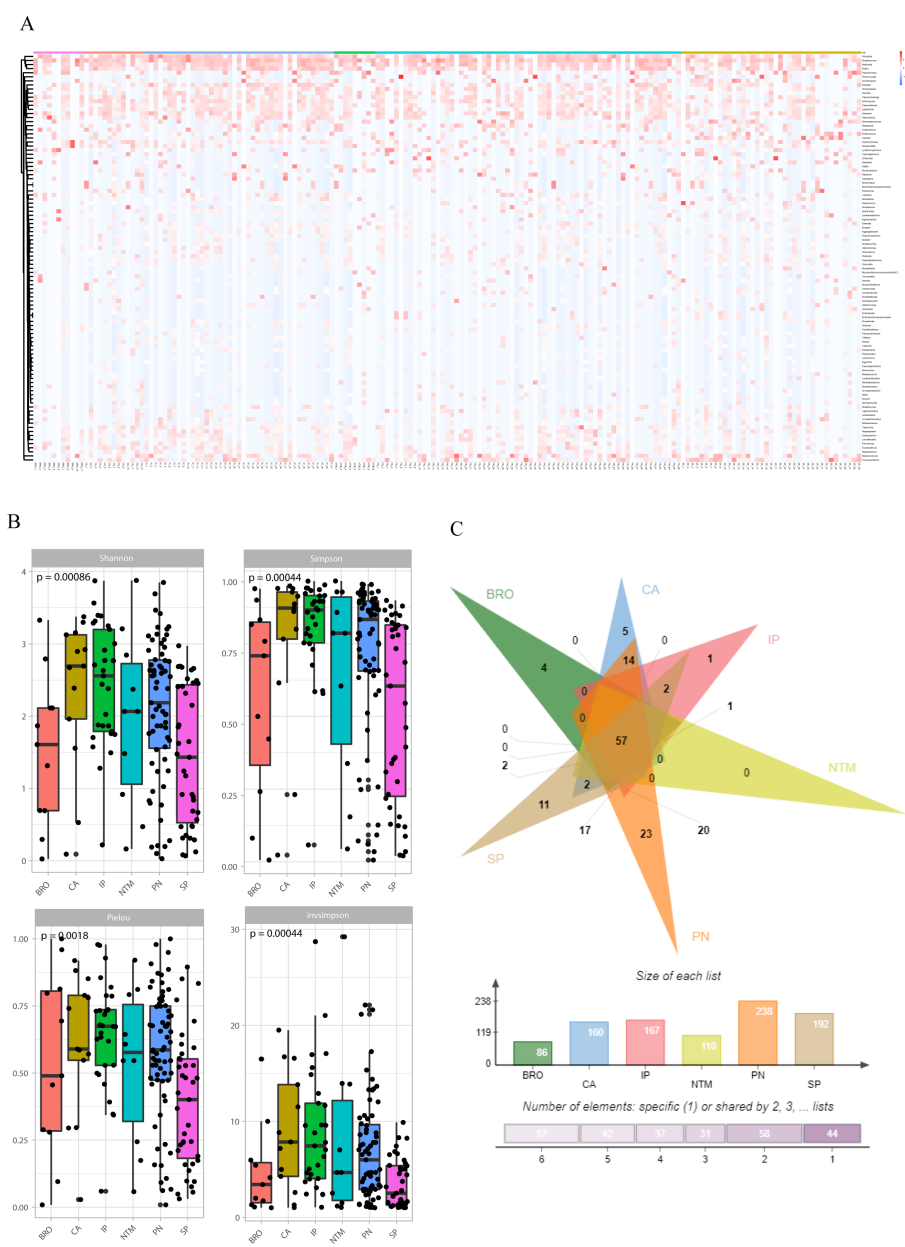
different groups. The results indicated that the common pneumonia group had the highest diversity with 238 species, contrasting with the bronchiectasis group, which had the lowest (86 species). Besides the shared species, unique species were identified in each group: four in the BRO group, 11 in the SP

group, 23 in the PN group, five in the CA group, and one in the IP group. Notably, no species were unique to the NTM group (Figure 4C).

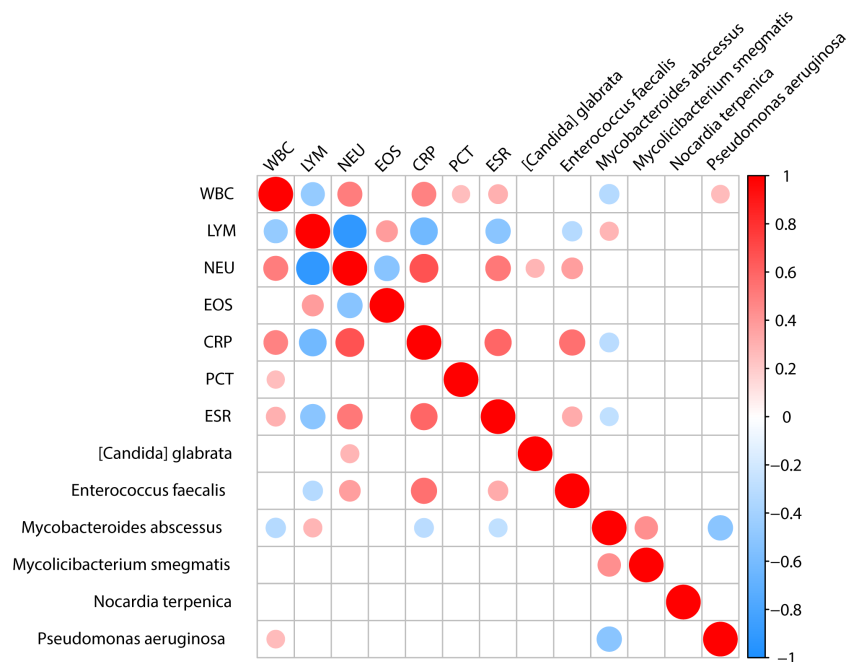
We analyzed the correlation between microorganisms and the clinical test results, such as the biochemical test of blood. We found that *Enterococcus faecalis* showed a positive correlation with the neutrophil count, CRP, and ESR, and a negative correlation with lymphocyte count. *Mycobacteroides abscessus* showed a positive correlation with lymphocyte count and a negative correlation with white blood cell count, CRP, and ESR. *Pseudomonas aeruginosa* exhibited a positive correlation with white blood cell count (Figure 5).

### 3.4 Statistical analysis of age and sex in relation to disease groups

We have performed a logistic regression analysis to evaluate the relationship between age, sex, and disease groups. The results showed that the coefficient for age was -0.07600, with a p-value of 0.08353, indicating a marginal effect on disease group classification that is not statistically significant at the 0.05 level. The negative coefficient suggests a slight decrease in the log-odds of being in a specific disease group with increasing age. Conversely, the coefficient for sex was -1.54491, with a significant p-value of



**FIGURE 4** Symbiotic microbiota research. (A) The distribution of symbiotic microbiota in different disease groups; (B) Analysis of microbiota diversity, alpha diversity research; (C) Venn diagrams show the comparison between different groups, the top shows the overlapping situation of symbiotic microbiota between different groups, and the bottom shows the total number of symbiotic microbiota and the number of unique symbiotic microbiota for each group.



**FIGURE 5** The association between clinical indicators and specific microbiota. Analysis of the correlation between important pathogens and clinical indicators shows that *[Candida] glabrata* is positively correlated with NEU. *Enterococcus faecalis* is positively correlated with NEU, CRP, and ESR clinical indicators while being negatively correlated with LYM. *Mycobacteroides abscessus* is positively correlated with LYM and negatively correlated with WBC, CRP and ESR. *Pseudomonas aeruginosa* is positively correlated with WBC.

0.03920, demonstrating that sex was a statistically significant predictor of disease group classification, with males having lower odds of being classified in certain disease groups compared to females. Combined with the result shown in Table 1, we found that males were more likely to get pneumonia and severe pneumonia than female.

### 4 Discussion

Despite the similarities in clinical presentations of various pulmonary diseases, such as fever, cough, and structural changes of the lung, the underlying pathological mechanisms differ significantly (Yilmaz et al., 2017; Ren et al., 2020; Hoshina and Takei, 2021; Mathew and Mugele, 2021). To study the pathological mechanisms of different pulmonary diseases, researchers typically used *ex vivo* methodologies, including cell and molecular biology as

well as *in vivo* animal models (Liao et al., 2019; Yang et al., 2021). Several studies have explored disease pathology through data mining and correlation analyses (Blanch et al., 2002; Rahman et al., 2022). Conventional microbiological tests have low detection rates, long turnaround time, and inability to identify a broad range of potential pathogens. Owing to advancements in diagnostic technology, mNGS has gained widespread acceptance for the diagnosis of infectious diseases (Fang et al., 2022). It has shown enhanced diagnostic efficacy compared to traditional methods, particularly in identifying rare or emerging pathogens. Therefore, in this study, we employed mNGS to evaluate the microbial profiles from patients with different pulmonary diseases.

Humans and microbes have coexisted in a symbiotic relationship. Termed as the second human genome, the human microbiome encompasses a diverse array of microorganisms, including bacteria, yeasts, archaea, fungi, protozoa, and viruses. These organisms and their byproducts are crucial in modulating and maturing the local

**TABLE 1** The patient information of different disease groups.

| Group                  | Age (years) | Sex            | Underlying disease(s) | Length of hospitalisation (Days) |
|------------------------|-------------|----------------|-----------------------|----------------------------------|
| Cancer                 | 44-79       | M (8), F (7)   | Yes (11), No (4)      | 4-50                             |
| Interstitial pneumonia | 38-81       | M (18), F (15) | Yes (25), No (6)      | 4-34                             |
| Severe pneumonia       | 46-88       | M (38), F (13) | Yes (48), No (3)      | 1-66                             |
| Pneumonia              | 14-86       | M (54), F (22) | Yes (63), No (13)     | 2-46                             |
| NTM infection          | 36-82       | M (5), F (5)   | Yes (6), No (4)       | 4-15                             |
| Bronchiectasis         | 23-82       | M (3), F (9)   | Yes (8), No (4)       | 3-15                             |



microenvironments, including the immune niches of the tissues and organs, and are pivotal in maintaining physiological homeostasis and regulating the functions of organs under pathological conditions (Tsay et al., 2021). Microorganisms inhabit nearly all surfaces of the human body and the lung microbiome typically maintains a dynamic equilibrium. However, this balance is perturbed during diseases, leading to various pathological conditions and clinical manifestations, including infection and inflammation. Consequently, the imbalance and subsequent rebalancing of the lung microbiome may represent pivotal mechanisms governing the pathological conditions and progression of pulmonary diseases (Dickson et al., 2018; Tsay et al., 2018).

The respiratory tract harbors ecological niches populated by commensal and pathogenic microorganisms that are crucial for the progression of diseases (Man et al., 2017; Goeteyn et al., 2023). In the microbiota of healthy human lungs, a variety of microorganisms are typically present, and their distribution and composition can provide insights into respiratory health. In the lungs of healthy individuals, the microbial landscape is predominantly shaped by bacteria, with a few dominant phyla, such as Firmicutes (Streptococcus and Staphylococcus), Proteobacteria (Haemophilus and Pseudomonas), Bacteroidetes (Prevotella and Porphyromonas) and Actinobacteria (Corynebacterium and Mycobacterium) (Dickson and Huffnagle, 2015). Bacteria like Streptococcus and Prevotella are often found in the lungs without causing disease, playing roles in maintaining the microbial balance and potentially stimulating the immune system, and Organisms like Pseudomonas aeruginosa and certain Staphylococcus species, while sometimes part of the normal microbiota, can become pathogenic under certain conditions, such as in individuals with compromised immune systems or underlying lung diseases (Dickson and Huffnagle, 2015).

The microbial diversity within the lungs can be modulated by various biotic and abiotic factors (Liu et al., 2022). An imbalance in the respiratory microbiome can facilitate colonization by opportunistic pathogens, culminating in respiratory infections, including pneumonia (Wypych et al., 2019; Hernández-Terán et al., 2021). Alterations in the microbiome are observed during infections of the lower respiratory tract and are closely correlated with the course and prognosis of pneumonia (Gu et al., 2019). Hence, a deeper understanding of alterations in microbiome composition is important for elucidating the role of pathogens in pulmonary infections. Studies have uncovered the microbial composition of the lungs in patients with bacterial meningitis (Moon et al., 2019), refractory *Mycoplasma pneumoniae* pneumonia (Shi et al., 2022; Deng et al., 2023), pulmonary tuberculosis (Chao et al., 2021; Zhang et al., 2023), and invasive pulmonary aspergillosis (Hérivaux et al., 2022) using untargeted pathogen metagenomics or 16S rRNA gene sequencing. While 16S rRNA sequencing can identify bacterial species, it lacks the resolution offered by metagenomic technologies, such as shotgun sequencing, particularly in closely related species (Gupta et al., 2019).

Distinct pulmonary diseases exhibited varying microbiological features; for instance, we found microbial compositions of the lower respiratory tract of patients having lung cancer, bronchiectasis, and NTM pneumonia primarily consisted of bacteria. This is consistent

with the understanding that bacterial communities can influence or be influenced by the pathophysiology of chronic respiratory diseases and cancer. For interstitial lung disease, common and severe pneumonia, these conditions had a more diverse microbial composition, including significant presences of fungi and viruses. This suggests a complex interplay in these diseases, potentially implicating more varied and severe pathogenic processes. Different types of disease showed enrichment of microorganisms, exhibiting a correlation with the clinical manifestations of each disease. For instance, the enrichment of certain bacteria in lung cancer could relate to the inflammatory and tumor-promoting environment, whereas the presence of fungi and viruses in interstitial lung disease could be linked to the disease's multifactorial etiology involving both immune response and environmental exposures.

Figures 3D, E illustrate the distribution of various bacteria across different pulmonary diseases, highlighting their potential roles in these conditions. Notably, Streptococcus pneumoniae appears frequently in a variety of lung diseases, including pneumonia, supporting its status as a primary pathogen (Shoar and Musher, 2020). The data in these charts align with research on the role of the microbiome in respiratory health, suggesting a potential causative relationship between specific microbial communities and diseases (Dietl et al., 2021). Additionally, Figures 3D, E highlight the significant presence of *Pseudomonas aeruginosa* in various pulmonary diseases, correlating with its known impact on chronic conditions like cystic fibrosis and non-CF bronchiectasis. The data from these charts show its prevalence and distribution across disease groups, underlining the challenges in managing this pathogen due to its resistance to multiple antibiotics. These findings emphasize the need for targeted antibiotic therapies to improve patient outcomes by effectively managing *Pseudomonas aeruginosa* infections, supporting ongoing research into tailored treatment strategies (Reynolds and Kollef, 2021; Eklof et al., 2024).

The diversity of microbiome in the lower respiratory tract varied among disease groups, with the highest diversity observed in the lung cancer and interstitial lung disease groups, succeeded by the common pneumonia and NTM pneumonia groups, and the lowest diversity in the bronchiectasis and severe pneumonia groups. High microbial diversity in lung cancer and interstitial lung disease groups could be linked to the chronic and progressive nature of these diseases, where a diverse microbial environment might influence disease progression and response to treatment. Common pneumonia and NTM pneumonia exhibited moderate diversity, this intermediate level might reflect the specific pathogenic involvement and the body's immune response to these conditions. Low diversity in severe pneumonia might indicate a dominance by pathogenic microbes that outcompete other members of the microbiota during acute disease phases. In bronchiectasis, reduced diversity could be associated with chronic infection or inflammation driven by a limited range of pathogens.

The AUC values obtained from our ROC analysis provide valuable insights into the performance of our classification models. The Combine group, with an AUC of 0.8696, demonstrates robust classification ability, indicating a strong relationship between disease types and pathogen abundance

within this group. This suggests that the combined dataset effectively captures relevant features, leading to better discrimination between classes. In contrast, the PN and SP groups exhibit lower AUC values of 0.7955 and 0.7470, respectively, revealing comparatively weaker model performance. These differences may highlight variations in how well the pathogen abundance correlates with disease types across different datasets or groups. The observed AUC values support the hypothesis that microbial ecology could be closely related to disease states, with the Combine group showing a more pronounced relationship. This aligns with the notion that a comprehensive dataset might enhance our ability to uncover associations between microbial communities and diseases. However, the lower AUCs in PN and SP groups suggest that additional factors or more nuanced features may be needed to improve classification performance in these contexts. Further investigation into these groups could reveal underlying complexities in the relationship between pathogen abundance and disease.

This study illustrates the use of mNGS in identifying pathogens and analyzing the lower respiratory microbiome. Our findings indicate that distinct pulmonary diseases show unique microbial features, which may be linked to the pathophysiology of different diseases.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

## Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

## Author contributions

RH: Data curation, Formal analysis, Validation, Writing – review & editing. SL: Investigation, Writing – review & editing. SZ:

Investigation, Writing – review & editing. YY: Data curation, Formal analysis, Visualization, Writing – review & editing. LL: Formal analysis, Visualization, Writing – review & editing. HY: Formal analysis, Visualization, Writing – review & editing. ZD: Data curation, Formal analysis, Visualization, Writing – review & editing. XC: Data curation, Formal analysis, Writing – review & editing. WW: Conceptualization, Writing – original draft. RR: Data curation, Formal analysis, Writing – original draft. XY: Conceptualization, Writing – original draft. BX: Conceptualization, Writing – original draft.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the Science Foundation for State Key Laboratory of Infectious Disease Prevention and Control (Grant No. 2022SKLID308).

## Conflict of interest

Authors YY, ZD, XC, WW and RR are employed by the company Matrixx Biotechnology Co., Ltd.

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2024.1320831/full#supplementary-material>

## References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1006/jmbi.1990.9999
- Blanch, L., Murias, G., and Nahum, A. (2002). Lung recruitment in unilateral lung disease. *Minerva anestesiologica* 68, 351–355.
- Chao, Y., Li, J., Gong, Z., Li, C., Ye, M., Hong, Q., et al. (2021). Rapid discrimination between tuberculosis and sarcoidosis using next-generation sequencing. *Int. J. Infect. Diseases: IJID: Off. Publ. Int. Soc. Infect. Dis.* 108, 129–136. doi: 10.1016/j.ijid.2021.05.028
- Chung, K. F. (2017). Airway microbial dysbiosis in asthmatic patients: A target for prevention and treatment? *J. Allergy Clin. Immunol.* 139, 1071–1081. doi: 10.1016/j.jaci.2017.02.004
- Cox, M. J., Turek, E. M., Hennessy, C., Mirza, G. K., James, P. L., Coleman, M., et al. (2017). Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients. *PLoS One* 12, e0170622. doi: 10.1371/journal.pone.0170622
- Deng, Q., Wang, Z., Wu, P., Liang, H., Wu, H., Zhang, L., et al. (2023). 16S rRNA gene sequencing reveals an altered composition of gut microbiota in children with *Mycoplasma pneumoniae pneumoniae* pneumonia treated with azithromycin. *J. Gen. Appl. Microbiol.* 68, 253–261. doi: 10.2323/jgam.2022.05.004
- Dickson, R. P., Erb-Downward, J. R., Falkowski, N. R., Hunter, E. M., Ashley, S. L., and Huffnagle, G. B. (2018). The lung microbiota of healthy mice are highly variable,

- cluster by environment, and reflect variation in baseline lung innate immunity. *Am. J. Respir. Crit. Care Med.* 198, 497–508. doi: 10.1164/rccm.201711-2180oc
- Dickson, R. P., Erb-Downward, J. R., and Huffnagle, G. B. (2013). The role of the bacterial microbiome in lung disease. *Expert Rev. Respir. Med.* 7, 245–257. doi: 10.1586/ers.13.24
- Dickson, R. P., and Huffnagle, G. B. (2015). The lung microbiome: new principles for respiratory bacteriology in health and disease. *PLoS Pathog.* 11, e1004923. doi: 10.1371/journal.ppat.1004923
- Dietl, B., Henares, D., Boix-Palop, L., Munoz-Almagro, C., Garau, J., and Calbo, E. (2021). Related factors to streptococcus pneumoniae invasive infection and clinical manifestations: the potential role of nasopharyngeal microbiome. *Front. Med. (Lausanne)* 8, 650271. doi: 10.3389/fmed.2021.650271
- Ding, L., Liu, Y., Wu, X., Wu, M., Luo, X., Ouyang, H., et al. (2021). Pathogen metagenomics reveals distinct lung microbiota signatures between bacteriologically confirmed and negative tuberculosis patients. *Front. Cell Infect. Microbiol.* 11, 708827. doi: 10.3389/fcimb.2021.708827
- Eklöf, J., Alispahic, I. A., Armbruster, K., Lapperre, T. S., Browatzki, A., Overgaard, R. H., et al. (2024). Systemic antibiotics for *Pseudomonas aeruginosa* infection in outpatients with non-hospitalised exacerbations of pre-existing lung diseases: a randomised clinical trial. *Respir. Res.* 25, 236. doi: 10.1186/s12931-024-02860-9
- Fang, C., Xu, L., Lu, J., Tan, H., Lin, J., and Zhao, Z. (2022). Clinical characteristics of chlamydia psittaci pneumonia confirmed by metagenomic next-generation sequencing. *Clin. Lab.* 68. doi: 10.7754/clin.lab.2022.220139
- Goeteyn, E., Grassi, L., Van den Bossche, S., Rigauts, C., Vande Weygaerde, Y., Van Braeckel, E., et al. (2023). Commensal bacteria of the lung microbiota synergistically inhibit inflammation in a three-dimensional epithelial cell model. *Front. Immunol.* 14, 1176044. doi: 10.3389/fimmu.2023.1176044
- Gu, L., Deng, H., Ren, Z., Zhao, Y., Yu, S., Guo, Y., et al. (2019). Dynamic changes in the microbiome and mucosal immune microenvironment of the lower respiratory tract by influenza virus infection. *Front. Microbiol.* 10, 2491. doi: 10.3389/fmicb.2019.02491
- Gupta, S., Mortensen, M. S., Schjørring, S., Trivedi, U., Vestergaard, G., Stokholm, J., et al. (2019). Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Commun. Biol.* 2, 291. doi: 10.1038/s42003-019-0540-1
- Hérivaux, A., Willis, J. R., Mercier, T., Lagrou, K., Gonçalves, S. M., Gonçalves, R. A., et al. (2022). Lung microbiota predict invasive pulmonary aspergillosis and its outcome in immunocompromised patients. *Thorax* 77, 283–291. doi: 10.1136/thoraxjnl-2020-216179
- Hernández-Terán, A., Mejía-Nepomuceno, F., Herrera, M. T., Barreto, O., García, E., Castillejos, M., et al. (2021). Dysbiosis and structural disruption of the respiratory microbiota in COVID-19 patients with severe and fatal outcomes. *Sci. Rep.* 11, 21297. doi: 10.1038/s41598-021-00851-0
- Hilty, M., Burke, C., Pedro, H., Cardenas, P., Bush, A., Bossley, C., et al. (2010). Disordered microbial communities in asthmatic airways. *PLoS One* 5, e8578. doi: 10.1371/journal.pone.00808578
- Hong, B. Y., Paulson, J. N., Stine, O. C., Weinstock, G. M., and Cervantes, J. L. (2018). Meta-analysis of the lung microbiota in pulmonary tuberculosis. *Tuberculosis (Edinb)* 109, 102–108. doi: 10.1016/j.tube.2018.02.006
- Hoshina, H., and Takei, H. (2021). Drug-induced interstitial lung disease after anthracycline-combined chemotherapy for breast cancer: A case report and literature review. *Case Rep. Oncol.* 14, 1671–1676. doi: 10.1159/000520126
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Liao, S. X., Sun, P. P., Gu, Y. H., Rao, X. M., Zhang, L. Y., and Ou-Yang, Y. (2019). Autophagy and pulmonary disease. *Ther. Adv. Respir. Dis.* 13, 1753466619890538. doi: 10.1177/1753466619890538
- Liu, C., Wu, K., Sun, T., Chen, B., Yi, Y., Ren, R., et al. (2022). Effect of invasive mechanical ventilation on the diversity of the pulmonary microbiota. *Crit. Care (London England)* 26, 252. doi: 10.1186/s13054-022-04126-6
- Llorens-Rico, V., Gregory, A. C., Van Weyenbergh, J., Jansen, S., Van Buyten, T., Qian, J., et al. (2021). Clinical practices underlie COVID-19 patient respiratory microbiome composition and its interactions with the host. *Nat. Commun.* 12, 6243. doi: 10.1038/s41467-021-26500-8
- Luan, Y., Hu, H., Liu, C., Chen, B., Liu, X., Xu, Y., et al. (2021). A proof-of-concept study of an automated solution for clinical metagenomic next-generation sequencing. *J. Appl. Microbiol.* 131, 1007–1016. doi: 10.1111/jam.15003
- Man, W. H., de Steenhuijsen Piters, W. A., and Bogaert, D. (2017). The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat. Rev. Microbiol.* 15, 259–270. doi: 10.1038/nrmicro.2017.14
- Mathew, A. E., and Mugele, J. (2021). Multifocal pneumonia: fever, cough, and dyspnea. *Am. Family Physician* 103, 503–504.
- Moon, J., Kim, N., Kim, T. J., Jun, J. S., Lee, H. S., Shin, H. R., et al. (2019). Rapid diagnosis of bacterial meningitis by nanopore 16S amplicon sequencing: A pilot study. *Int. J. Med. Microbiology: IJMM* 309, 151338. doi: 10.1016/j.ijmm.2019.151338
- Natalini, J. G., Singh, S., and Segal, L. N. (2023). The dynamic lung microbiome in health and disease. *Nat. Rev. Microbiol.* 21, 222–235. doi: 10.1038/s41579-022-00821-x
- Rahman, M. M., Bibi, S., Rahaman, M. S., Rahman, F., Islam, F., Khan, M. S., et al. (2022). Natural therapeutics and nutraceuticals for lung diseases: Traditional significance, phytochemistry, and pharmacology. *Biomedicine pharmacotherapy = Biomedecine pharmacotherapie* 150, 113041. doi: 10.1016/j.biopha.2022.113041
- Ramsheh, M. Y., Haldar, K., Esteve-Codina, A., Purser, L. F., Richardson, M., Muller-Quernheim, J., et al. (2021). Lung microbiome composition and bronchial epithelial gene expression in patients with COPD versus healthy individuals: a bacterial 16S rRNA gene sequencing and host transcriptomic analysis. *Lancet Microbe* 2, e300–ee10. doi: 10.1016/s2666-5247(21)00035-5
- Ren, L. L., Wang, Y. M., Wu, Z. Q., Xiang, Z. C., Guo, L., Xu, T., et al. (2020). Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chin. Med. J.* 133, 1015–1024. doi: 10.1097/CM9.0000000000000722
- Reynolds, D., and Kollef, M. (2021). The epidemiology and pathogenesis and treatment of *Pseudomonas aeruginosa* infections: an update. *Drugs* 81, 2117–2131. doi: 10.1007/s40265-021-01635-6
- Schlager, R., Chiu, C. Y., Miller, S., Procop, G. W., Weinstock, G., Professional Practice, C., et al. (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch. Pathol. Lab. Med.* 141, 776–786. doi: 10.5858/arpa.2016-0539-ra
- Shi, D. W., Wang, D. M., Ning, L. H., Li, J., Dong, Y., Zhang, Z. K., et al. (2022). Using 16S rDNA Sequencing Technology to Preliminarily Analyze Intestinal Flora in Children with *Mycoplasma pneumoniae* Pneumonia. *Biomed. Environ. sciences: BES* 35, 528–537. doi: 10.3967/bes2022.070
- Shoar, S., and Musher, D. M. (2020). Etiology of community-acquired pneumonia in adults: a systematic review. *Pneumonia (Nathan)* 12, 11. doi: 10.1186/s41479-020-00074-3
- Tian, X., Duan, W., Zhang, X., Wu, X., Zhang, C., Wang, Z., et al. (2022). Metagenomic next-generation sequencing reveals the profile of viral infections in kidney transplant recipients during the COVID-19 pandemic. *Front. Public Health* 10, 888064. doi: 10.3389/fpubh.2022.888064
- Tsay, J. J., Wu, B. G., Badri, M. H., Clemente, J. C., Shen, N., Meyn, P., et al. (2018). Airway microbiota is associated with upregulation of the PI3K pathway in lung cancer. *Am. J. Respir. Crit. Care Med.* 198, 1188–1198. doi: 10.1164/rccm.201710-2118oc
- Tsay, J. J., Wu, B. G., Sulaiman, I., Gershner, K., Schluger, R., Li, Y., et al. (2021). Lower airway dysbiosis affects lung cancer progression. *Cancer Discovery* 11, 293–307. doi: 10.1158/2159-8290.CD-20-0263
- Wilson, M. R., Sample, H. A., Zorn, K. C., Arevalo, S., Yu, G., Neuhaus, J., et al. (2019). Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N Engl. J. Med.* 380, 2327–2340. doi: 10.1056/NEJMoa1803396
- Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46. doi: 10.1186/gb-2014-15-3-r46
- Wypych, T. P., Wickramasinghe, L. C., and Marsland, B. J. (2019). The influence of the microbiome on respiratory health. *Nat. Immunol.* 20, 1279–1290. doi: 10.1038/s41590-019-0451-9
- Yang, L., Zhou, F., Zheng, D., Wang, D., Li, X., Zhao, C., et al. (2021). FGF/FGFR signaling: From lung development to respiratory diseases. *Cytokine Growth factor Rev.* 62, 94–104. doi: 10.1016/j.cytogfr.2021.09.002
- Yılmaz, N., Uçar, E. Y., and Sağlam, L. (2017). Mycobacterium tuberculosis and nontuberculous mycobacteria coinfection of the lungs. *Turkish Thorac. J.* 18, 23–26. doi: 10.5152/TurkThoracJ.2017.16034
- Zhang, Q., Bai, J. S., Liu, J. X., Fu, A. S., Wang, J. M., Zhou, X. Y., et al. (2023). T-SPOT.TB negative and diagnosed as interstitial pulmonary tuberculosis by NGS: a case report and literature review. *Clin. Lab.* 69. doi: 10.7754/clin.lab.2022.220704