

Comprehensive Analysis of Pathogen Diversity and Diagnostic Biomarkers in Patients with Suspected Pulmonary Tuberculosis Through Metagenomic Next-Generation Sequencing

Yuecui Li^{1,*}, Chenghang Li^{1,*}, Yu Fang^{2,*}, LiLi Zhang¹, Xiaoyan Ying¹, Ruotong Ren³, Yinghui Zang¹, Dandan Ying¹, Shengwei Zhu¹, Jiao Liu¹, Xuefang Cao³

¹Department of Infectious Disease, The First People's Hospital of Yongkang, affiliated to Hangzhou Medical College, Jinhua, People's Republic of China;

²Zhejiang Key Laboratory of Digital Technology in Medical Diagnostics, Hangzhou, People's Republic of China; ³MatriDx Biotechnology Co., Ltd, Hangzhou, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yuecui Li, Department of Infectious Disease, The First People's Hospital of Yongkang, affiliated to Hangzhou Medical College, 599 Jinshan West Road, Jinhua, 321300, People's Republic of China, Email yklycwh@126.com; Xuefang Cao, MatriDx Biotechnology Co., Ltd, Hangzhou, 311112, People's Republic of China, Email xfczju@163.com

Background: This study aimed to investigate the co-infecting pathogens and lung microbiomes in patients with clinically confirmed pulmonary tuberculosis (TB) and explore potential diagnostic biomarkers to differentiate between varied infection patterns.

Methods: We conducted a retrospective cohort study by analyzing 198 bronchoalveolar lavage fluid (BALF) samples collected from patients with suspected pulmonary TB. All BALF samples were sequenced using metagenomic next-generation sequencing (mNGS).

Results: A total of 63 pathogens were detected in all samples. The TB group exhibited a higher diversity of pathogens ($n=51$) than the Non-TB group ($n=37$). The analysis revealed that TB patients had significantly higher pathogen counts ($P=0.014$), and specific microorganisms, such as *Mycobacterium tuberculosis* complex (MTBC), MTB, *Streptococcus infantis*, and *Campylobacter curvus*, were significantly enriched. Furthermore, the abundance of MTBC was negatively correlated with hemoglobin levels ($R=-0.17$, $P=0.015$) and positive correlated with C-reactive protein (CRP) levels ($R=0.16$, $P=0.029$). The random forest model combined eight differential microbes and five clinical parameters, yielding an area under the curve (AUC) of 0.86 for differentiating TB from Non-TB cohorts, whereas subgroup differentiation yielded an AUC of 0.571, demonstrating the potential for targeted diagnostics in pulmonary infections.

Conclusion: Our findings highlight the complexity of co-infection patterns in pulmonary TB and emphasize the potential of integrating microbial and clinical markers to improve diagnostic accuracy. This study provides valuable insights into the role of the lung microbiome in TB and informs future research on targeted therapies for this disease.

Keywords: tuberculosis, bronchoalveolar lavage fluid, lung microbiome, biomarkers, metagenomic next-generation sequencing

Introduction

Tuberculosis (TB), caused by the *Mycobacterium tuberculosis* (MTB) complex, remains a significant global public health challenge and contributes to high morbidity and mortality rates worldwide.¹ Historically, TB has been one of the deadliest infectious diseases and continues to rank among the leading causes of global infectious mortality.² Although the incidence of TB has gradually declined over the past decade, with mortality rates decreasing by nearly one-third, this positive trend has been significantly disrupted by the COVID-19 pandemic.³ In many regions, the pandemic has led to a substantial reduction in TB testing and case notifications, resulting in an associated increase in mortality and the reversal of global TB control efforts by approximately ten years.⁴ It was estimated that 10.6 million people are estimated

to be infected with TB globally in 2021, corresponding to an incidence rate of 134 cases per 100,000 people.² The diagnosis and treatment of TB are complex and often controversial.⁵ Numerous studies have reported co-infections with various strains of *Mycobacterium* or other pathogenic species in patients with pulmonary TB.^{6,7} Given that prompt and accurate diagnosis is essential for effective treatment, there is an urgent need for rapid screening and diagnostic methods to enhance TB management.

Metagenomic next-generation sequencing (mNGS) has recently emerged as a promising complementary approach and has been widely adopted for the diagnosis of various infectious diseases in clinical settings. It offers several advantages, including shortened turnaround time, unbiased detection, and semiquantitative assessment, allowing the theoretical identification of all pathogens present in a clinical sample. This technique is particularly valuable for detecting rare, novel, and atypical etiologies associated with complicated infectious diseases, thus facilitating more precise targeted antimicrobial therapy.^{8,9} Extensive research has demonstrated that mNGS can serve as an important complement to traditional etiological diagnostic methods for patients with TB owing to its superior performance.^{10,11} However, few studies have systematically characterized the spectrum of co-infecting pathogens in patients with suspected pulmonary TB.

The microbiota plays an important role in human health and the development of various diseases.¹² High-throughput sequencing technologies have been applied to characterize human microbiota across different body habitats, including the gut,¹³ oral cavity¹⁴ and respiratory tract.¹⁵ The lungs and entire lower respiratory tract (LRT) host niche-specific microbial communities.^{12,16} Most lung microbiome studies have relied on 16S rRNA sequencing, which has several limitations, including potential amplification bias and inability to detect many microorganisms at the species and strain levels.¹⁷ Understanding lung microbiomes is essential for the early diagnosis of pulmonary TB, facilitating timely treatment, and improving prognosis. Some studies have used mNGS to investigate the respiratory microbiome in COVID-19 patients,¹⁸ community-acquired pneumonia (CAP) patients¹⁹ and people living with HIV.²⁰ Most TB research has focused on microbial communities in gut or sputum samples using 16S rRNA sequencing, which provides limited information on the bacterial composition. Few studies utilizing mNGS have reported disruption of bronchoalveolar lavage fluid (BALF) in TB patients.^{5,21} However, it remains unclear whether these disruptions correlate with different infection patterns. Given that the respiratory microbiome encompasses both commensals and pathogenic organisms, integrating pathogen identification with comprehensive microbiome characterization using mNGS could enhance our understanding of pulmonary TB.

This retrospective cohort study extensively investigate the co-infected pathogens and lung microbiomes of patients clinically confirmed to have pulmonary TB, as well as patient cohorts with different infection patterns. Furthermore, we explored potential diagnostic biomarkers that could be used to differentiate between the patient cohorts. This study is the first to provide comprehensive information on the lung microbiomes of patients with pulmonary TB based on sensitive BALF mNGS compared to those without pulmonary TB, thereby offering valuable insights into the etiology of pulmonary infection and informing targeted therapy for patients with pulmonary TB.

Methods

Study Design and Patients

This study retrospectively enrolled patients with suspected pulmonary TB between January 2022 and June 2023 at The First People's Hospital of Yongkang. Suspected pulmonary TB was diagnosed based on clinical manifestations and imaging examinations. The results of etiological screening of the enrolled patients were evaluated by a panel of clinical experts (including three experienced physicians). The sputum tests for these patients included Ziehl–Neelsen staining (AFS), Xpert, and bacterial culture. Given that conventional clinical examinations were unable to provide a definitive diagnosis, all enrolled patients underwent electronic bronchoscopy at our hospital. Chest CT scan revealed lesions with a brush-like appearance, prompting the collection of BALF for mNGS detection. Infectious diseases are diagnosed on the basis of microbiological tests, mNGS results, and clinical review results. Patients were classified into TB and non-TB groups based on comprehensive diagnostic criteria. The non-TB groups refer to patients who have not been diagnosed with TB. Notably, patients in whom nontuberculous mycobacteria (NTM) were detected, without the presence of MTBC, were also classified as non-TB. Combined with the mNGS results, TB groups were sub-divided

into only MTB infection (TB) and co-infection with MTB and NTM (TB+NTM) subgroups. Non-TB groups were subdivided into NTM infection without MTB infection (NTM) and non-MTB-and-non-NTM infection (Non-TB-NTM) subgroups. Patients with incomplete clinical data or without any microbial results, such as smear, culture, PCR, or Xpert, were excluded.

Clinical Data

Demographic information, clinical symptoms, laboratory test results, imaging examination results, diagnosis and treatment history, and patient outcomes were collected from the electronic medical records. Laboratory data included white blood cells (WBC), hemoglobin (HGB), C-reactive protein (CRP), and procalcitonin (PCT).

Clinical Sample Collection and DNA Extraction

BALF samples were collected from each patient by experienced bronchoscopists through bronchoscopy under midazolam anesthesia, following standard operational procedures,²² with the consent of the patient or their parents/guardians. DNA extraction and library preparation from clinical samples were performed using an NGS Automatic Library Preparation System (MatriDx Biotech Corp. Hangzhou). DNA quality was assessed using BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, United States) combined with quantitative PCR to measure adapters before sequencing.

Metagenomic Next-Generation Sequencing

Qualified DNA libraries were pooled and sequenced on an Illumina NextSeq500 system (50 bp single-end; San Diego, CA, United States). To control the quality of each sequencing run, a negative control and positive control were conducted in parallel (Supplemental Methods). Pathogen identification followed the data-filtering criteria established by MatriDx Biotechnology Co. Ltd.

Statistical Analysis

All collected data were statistically analyzed using the R package. Categorical variables, shown as frequencies and percentages, were compared using the Fisher's exact test. Continuous measurement data following a normal distribution are shown as the mean (standard deviation) or mean (standard error), and non-normal distribution is shown as the median (range). Differences and significance between groups were calculated using Student's *t*-test (for normally distributed data) and Wilcoxon rank-sum test or Kruskal–Wallis test (for non-normally distributed data). The data were visualized using R software (version 4.2.1) (Supplemental Methods).

Results

Demographics and Clinical Characteristics of Patients

A total of 264 BALF samples were collected from patients with suspected pulmonary TB for this study (Figure 1A). Of these, 48 had incomplete clinical data, and 18 lacked microbial results (cultures, PCR, or Xpert), leaving 198 patients in the study, including 87 clinically confirmed TB cases (TB, *n*=87) and 111 Non-TB cases (Non-TB, *n*=111). Among the 87 TB patients, the positive rates for smear, culture, Xpert, and X-rays were as follows: 8 (9.20%), 30 (34.48%), 31 (35.63%), and 65 (74.71%), respectively. The four subgroups, TB, TB+NTM, NTM, and Non-TB-NTM, have 76, 11, 21, and 90 cases, respectively. Among the 198 patients, 93 (47.0%) were females and 105 (53.0%) were males, with a median age of 44.50 (IQR: 27.00–60.75) years. There were statistically significant differences in patient age between the two groups (*P*=0.011) (Table 1) and the four subgroups (*P*=0.028) (Table 2). Although no significant differences were observed in WBC, HGB, and PCT between the two groups, CRP levels (*P*=0.017) were significantly higher in the TB group than in the Non-TB group. However, the laboratory findings of the four subgroups showed no significant differences.

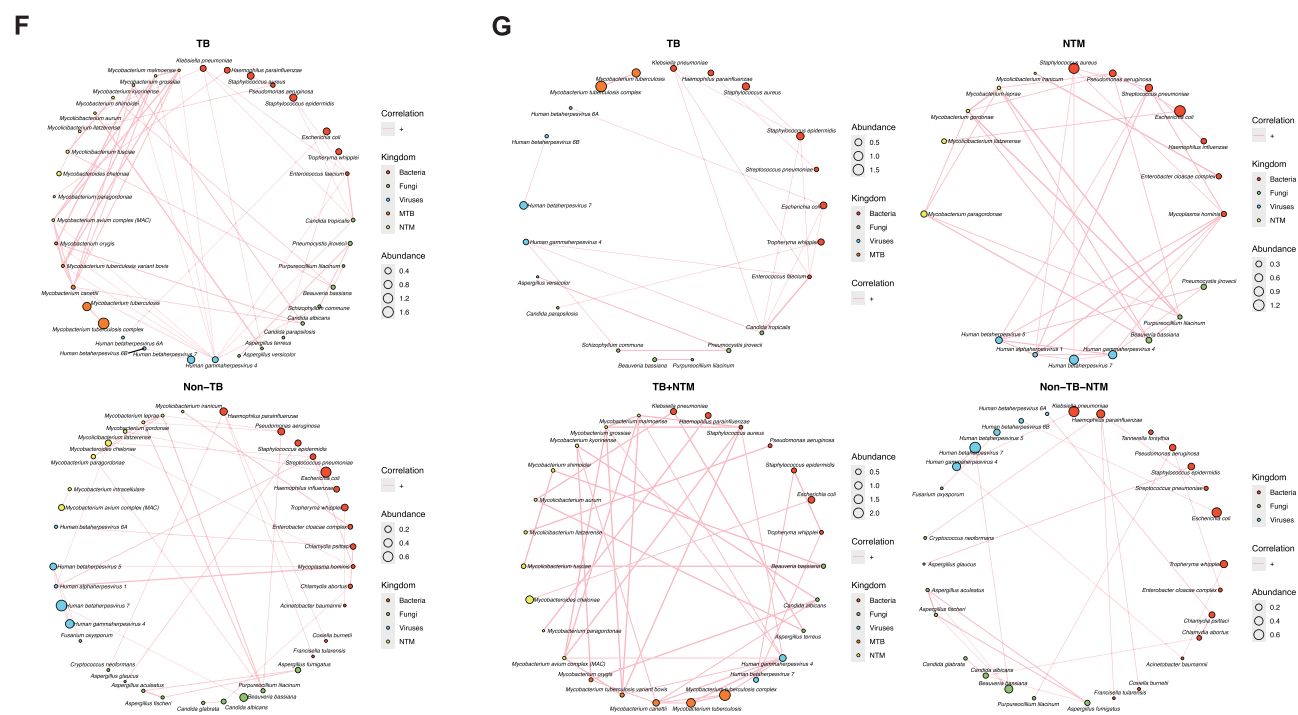


Figure 1 Profile and characteristics of pathogens identified in patients from different cohorts based on mNGS. **(A)** Flowchart of patients enrolled in this study. **(B)** Overview and distribution of pathogenic in different cohorts. The heatmap shows the abundance of the different pathogens (bacteria, fungi, viruses, MTB and NTM) with log10-transformed RPM of pathogens were applied. Group names, subgroup names and microbial taxonomy are indicated by color bars at the bottom. The barplot on the top shows the total counts of pathogens in each sample from the different groups, and the barplot on the right shows the total frequency of each pathogen found in all samples. **(C)** Comparison of the counts and burdens of pathogens in patients with or without MTB infections. Differences between groups were assessed using Wilcoxon-test. **(D)** Comparison of the counts and burdens of pathogens in patients from different subgroups. Differences between groups were assessed using Wilcoxon-test. **(E)** Venn diagram and UpSet plot show the shared and unique pathogenic in the two groups and the four subgroups, respectively. **(F)** Co-occurrence network of pathogenic species among patients with or without MTB infections. **(G)** Co-occurrence network of pathogenic species among patients from different subgroups. Pathogenic species with $r > 0.3$ and p value < 0.05 were selected. Node size indicates the abundance (log10-transformed RPM) of species and node color indicates the different taxonomy. The color of connections was used to distinguish the different microbial interactions, where red represents significantly positive interactions between species and blue represents significantly negative interactions. The thickness of connections is proportional to the strengthen of correlations ($\text{cor}^2 > 0.75$).

species or at least two species ([Supplementary Figure 2B](#)). Among the 87 cases with confirmed pathogens in the Non-TB group, 41 (37%) were found to be mono-infections and 46 (41%) cases were categorized as mixed infections ([Supplementary Figure 2C](#)).

Pathogen counts (the number of different pathogenic species detected) and pathogen burden (the quantity of pathogens, measured using Reads Per Million, RPM) in each sample from different groups and subgroups were analyzed. Interestingly, the pathogen counts in the TB group were significantly higher than those in the Non-TB group ($P=0.014$), despite no significant difference in pathogen burden ([Figure 1C](#)). In terms of pathogen counts among the

Table 1 Demographic and Clinical Characteristics of Patients in Different Groups

Characteristics	Overall(N=198)	TB(N=87)	Non-TB(N=111)	P-value
Demographics				
Age ^a	44.50 [27.00, 60.75]	35.00 [23.00, 59.50]	49.00 [35.00, 61.00]	0.011
Gender ^b				
Female	93 (47.0)	41 (47.1)	52 (46.8)	>0.999
Male	105 (53.0)	46 (52.9)	59 (53.2)	

(Continued)

Table 1 (Continued).

Characteristics	Overall(N=198)	TB(N=87)	Non-TB(N=111)	P-value
Laboratory examination				
WBC (*10 ⁹ /L) ^a	5.72 [4.34, 7.10]	5.59 [4.46, 6.54]	5.91 [4.30, 7.38]	0.475
HGB (g/L) ^c	131.38 (19.25)	128.74 (18.25)	133.45 (19.84)	0.087
CRP (ng/mL) ^a	0.50 [0.50, 5.24]	1.10 [0.50, 15.62]	0.50 [0.50, 2.62]	0.017
PCT (ng/mL) ^a	0.03 [0.01, 0.05]	0.02 [0.01, 0.05]	0.03 [0.01, 0.05]	0.705

Notes: ^aData were presented as median [IQR]. ^bData were presented as frequency (%). ^cData were presented as means (SD).

Abbreviations: WBC, White blood cells; HGB, Hemoglobin; CRP, C-reactive protein; PCT, Procalcitonin.

Table 2 Demographic and Clinical Characteristics of Patients in Different Subgroups

Characteristics	Overall(N=198)	TB(N=76)	NTM(N=21)	TB+NTM(N=11)	Non-TB-NTM(N=90)	P-value
Demographics						
Age ^a	44.50 [27.00, 60.75]	31.00 [22.75, 58.25]	50.00 [35.00, 59.00]	55.00 [36.00, 63.00]	49.00 [35.25, 61.00]	0.028
Gender ^b						
Female	93 (47.0)	38 (50.0)	11 (52.4)	3 (27.3)	41 (45.6)	0.510
Male	105 (53.0)	38 (50.0)	10 (47.6)	8 (72.7)	49 (54.4)	
Laboratory examination						
WBC (*10 ⁹ /L) ^a	5.72 [4.34, 7.10]	5.62 [4.56, 6.62]	5.92 [3.68, 6.99]	4.54 [4.09, 6.19]	5.91 [4.31, 7.82]	0.638
HGB (g/L) ^c	131.38 (19.25)	127.59 (18.44)	128.76 (21.18)	136.64 (15.38)	134.54 (19.47)	0.085
CRP (ng/mL) ^a	0.50 [0.50, 5.24]	1.45 [0.50, 14.84]	0.50 [0.50, 3.34]	0.92 [0.50, 25.95]	0.50 [0.50, 2.34]	0.113
PCT (ng/mL) ^a	0.03 [0.01, 0.05]	0.02 [0.01, 0.05]	0.03 [0.01, 0.04]	0.04 [0.02, 0.07]	0.02 [0.01, 0.06]	0.410

Notes: ^aData were presented as median [IQR]. ^bData were presented as frequency (%). ^cData were presented as means (SD).

Abbreviations: WBC, White blood cells; HGB, Hemoglobin; CRP, C-reactive protein; PCT, Procalcitonin.

four subgroups, the TB+NTM subgroup exhibited significantly higher counts than both the TB subgroup ($P<0.01$) and Non-TB-NTM ($P<0.01$) subgroups (Figure 1D). Compared with the Non-TB-NTM subgroup, pathogen counts in the NTM subgroup were significantly higher ($P<0.01$). Conversely, the pathogen burden was significantly higher in the TB+NTM subgroup than in the Non-TB-NTM subgroup ($P<0.01$) (Figure 1D). As expected, the MTBC and MTB were predominant in patients with TB (Supplementary Figure 3A). In contrast, *Human betaherpesvirus 7* and *Escherichia coli* were the most abundant species in the Non-TB group. Consistently, MTBC and MTB were predominant in the TB and TB+NTM subgroups. The most dominant species in the NTM subgroup included *Escherichia coli*, *Staphylococcus aureus* and *Mycobacteroides chelonae*, whereas *Human betaherpesvirus 7*, *Klebsiella pneumoniae*, and *Escherichia coli* were predominant in the Non-TB-NTM subgroup (Supplementary Figure 3B).

Spearman correlation analysis was performed to compare the co-occurrence network of pathogens between the groups and subgroups. The results showed significant positive correlations among the pathogens in the different patient cohorts. After excluding microbial species with correlations less than 0.3 or P values greater than 0.05, 36 interaction nodes and 98 connections were retained in the TB group, while 37 interaction nodes and 76 connections were retained in the Non-TB group. This indicates a more complex correlation among the pathogens in the TB group. Moreover, a strong correlation was observed between the MTB and NTM species in the TB group (Figure 1F, Supplementary Table 2). Among different subgroups, 21 interaction nodes and 40 connections were retained in the different

subgroups, the TB subgroup retained 21 interaction nodes and 40 connections, the NTM subgroup had 19 nodes and 70 connections, the TB+NTM subgroup included 27 nodes and 78 connections, and the Non-TB-NTM subgroup contained 29 nodes and 50 connections. These results indicate that the TB+NTM and Non-TB-NTM subgroups were more diverse and complex, with significant positive correlations, particularly in the TB+NTM subgroup. (Figure 1G, [Supplementary Table 3](#)).

Microbial Composition and Diversity of the Lower Respiratory Tract in Patients with Suspected Pulmonary Infection Identified Through mNGS

The composition and diversity of microorganisms in BALF samples from different patient cohorts were explored. After filtering out microbial species with low frequency (overall less than 10%), 261 microorganisms at the species level (235 bacteria, 20 fungi, and 6 viruses) were included for further analyses. The heatmap in [Figure 2A](#) displays the top 50 bacterial species, along with all fungal and viral species detected in the BALF samples from different groups and subgroups. *Prevotella melaninogenica*, *Streptococcus mitis*, and *Prevotella jejuni* were the most frequent bacterial species in all samples, whereas *Malassezia restricta* and *Human betaherpesvirus 7* were the most prevalent fungal and viral species, respectively ([Figure 2A](#), [Supplementary Table 4](#)). A high proportion of microbial species (89.7%) were shared between TB and Non-TB groups. However, 20 species were uniquely observed in the TB group and seven species were uniquely found in the Non-TB group ([Figure 2B](#)). Similarly, a substantial number of shared species (n=202) was identified across all subgroups, with fewer unique species observed in the various subgroups ([Figure 2C](#)). However, both alpha and beta diversity indices were not significantly different between the groups ([Supplementary Figure 4A](#) and [B](#)) and subgroups ([Supplementary Figure 4C](#) and [D](#)).

Microbial Species with Differential Abundance and Their Correlation with Clinical Parameters of Patients from Different Cohorts

To identify microbes that were significantly enriched in different groups and subgroups, linear discriminant analysis effect size (LEfSe) analysis was conducted. Several microbes, including MTBC, MTB, *Streptococcus infantis* and *Campylobacter curvus*, were significantly abundant in the TB group, whereas *Corynebacterium striatum*, *Staphylococcus epidermidis*, and *Ralstonia mannitolilytica* were significantly enriched in the Non-TB group ([Figure 3A](#)). When comparing the different subgroups, only four microorganisms were identified as significantly enriched: one in the TB+NTM subgroup and three in the NTM subgroup ([Figure 3B](#)). Specifically, *Mycobacterium canettii* was enriched in the TB+NTM subgroup, whereas *Bacteroides zoogloformans*, *Campylobacter curvus* and *Enterobacter cloacae* complex were found to be significantly abundant in the NTM subgroup. To further investigate the correlation between key microbial species in BALF and clinical characteristics, a correlation analysis was performed using Spearman's rank-based test. Interestingly, the abundance of MTBC was significantly negatively correlated with the level of HGB ($R=-0.17$, $P=0.015$), whereas it exhibited a significant positive correlation with the level of CRP ($R=0.16$, $P=0.029$) ([Figure 3C](#), [Supplementary Figure 5](#) and [Supplementary Table 5](#)). However, no significant differences were observed between the four key microbes screened from different subgroups and their clinical indices ([Figure 3D](#), [Supplementary Table 6](#)).

Random Forest-Based Classification Model for Screening Potential Biomarkers

A random forest model was constructed based on differential microorganisms and potential clinical indices. Ten-fold cross-validation was employed, with a training-to-testing set ratio of 3:1. A combination of eight differential microorganisms derived from the LEfSe results and all five clinical parameters was selected for the random forest classification model to differentiate between the TB and Non-TB cohorts. The importance of the 13 features mentioned above was ranked based on the mean decrease in accuracy, as shown in [Figure 3E](#). The top four important features were MTBC, *Campylobacter curvus*, MTB and age. We further assessed the diagnostic performance of the classifiers based on eight species combined with five clinical indices using the ROC curves ([Figure 3F](#)). The results indicated that the current classifier demonstrated satisfactory diagnostic performance, with an AUC of 0.86, a specificity of 81.2%, and

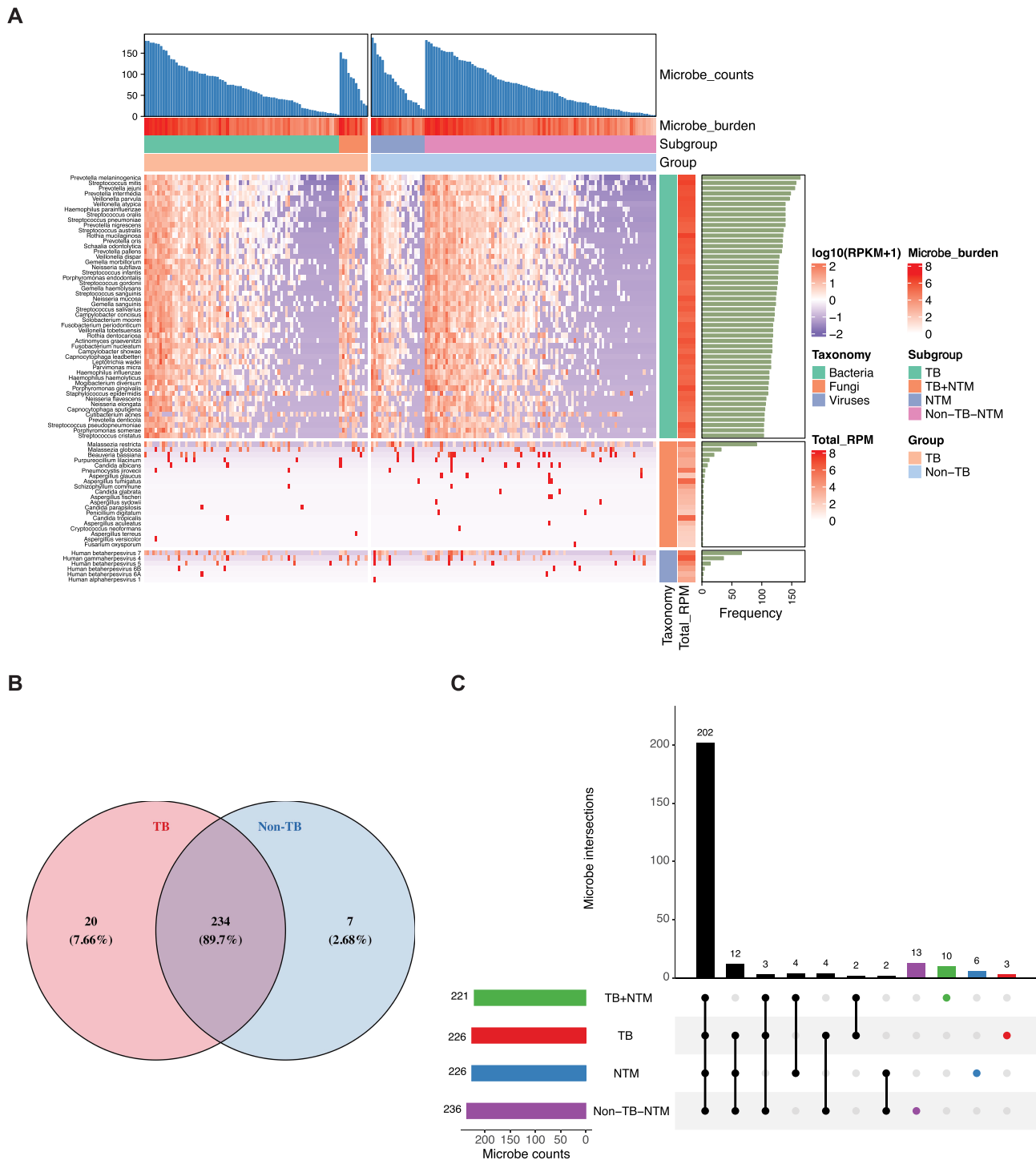


Figure 2 Overview and distribution of microbial species in BALF samples from patients in different cohorts. **(A)** The heatmap shows the abundance of the top 50 bacteria, all fungi and all viruses species, with log10-transformed RPM of microbes were applied. Group names, subgroup names and microbial taxonomy are indicated by color bars on the right. The barplot on the top shows the total counts of microbes in each sample from the different groups, and the barplot on the right shows the total frequency of each microbe found in all samples. **(B)** Venn diagram shows the shared and unique pathogenic in the two groups. **(C)** Upset plot shows the shared and unique pathogenic in the four subgroups.

a sensitivity of 78.6%. A combination of four differential microbes based on the LEfSe results and all five clinical parameters was selected for the random forest classification model to differentiate the different patient subgroups. As shown in **Figure 3G**, *Campylobacter curvus*, age, *Mycobacterium canettii* and CRP level were ranked as the four most important variables based on the mean decrease in accuracy. Similarly, the diagnostic performance of the classifiers based

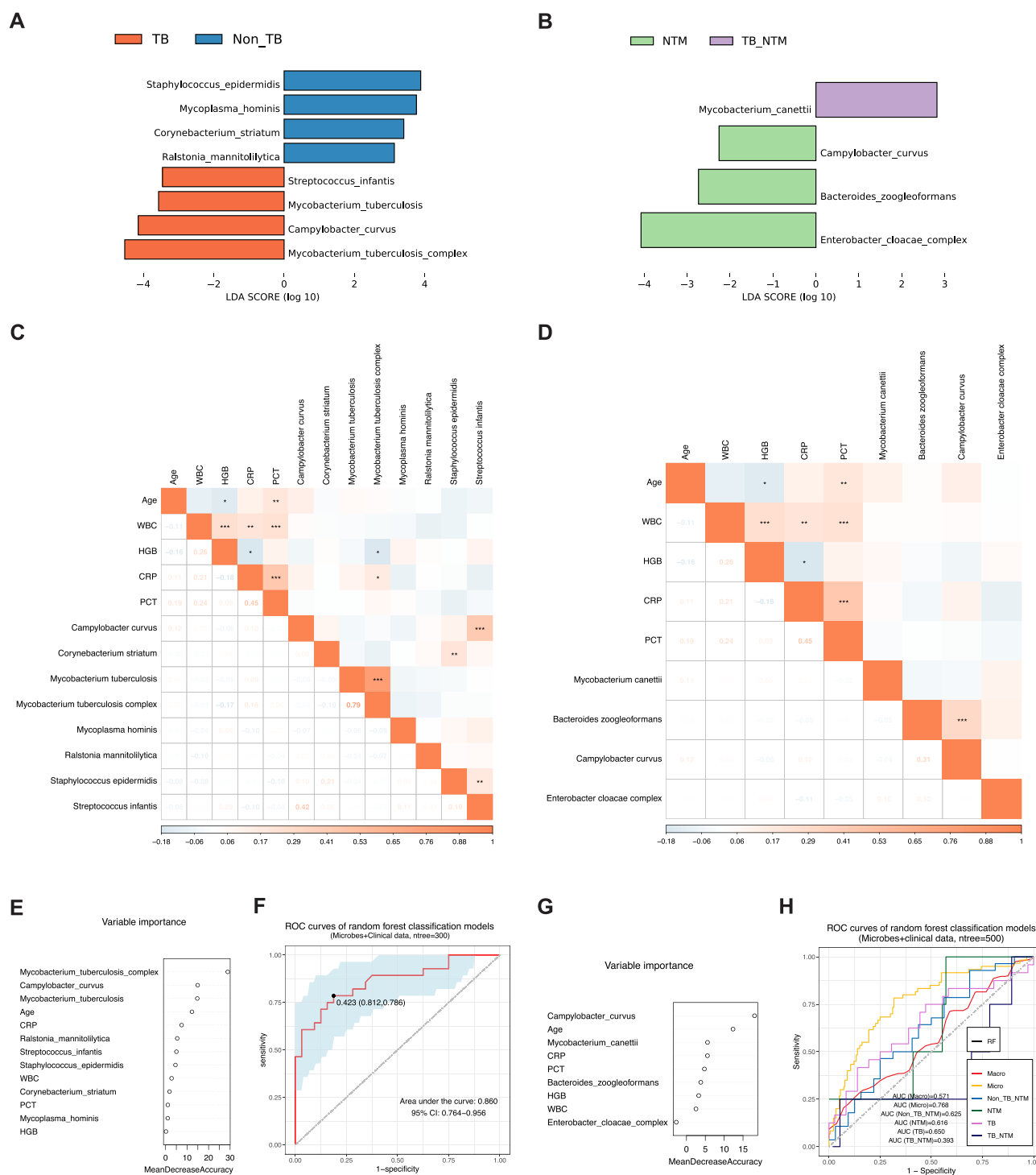


Figure 3 Predictive model for *Mycobacterium tuberculosis* infection based on microbial species and clinical parameters. **(A)** Microbial species with differential abundance between TB and Non-TB groups identified through LEfSe analysis with the thresholds of $\log_{10}$ LDA score ≥ 2 and P value < 0.05 . **(B)** Microbial species with differential abundance between different subgroups identified through LEfSe analysis with the thresholds of $\log_{10}$ LDA score ≥ 2 and P value < 0.05 . **(C)** Spearman correlation analysis of differential microbes between groups and clinical parameters of patients. **(D)** Spearman correlation analysis of differential microbes between subgroups and clinical parameters of patients. Orange values indicate species were positively correlated with clinical data, while blue ones indicate the species were negatively correlated with clinical data. Significant correlations are denoted with asterisks, where * represents $P < 0.05$, ** represents $P < 0.01$ and *** represents $P < 0.001$. **(E)** Random forest-based classification model and contribution of differential microbial species and clinical parameters between groups were ranked by mean decrease accuracy. **(F)** Random forest-based classification model and contribution of differential microbial species and clinical parameters between subgroups were ranked by mean decrease accuracy. **(G)** Receiver operating characteristic curve (ROC) of random forest model for classifying different groups. **(H)** Receiver operating characteristic curve (ROC) of random forest model for classifying different subgroups.

on the nine species combined with the five clinical indices was evaluated using the ROC curves (Figure 3H). The results showed that the current classification model was effective in distinguishing the different patient subgroups, with an average AUC of 0.571.

Discussion

To our knowledge, this is the first study to comprehensively investigate the spectrum of co-infectious pathogens in patients with suspected TB. Using mNGS technology, we identified 5 kinds of MTB in patients diagnosed with pulmonary TB. Previous report confirmed that the 16S rRNA gene sequencing approach lacks sensitivity for detecting NTM in culture-positive respiratory samples.²³ Furthermore, mNGS effectively identifies various NTM species, a task that is often challenging using culture-based methods. Our study revealed that TB patients were frequently co-infected with NTM and other pathogenic bacteria, fungi, and virus species. Previous studies utilizing mNGS have highlighted its unique advantages in identifying pulmonary TB along with other co-infections, including *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Human herpesvirus*, *Cryptococcus neoformans* and *Candida albicans*.^{7,24} Moreover, we observed an instance of mixed MTB infection, including co-infection with different strains or *Mycobacterium Avium* complex (MAC).⁶ Generally, MTB and NTM exhibit similar clinical manifestations, complicating the differential diagnosis. The most abundant species identified in Non-TB-NTM subgroups were *Human betaherpesvirus 7*, *Klebsiella pneumoniae*, and *Escherichia coli*. *Human betaherpesvirus 7* is a common virus that can cause latent infections and may contribute to immune dysregulation, particularly in immunocompromised individuals.²⁵ *Klebsiella pneumoniae*, an opportunistic pathogen, is known for causing a wide range of infections, including pneumonia, which can complicate TB diagnosis and treatment due to similar clinical manifestations.²⁶ *Escherichia coli*, while primarily associated with gastrointestinal infections, can also cause respiratory infections, especially in immunocompromised patients.²⁷ In TB patients, the presence of these pathogens can exacerbate the clinical course and complicate treatment. For instance, *Klebsiella pneumoniae* is known for its ability to evade the host immune system through various virulence factors, such as capsules and siderophores, which enhance its pathogenicity.²⁶ The pathogen spectrum in patients was more diverse and complex, with co-infected patients showing stronger associations between pathogens. We propose that these findings may be linked to an increased risk of opportunistic infections resulting from immune dysfunction in TB patients, particularly among those co-infected with NTM. Further validation of this hypothesis will require additional clinical data.²⁸

Dysbiosis of lung microbiota may play a significant role in the pathophysiological processes associated with TB. Research on TB-associated lung microbiota is still in its infancy, in contrast to the extensively studied gut microbiota in patients with TB patients.²⁹ This study used mNGS to evaluate the lung microbiota across different cohorts of patients with suspected pulmonary TB. Our aim was to identify variations in the composition and diversity of microbial communities in patients with or without TB as well as among patients with different infection patterns. BALF samples were selected for analysis to accurately reflect the lung microbiota, BALF samples were selected for analyses.³⁰ Our findings revealed that different patient cohorts predominantly shared similar lung microbial species, with the most frequently identified being *Prevotella melaninogenica*, *Streptococcus mitis*, *Prevotella jejuni*, and *Veillonella parvula*. Notably, *Prevotella* and *Veillonella* spp. are known to produce short-chain fatty acids, enhance immune responses, and suppress inflammation at other mucosal sites.³¹ Although previous studies have reported significant differences in the alpha and beta diversities of microbial communities in the respiratory tract of patients with TB compared to healthy individuals,^{29,32} our results did not align, revealing no significant differences in microbial diversity. Variations in study outcomes may be attributed to differences in populations, experimental protocols, geographic factors, sample size, and sample types (eg, BALF versus sputum), as well as differences in the study cohorts used for comparison.

Although the overall microbial composition and diversity did not exhibit significant alterations, specific microorganisms were selectively enriched in different patient groups, as determined by LEfSe analysis. The MTBC, MTB, *Streptococcus infantis* and *Campylobacter curvus* were significantly more abundant in the TB group. Conversely, *Corynebacterium striatum*, *Staphylococcus epidermidis*, *Ralstonia mannitolilytica* and *Mycoplasma hominis* were significantly enriched in the Non-TB group. In addition to MTB, the most abundant species in patients with TB include

Streptococcus infantis and *Campylobacter curvus*. In contrast, the enriched microbes in patients were primarily commensal microorganisms, which may have the potential to become opportunistic pathogens. *Ralstonia* spp. are aerobic, gram-negative, non-fermenting bacteria that constitute part of the normal microbiota of the oral and upper respiratory tracts but are increasingly recognized as opportunistic pathogens in the lower respiratory tract.²⁹ Subgroup comparisons showed that the TB+NTM subgroup was enriched in *Mycobacterium canettii*, whereas the NTM subgroup showed higher levels of *Bacteroides zoogloformans*, *Campylobacter curvus* and *Enterobacter cloacae* complex. However, the underlying mechanisms for the prevalence of these species with different infection patterns warrant further investigation.

Our analysis revealed that the abundance of the MTBC was significantly negatively correlated with HGB levels, whereas it was significantly positively correlated with CRP levels. TB is a chronic infectious disease characterized by elevated levels of inflammatory markers that may lead to metabolic defects that increase HGB consumption.⁵ Low HGB levels are often associated with anemia, which can increase susceptibility to infections, including TB, by impairing immune function.^{33,34} Thus, our findings suggest that anemia may be a risk factor for TB. Biomarkers represent an emerging area of research in the diagnosis and treatment. A recent investigation employed a random forest model for feature selection and biomarker screening in patients with lung cancer based on microbial profiles identified through mNGS.³⁵ To further evaluate the diagnostic performance of specific microbes in patients with TB, we constructed random forest classifiers using these differential microbial species, in conjunction with five clinical indices identified as potential biomarkers. The classification model demonstrated promising diagnostic capabilities for pulmonary TB, achieving an AUC $\geq$ 0.8. Previous studies have employed random forest screening to predict biomarkers for TB and have reported similar diagnostic performance.¹³

Nevertheless, several limitations must be considered in the present study. First, healthy lung microbiota were not included, as obtaining BALF samples from healthy individuals poses significant challenges. Future investigations should aim to recruit healthy volunteers for BALF sampling to facilitate comparative analysis of the lung microbiota. Second, this was a retrospective single-center study with a relatively small cohort of patients. Future research should include a larger sample size from diverse regions of China to validate our findings further. Third, our study did not include a longitudinal analysis of TB patients. Analyzing the lung microbiota and immune response of patients with TB throughout the treatment course would likely provide valuable insights into the dynamics of the lung microbiome during infection and the effects of various treatment modalities.

Conclusion

Our findings highlight the complexity of co-infection patterns in pulmonary TB and emphasize the potential of integrating microbial and clinical markers to improve diagnostic accuracy. This study provides valuable insights into the role of the lung microbiome in TB and informs future research on targeted therapies for this disease. In conclusion, the present study identified differences in the pathogen spectra and specific microbial species between TB and Non-TB patients. Several differential lung microbes and clinical parameters, such as MTBC, MTB, *Streptococcus infantis*, and *Campylobacter curvus*, may serve as potential biomarkers to distinguish patients with TB from those without TB. These findings may enhance the diagnosis and management of both TB and Non-TB patients, thereby improving clinical outcomes.

Data Sharing Statement

Metagenomic sequencing data without reads of the human genome are publicly available in NCBI under the SRA accession PRJNA1052360.

Ethical Approval Statement

This study was approved by the Ethics Committee of Yongkang First People's Hospital (YKSDYRMYYEC2022-KT-HS-001-01) and conducted in accordance with the principles of the Declaration of Helsinki and relevant ethical and legal requirements.

Funding

This work was supported by grants from the Jinhua Key Science and Technology Program Projects (Grant No. 2022-3-050).

Disclosure

The authors report no conflicts of interest in this work.

References

- Huang Y, Ai L, Wang X, Sun Z, Wang F. Review and updates on the diagnosis of tuberculosis. *J Clin Med*. 2022;11(19):5826. doi:10.3390/jcm11195826
- Global tuberculosis report;2021. Available from: <https://www.who.int/publications/i/item/9789240037021>. Accessed on April 11, 2025.
- Falzon D, Zignol M, Bastard M, Floyd K, Kasaeva T. The impact of the COVID-19 pandemic on the global tuberculosis epidemic. *Front Immunol*. 2023;14:1234785. doi:10.3389/fimmu.2023.1234785
- Dheda K, Perumal T, Moultrie H, et al. The intersecting pandemics of tuberculosis and COVID-19: population-level and patient-level impact, clinical presentation, and corrective interventions. *Lancet Respir Med*. 2022;10(6):603–622. doi:10.1016/S2213-2600(22)00092-3
- Ding L, Liu Y, Wu X, et al. Pathogen metagenomics reveals distinct lung microbiota signatures between bacteriologically confirmed and negative tuberculosis patients. *Front Cell Infect Microbiol*. 2021;11:708827. doi:10.3389/fcimb.2021.708827
- Khan Z, Miller A, Bachan M, Donath J. Mycobacterium Avium Complex (MAC) lung disease in two inner city community hospitals: recognition, prevalence, co-infection with Mycobacterium Tuberculosis (MTB) and Pulmonary Function (PF) improvements after treatment. *Open Respir Med J*. 2010;4:76–81. doi:10.2174/1874306401004010076
- Shi CL, Han P, Tang PJ, et al. Clinical metagenomic sequencing for diagnosis of pulmonary tuberculosis. *J Infect*. 2020;81(4):567–574. doi:10.1016/j.jinf.2020.08.004
- Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet*. 2019;20(6):341–355. doi:10.1038/s41576-019-0113-7
- Ramachandran PS, Wilson MR. Metagenomics for neurological infections - expanding our imagination. *Nat Rev Neurol*. 2020;16(10):547–556. doi:10.1038/s41582-020-0374-y
- Sun W, Lu Z, Yan L. Clinical efficacy of metagenomic next-generation sequencing for rapid detection of mycobacterium tuberculosis in smear-negative extrapulmonary specimens in a high tuberculosis burden area. *Int J Infect Dis*. 2021;103:91–96. doi:10.1016/j.ijid.2020.11.165
- Fu M, Cao LJ, Xia HL, et al. The performance of detecting mycobacterium tuberculosis complex in lung biopsy tissue by metagenomic next-generation sequencing. *BMC Pulm Med*. 2022;22(1):288. doi:10.1186/s12890-022-02079-8
- Wypych TP, Wickramasinghe LC, Marsland BJ. The influence of the microbiome on respiratory health. *Nat Immunol*. 2019;20(10):1279–1290. doi:10.1038/s41590-019-0451-9
- Hu Y, Feng Y, Wu J, et al. The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. *Front Cell Infect Microbiol*. 2019;9:90. doi:10.3389/fcimb.2019.00090
- Li H, Wu X, Zeng H, et al. Unique microbial landscape in the human oropharynx during different types of acute respiratory tract infections. *Microbiome*. 2023;11(1):157. doi:10.1186/s40168-023-01597-9
- Beck JM, Schloss PD, Venkataraman A, et al. Multicenter comparison of lung and oral microbiomes of HIV-infected and HIV-uninfected individuals. *Am J Respir Crit Care Med*. 2015;192(11):1335–1344. doi:10.1164/rccm.201501-0128OC
- Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol*. 2017;15(5):259–270. doi:10.1038/nrmicro.2017.14
- Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One*. 2012;7(8):e43093. doi:10.1371/journal.pone.0043093
- Miao Q, Ma Y, Ling Y, et al. Evaluation of superinfection, antimicrobial usage, and airway microbiome with metagenomic sequencing in COVID-19 patients: a cohort study in Shanghai. *J Microbiol Immunol Infect*. 2021;54(5):808–815. doi:10.1016/j.jmii.2021.03.015
- Ao Z, Xu H, Li M, Liu H, Deng M, Liu Y. Clinical characteristics, diagnosis, outcomes and lung microbiome analysis of invasive pulmonary aspergillosis in the community-acquired pneumonia patients. *BMJ Open Respir Res*. 2023;10(1):e001358. doi:10.1136/bmjresp-2022-001358
- Tan Y, Chen Z, Zeng Z, et al. Microbiomes detected by bronchoalveolar lavage fluid metagenomic next-generation sequencing among HIV-infected and uninfected patients with pulmonary infection. *Microbiol Spectr*. 2023;11(4):e0000523. doi:10.1128/spectrum.00005-23
- Xiao G, Cai Z, Guo Q, et al. Insights into the unique lung microbiota profile of pulmonary tuberculosis patients using metagenomic next-generation sequencing. *Microbiol Spectr*. 2022;10(1):e0190121. doi:10.1128/spectrum.01901-21
- Meyer KC, Raghu G, Baughman RP, et al. An official American thoracic society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *Am J Respir Crit Care Med*. 2012;185(9):1004–1014. doi:10.1164/rccm.201202-0320ST
- Sulaiman I, Wu BG, Li Y, et al. Evaluation of the airway microbiome in nontuberculous mycobacteria disease. *Eur Respir J*. 2018;52(4):1800810. doi:10.1183/13993003.00810-2018
- Zhou X, Wu H, Ruan Q, et al. Clinical evaluation of diagnosis efficacy of active mycobacterium tuberculosis complex infection via metagenomic next-generation sequencing of direct clinical samples. *Front Cell Infect Microbiol*. 2019;9:351. doi:10.3389/fcimb.2019.00351
- Verbeek R, Vandekerckhove L, Van Cleemput J. Update on human herpesvirus 7 pathogenesis and clinical aspects as a roadmap for future research. *J Virol*. 2024;98(6):e00437–24. doi:10.1128/jvi.00437-24
- Abbas R, Chakkour M, Zein El Dine H, et al. General overview of Klebsiella pneumonia: epidemiology and the role of siderophores in its pathogenicity. *Biology*. 2024;13(2):78. doi:10.3390/biology13020078
- Mueller M, Tainter CR. Escherichia coli Infection. In: *StatPearls*. StatPearls Publishing; 2025.
- José RJ, Brown JS. Opportunistic bacterial, viral and fungal infections of the lung. *Medicine*. 2016;44(6):378–383. doi:10.1016/j.mpm.2016.03.015
- Valdez-Palomares F, Muñoz Torrico M, Palacios-González B, Soberón X, Silva-Herzog E. Altered microbial composition of drug-sensitive and drug-resistant tb patients compared with healthy volunteers. *Microorganisms*. 2021;9(8):1762. doi:10.3390/microorganisms9081762

30. Vázquez-Pérez JA, Carrillo CO, Iñiguez-García MA, et al. Alveolar microbiota profile in patients with human pulmonary tuberculosis and interstitial pneumonia. *Microb Pathog.* 2020;139:103851. doi:10.1016/j.micpath.2019.103851
31. Kim CH. Control of lymphocyte functions by gut microbiota-derived short-chain fatty acids. *Cell mol Immunol.* 2021;18(5):1161–1171. doi:10.1038/s41423-020-00625-0
32. Hu Y, Kang Y, Liu X, et al. Distinct lung microbial community states in patients with pulmonary tuberculosis. *Sci China Life Sci.* 2020;63(10):1522–1533. doi:10.1007/s11427-019-1614-0
33. Gelaw Y, Getaneh Z, Melku M. Anemia as a risk factor for tuberculosis: a systematic review and meta-analysis. *Environ Health Prev Med.* 2021;26(1):13. doi:10.1186/s12199-020-00931-z
34. Araújo-Pereira M, Krishnan S, Salgame P, et al. Effect of the relationship between anaemia and systemic inflammation on the risk of incident tuberculosis and death in people with advanced HIV: a sub-analysis of the REMEMBER trial. *EClinicalMedicine.* 2023;60:102030. doi:10.1016/j.eclinm.2023.102030
35. Chen Q, Hou K, Tang M, et al. Screening of potential microbial markers for lung cancer using metagenomic sequencing. *Cancer Med.* 2023;12(6):7127–7139. doi:10.1002/cam4.5513

Infection and Drug Resistance

Publish your work in this journal

Infection and Drug Resistance is an international, peer-reviewed open-access journal that focuses on the optimal treatment of infection (bacterial, fungal and viral) and the development and institution of preventive strategies to minimize the development and spread of resistance. The journal is specifically concerned with the epidemiology of antibiotic resistance and the mechanisms of resistance development and diffusion in both hospitals and the community. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/infection-and-drug-resistance-journal>

Dovepress
Taylor & Francis Group