Clinical characteristics, diagnosis, outcomes and lung microbiome analysis in invasive pulmonary aspergillosis in the community-acquired pneumonia patients

Zhi Ao,¹ Huan Xu,² Mengqi Li,¹ Huifang Liu,² Min Deng,¹ Yuliang Liu¹

ABSTRACT
Background Invasive pulmonary aspergillosis (IPA) remains underestimated in patients with community-acquired pneumonia (CAP). This study aims to describe clinical features and outcomes of IPA in CAP patients, assess diagnostic performance of metagenomic next-generation sequencing (mNGS) for IPA and analyse lung microbiome via mNGS data.

Methods This retrospective cohort study included CAP patients from 22 April 2019 to 30 September 2021. Clinical and microbiological data were analysed. Diagnostic performance of mNGS was compared with traditional detection methods. The lung microbiome detected by mNGS was characterised and its association with clinical features was evaluated.

Main results IPA was diagnosed in 26 (23.4%) of 111 CAP patients. Patients with IPA displayed depressed immunity, higher hospital mortality (30.8% vs 11.8%) and intensive care unit mortality (42.1% vs 17.5%) compared with patients without IPA. The galactomannan (GM) antigen test had the highest sensitivity (57.7%) in detecting the Aspergillus spp, followed by mNGS (42.3%), culture (30.8%) and smear (7.7%). The mNGS, culture and smear had 100% specificity, while GM test had 92.9% specificity. The microbial structure of IPA significantly differed from non-IPA patients (p<0.001; Wilcoxon test). Nineteen different species were significantly correlated with clinical outcomes and laboratory biomarkers, particularly for Streptococcus salivarius, Prevotella timonensis and Human beta herpes virus 5.

Conclusions Our results reveal that patients with Aspergillus infection tend to have a higher early mortality rate. The mNGS may be suggested as a complement to routine microbiological test in diagnosis of patients at risk of Aspergillus infection. The lung microbiota is associated with inflammatory, immune and metabolic conditions of IPA, and thus influences clinical outcomes.

INTRODUCTION
Invasive pulmonary aspergillosis (IPA) is an important cause of morbidity and mortality, typically affecting immunocompromised hosts, such as patients with HIV infection as well as patients receiving corticosteroids therapy, transplant recipients or haematologic cancer.1 Numerous cases of misdiagnosis or missed diagnosis and only diagnosed post-mortem due to lack of medical resource and new risk factors are underestimated.² Clinical diagnosis of IPA comprises host factors, clinical features and mycological evidence.³ Problems arising from the inaccessibility of tissue
samples, low sensitivity of culture, poor applicability of traditional host factors and absence of typical clinical and radiological findings lead to delayed and missed diagnosis in IPA patients. The recently published American Thoracic Society clinical practice guidelines for the diagnosis of fungal infection highlight the culture-independent diagnostic tests, with strong recommendations given to the detection of Aspergillus DNA in bronchoalveolar lavage fluid (BALF) by PCR.

In the past decade, development in molecular methods for the quantification and sequencing of pathogenic DNA has greatly facilitated the detection of various microorganisms. Metagenomic next-generation sequencing (mNGS) has been now widely applied to the diagnosis of infectious diseases in clinical settings, especially with specific or unidentified or mixed pathogens in a non-targeted approach. For example, combining mNGS with conventional detection methods could increase the detection rate for mycobacterium tuberculosis complex with conventional detection methods could increase the detection rate for mycobacterium tuberculosis complex.5

Aspergillus DNA in bronchial secretions given to the detection of IPA has been discovered patients with IPA exhibited a differential abundance of specific bacterial taxa.11 With newer diagnostic techniques available, it is feasible to evaluate clinical characteristics and lung microbiome of IPA from a new perspective for a comprehensive understanding of the disease. This study was designed to describe clinical characteristics and outcomes of IPA in CAP patients, assess diagnostic performance of mNGS for IPA and reveal the relationship between clinical characteristics and lung microbiome in IPA.

METHODS
Study design and participants
This was a retrospective cohort study that consisted of hospitalised patients (≥18 years of age) diagnosed as CAP at The First Affiliated Hospital of Chongqing Medical University between 22 April 2019 and 30 September 2021. Patients enrolled in this study should meet the following inclusive criteria: (1) diagnosed with CAP meeting the 2007 IDSA/ATS guideline, (2) undergoing bronchoalveolar lavage, (3) no antibiotic exposure in the last 2 weeks before sample collection and (4) having galactomannan (GM) test, smear, culture and BALF mNGS results of BALF. Cases with probable IPA included patients with no evidence of clinical or radiological evidence or the presence of Aspergillus spp in the BALF. All the enrolled patients were followed up by telephone to inquire about survival or date of death.

Data collection and microbiological testing
Clinical information was extracted from the electronic medical records using a standardised data collection form, which included demographic information, clinical data, laboratory findings, CT of the chest, antifungal treatments and outcomes. All the data were reviewed by two investigators independently to verify data accuracy. The 7-category ordinal scale was modified based on the study design, which consisted of mutually exclusive categories as follows: category 7, death; category 6, requiring invasive ventilation and additional organ support (eg, continuous renal replacement therapy or extracorporeal membrane oxygenation); category 5, requiring invasive mechanical ventilation; category 4, requiring non-invasive ventilation or high-flow oxygen; category 3, requiring supplemental oxygen by mask or nasal prongs; category 2, hospitalised, no oxygen; category 1, discharged. Improvement in clinical status was defined as a decline of 2 categories assessed by the 7-level ordinal scale at fixed time points (day 1, 7, 14, 21 and 28).

Testing for GM antigen in serum and GM in BALF is generally performed in all enrolled patients. A result was considered positive when optical index values were ≥1.0 in serum or ≥0.7 in BALF or ≥0.8 in BALF. BALF specimens were subjected to mNGS, smear by microscopy and ordinary fungal culture.

Metagenomics next-generation sequencing
A 1.5–3 mL of BALF was collected from each patient according to standard procedures. A 1.5 mL microcentrifuge tube containing 0.6 mL BALF sample, enzyme and 1 g of 0.5 mm glass beads was agitation by a vortex mixer, and a 0.3 mL sample was separated into a 1.5 mL microcentrifuge tube and DNA was extracted using the TIANamp Micro DNA Kit (DP316, Tiangen Biotech) following the manufacturer’s manual. DNA libraries were constructed through DNA fragmentation, end repair, adapter ligation and PCR amplification. Then, libraries with confirmed quality were sequenced by the illumina NextSeq 550 platform with single reads of 75 bp (Illumina, San Diego, California, USA).

The internal control, named unique molecular spiked-in (UMSI), was added to the sample before the DNA extraction. The sequence of UMSI varied in at least one host factor, a clinical manifestation and mycologic evidence in BALF, according to the 2019 criteria from the European Organization for Research and Treatment of Cancer/Mycoses Study Group (EORTIC/MSG). In this study, a modification in diagnostic criteria was the addition of mNGS as one of mycologic testing methods including GM test, smear and culture. Cases without probable IPA included patients with no evidence of clinical or radiological evidence or the presence of Aspergillus spp in the BALF. All the enrollees were followed up by telephone to inquire about survival or date of death.
different samples. Each mNGS assay run included an external negative control that ran in parallel with clinical samples. During analysis, the contamination between samples could be found if the UMSI sequence was the same or the reads of some pathogens in the external control were very high.

High-quality sequencing data were generated by removing low-quality reads, followed by computational subtraction of human host sequences mapped to the human reference genome (hg19) using Burrows-Wheeler alignment. The remaining data by removal of low-complexity reads were classified by simultaneously aligning to four microbial genome databases (bacteria, fungi, viruses and parasites), which were downloaded from the National Center for Biotechnology Information (ftp://ftp.ncbi.nlm.nih.gov/genomes/). Interpretation of mNGS results can be seen in online supplemental methods.

Patient and public involvement
Patients and the public were not involved in the design and conduct of this study, choice of outcome measures, nor recruitment. The results of this study will not be disseminated to the participant and linked communities.

Statistical analyses
Continuous variables were expressed as median (IQR) and compared with the Mann-Whitney U test. Categorical variables were expressed as number (%) and compared by χ² test or Fisher’s exact test. The 2×2 contingency tables were established to determine sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). A two-sided α of less than 0.05 was considered statistically significant. Statistical analyses were done using the IBM SPSS Statistics, V.23.0, unless otherwise indicated.

The high-quality reads were first mapped to the human reference genome (hg19) and then non-human reads were aligned to four microbial genome databases (bacteria, fungi, viruses and parasites). After removing the background microbial which appeared in negative controls (online supplemental table 1), Bayesian re-estimation of abundance was used to estimate the relative abundance of species (online supplemental table 2). The alpha diversity index for each sample was calculated based on Shannon and Simpson indexes, and beta diversity was calculated based on weighted UniFrac distance and Bray-Curtis distance. P values for alpha and beta analysis were calculated using Wilcoxon rank-sum test. Linear discriminant analysis (LDA) effect size was used to find significantly different species between groups, with thresholds of log10 LDA Score≥2 and p values<0.05. The correlation between the representative microbiota and clinical data was evaluated by Spearman.

RESULTS
Patient recruitment and clinical characteristics
A total of 123 patients with 128 BALF specimens were reviewed and 111 patients were enrolled in this study. Twenty-six patients were diagnosed with probable IPA (named IPA) and 85 controls without IPA (named control) (figure 1). Comparisons of demographic and clinical characteristics between IPA group and control group were summarised in table 1. The study population had a median age of 60 (IQR, 53–73) years and was 66.7% were men. Overall, 33 patients (29.7%) had a history of smoking. Twenty-two patients (19.8%) had chronic obstructive pulmonary disease. Fifteen patients (13.5%) had received corticosteroids and 36 patients (32.4%) were considered to be immunocompromised (online supplemental table 3). Patients with probable IPA had a significantly higher rate with known risk factors for aspergillosis than those without IPA, including treatment with corticosteroids longer than 28 days (38.5% vs 5.9%, p=0.000) and immunocompromised status (33.8% vs 25.9%, p=0.008). No significant differences were found in sequential organ failure assessment scoring and confusion, urea, respiratory rate, blood pressure, age 65 scoring of patients between two groups on admission.

Microbiological, CT images and laboratory findings
Aspergillus fumigatus isolates were obtained for 6 of 26 (23.1%) probable IPA cases, followed by Aspergillus flavus for 3 of 26 (11.5%) and Aspergillus ustus for 1 of 26 (3.8%) (table 2). Bronchial wall thickening was more often shown on CT of the chest in the IPA (44.0% vs 12.8%, p=0.000) (table 2). IPA patients have significantly lower level of total lymphocyte count (774.1 vs 1088.9, p=0.009) and lymphocyte subpopulation, including CD3⁺ T lymphocyte count (553.0 vs 782.6, p=0.016), CD4⁺ T lymphocyte count (287.9 vs 409.5, p=0.031) and CD4⁻ CD8⁻ T lymphocyte count (25.9 vs 45.0, p=0.007) than control patients (online supplemental table 4), indicating the compromised immune function of IPA patients.

Comparison of diagnostic performance
To compare the diagnostic performance of BALF mNGS and conventional tests for distinguishing cases with probable IPA from cases without probable IPA, the results were...
showed in figure 2. GM test (57.7%) had the highest sensitivity in detecting the *Aspergillus* spp, followed by mNGS (42.3%), culture (30.8%) and smear (7.7%). Comparing with smear, mNGS had the significantly higher sensitivity 42.3% vs 7.7%, p=0.01. The mNGS, culture and smear had 100% specificity, while GM test had 92.9% specificity, with significant difference between mNGS (or culture or smear) and GM test (p=0.029). The PPV of mNGS, culture and smear for identifying *Aspergillus* spp was 100%, while GM test was 71.4%. The NPV ranked as GM test (87.8%), mNGS (85.0%), culture (82.5%) and smear (78.0%), with no significant difference among these methods. We have analysed differences in sensitivity/specificity/PPV/NPV of GM on serum versus BALF, using a BAL fluid GM index of ≥1.0 or serum GM index of ≥1.0 as positive, according to 2019 EORTIC/MSG. It was significantly higher in sensitivity but lower in specificity of GM BALF than that of GM serum (57.7% vs 26.9%, p=0.048, 92.9% vs 100%, 0.029, respectively), which was shown in online supplemental table 5. There were no differences in PPV and NPV.

### Clinical outcomes

There were no significant differences in 28-day, 60-day and 90-day mortality between the two groups (table 3). Hospital mortality (30.8% vs 11.8%, p=0.021) and intensive care unit (ICU) mortality rates (42.1% vs 17.5%,
p=0.029) were significantly higher in the IPA group than in the control group. The distribution of patients falling into each category of the 7-category scale showed statistical difference between the two groups by day 28 (p=0.017); however, significance with statistical testing was not reached in the 7-category ordinal scale by days 1, 7 and 14 (figure 3). Multiple antifungal drugs were administered for 23.1% of patients with probable IPA, compared with 5.9% of patients without probable IPA (p=0.01). The proportion of patients receiving voriconazole was remarkably higher in cases with probable IPA (p=0.01). The proportion of patients receiving voriconazole was remarkably higher in cases with probable IPA (53.9% vs 8.2%, p=0.000). There was no significant difference in the use of amphotericin B and caspofungin compared with 5.9% of patients without probable IPA, administered for 23.1% of patients with probable IPA, 1, 7 and 14 (figure 3). Multiple antifungal drugs were administered for 23.1% of patients with probable IPA, compared with 5.9% of patients without probable IPA (p=0.01). The proportion of patients receiving voriconazole was remarkably higher in cases with probable IPA (53.9% vs 8.2%, p=0.000). There was no significant difference in the use of amphotericin B and caspofungin between the two groups (table 3).

Lung microbiome analysis

To compare the overall composition and diversity of the lung microbial signature in patients with IPA and the control, we analysed BALF specimens collected from 109 CAP patients, including 24 cases (missing 2 patients due to loss of mNGS data) diagnosed with probable IPA (named IPA) and 85 cases without probable IPA (named control). We assessed α diversity of the lung communities using the Shannon and Simpson diversity index. There was a decrease in α diversity both in Shannon and Simpson diversity index in IPA patients as compared with non-IPA patients in spite of significant difference (figure 4A,B). Using the weighted UniFrac metric and Bray-Curtis metric, we observed that β diversity of IPA patients significantly differed from non-IPA patients (p<0.001; Wilcoxon test), suggesting that lung community structure of patients diagnosed with probable IPA differed substantially from those without probable IPA (figure 4C,D). Based on the average relative abundance, we plotted the top 10 phyla, genus and species among 2 groups. In the overall cohort, Firmicutes, Proteobacteria, Actinobacteria, Bacteroidetes and Ascomycota were the most abundant phyla, whereas Prevotella, Streptococcus, Acinetobacter, and Pneumocystis were the most common genera (figure 5A,B). The relative abundance of the top 10 in species ranked as Pneumocystis jirovecii, Acinetobacter baumannii, Lautropia mirabilis, Streptococcus oralis, Corynebacterium striatum, Human betaherpesvirus 5, Rothia mucilaginosa, Staphylococcus aureus, Prevotella melaninogenica, Nocardia Farcinica, with only H. betaherpesvirus 5 was the significantly different between two groups (figure 5C). Focusing on the differential species in the IPA compared with their controls, we found 21 species with LDA Score ≥2 and p<0.05, which H. betaherpesvirus 5, A. fumigatus, Aspergillus niger, Citrobacter braakii, Barillus thermoamylovorans, Helcococcus kunzii, Lactobacillus

---

**Table 2** Microbiological findings and CT images of the study cohort

<table>
<thead>
<tr>
<th>Thoracic CT findings</th>
<th>Suspected CAP cohort (n=111)</th>
<th>With probable Aspergillus infection (n=26; 23.4%)</th>
<th>Without probable Aspergillus infection (n=85; 76.8%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consolidation</td>
<td>24 (21.6)</td>
<td>7 (28.0)</td>
<td>17 (19.8)</td>
<td>0.453</td>
</tr>
<tr>
<td>Micronodules</td>
<td>11 (9.9)</td>
<td>2 (8.0)</td>
<td>9 (10.5)</td>
<td>0.954</td>
</tr>
<tr>
<td>Cavity or air-crescent sign</td>
<td>17 (15.3)</td>
<td>7 (28.0)</td>
<td>10 (11.6)</td>
<td>0.060</td>
</tr>
<tr>
<td>Ground glass opacity</td>
<td>64 (57.7)</td>
<td>17 (64.0)</td>
<td>47 (55.8)</td>
<td>0.362</td>
</tr>
</tbody>
</table>

Data are given as n (%), unless otherwise specified. When no denominator is specified, the denominator is the number of participants in the corresponding cell in the first row. P values were calculated by Fisher’s exact probability t-test for categorical variables.

CAP, community-acquired pneumonia.

---

**Figure 2** Diagnostic performance of bronchoalveolar lavage fluid mNGS and conventional testing for differentiating probable invasive pulmonary aspergillosis (IPA) from non-probable IPA. Gm, galactomannan; mNGS, metagenomic next-generation sequencing; neg, negative; NPV, negative predictive value; pos, positive; PPV, positive predictive value.
delbrueckii, Burkholderia dolosa, Marinobacter hydrocarbonoclasticus, Riemerella anatipestifer, Corynebacterium halotolerans and Lactobacillus plantarum were significantly abundant in the cases diagnosed with probable IPA, and Streptococcus salivarius, Citrobacter freundii, Paraburkholderia fungorum, Dolosigranulum pigrum, Prevotella timonensis, Sphingobium yanoikuyae, Serratia marcescens and Corynebacterium oculi were enriched in cases without probable IPA (figure 5D, online supplemental figure 1).

To further investigate the correlation between clinical data and the lung microbiota of IPA, we performed Spearman’s rank-based correlation test on the 21 significantly different species with clinical outcome and laboratory findings (figure 6A). We observed 19 different species were significantly correlated with 7 clinical outcomes and 22 laboratory biomarkers, particularly S. salivarius, P. timonensis and H. betaherpesvirus 5. S. salivarius was primarily positively correlated with eight laboratory biomarkers, including red blood cell, total lymphocytes, T lymphocytes, T helper cells, cytotoxic T lymphocytes, B lymphocytes, serum potassium and albumin, and negatively relative with aspartate aminotransferase, direct bilirubin and D-dimer. Whereas, S. salivarius was mainly negatively associated with duration of ICU stay and hospital mortality, but positively with ventilator-free days within 28 days. Opposite to S. salivarius and P. timonensis, H. betaherpesvirus 5.

<table>
<thead>
<tr>
<th>Table 3 Clinical outcomes of the study cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspected CAP cohort (n=111)</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td><strong>Mortality</strong></td>
</tr>
<tr>
<td>Mortality (28 days)</td>
</tr>
<tr>
<td>Mortality (60 days)</td>
</tr>
<tr>
<td>Mortality (90 days)</td>
</tr>
<tr>
<td>Hospital mortality</td>
</tr>
<tr>
<td>ICU mortality (76)</td>
</tr>
<tr>
<td><strong>Prognosis</strong></td>
</tr>
<tr>
<td>Improvement in clinical status, median (IQR), days</td>
</tr>
<tr>
<td>Ventilator-free days within 28 days, median (IQR)</td>
</tr>
<tr>
<td>Supplemental oxygen independence days within 28 days, median (IQR), days</td>
</tr>
<tr>
<td>Duration of ICU stay, median (IQR), days</td>
</tr>
<tr>
<td>Duration of hospital stay, median (IQR), days</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
</tr>
<tr>
<td>Multiple antifungal drugs*</td>
</tr>
<tr>
<td>Main antifungal therapy†</td>
</tr>
<tr>
<td>Voriconazole</td>
</tr>
<tr>
<td>Amphotericin B</td>
</tr>
<tr>
<td>Caspofungin</td>
</tr>
</tbody>
</table>

Data are n (%) or median (IQR), unless otherwise specified. When no denominator is specified, the denominator is the number of participants in the corresponding cell in the first row. Parenthetical values in the stub column represent the number of patients for whom data were available.

*Multiple antifungal drugs: used in combination or consecutively.†Main antifungal therapy: first antifungal drug administered for ≥5 consecutive days following invasive pulmonary aspergillosis diagnosis.

CAP, community-acquired pneumonia; ICU, intensive care unit.
betaherpesvirus 5 was negatively associated with six laboratory biomarkers, including total lymphocyte, T lymphocytes, T helper cells, CD4+CD8+ T cells, B lymphocytes and serum potassium, and positively relative with blood urea nitrogen. Meanwhile, H. betaherpesvirus 5 was mainly negatively correlated with duration of ICU stay and time to supplemental oxygen independence within 28 days. These observations suggested the dysbiosis of microbial communities in the lung is associated with the patient’s pathophysiological conditions and thus can influence the clinical outcomes.

To further investigate the clinical impact of 21 significantly different species, we made the survival analysis by plotting the cumulative probability curve. As shown in figure 6B,C, detection of A. fumigatus and H. betaherpesvirus 5 was significantly predictive of worse ICU outcomes (p=0.045, 0.032). While in figure 6D,E, we found that detection of S. salivarius and P. timonensis was extremely significantly predictive of improved ICU outcomes (p=0.0031, 0.0054). We thus concluded that in the lung microbiota of CAP, ICU outcomes may be predicted by community composition, specifically A. fumigatus, H. betaherpesvirus 5, S. salivarius and P. timonensis.

DISCUSSION
In this retrospective cohort study, we examined clinical characteristics of IPA. Interestingly, we noted that patients categorised as having probable Aspergillus infection had a significantly higher hospital mortality (30.8%) and ICU mortality (42.1%) compared with the control. The mortality at day 28, though not statistically significant,
was numerically higher in IPA patients than the control. These data suggested that patients with *Aspergillus* infection tend to have a higher early mortality rate. Previous study reported an estimated 45% of ICU mortality and 26% of hospital mortality. Among invasive tracheobronchial aspergillosis, a rare entity of IPA, 30-day and 90-day mortality rates were both as high as 90%. Loughlin *et al* reported that ICU mortality in patients with probable *Aspergillus* infection was not significantly higher than those without *Aspergillus* infection, which possibly due to generally critical status in ventilator-associated pneumonia patient cohort in the ICU setting. In this study, the overall condition of the enrolled patients with CAP was relatively mild, which can be remarkably aggravated by *Aspergillus* infection and potentially life-threatening.

Cornillet *et al* have suggested that IPA patients’ non-specific clinical features and the diagnostic difficulties particularly when patients without neutropenia easily lead to suboptimal management and delayed initiation of treatment, resulting in high mortality and morbidity. Given the devastating mortality rates of patients with IPA, timely and accurate diagnosis is highlighted and urgently needed for effective antifungal treatment to decrease the morbidity and mortality rate. In the past decade, GM detection has been widely used in the diagnosis of IPA. However, GM testing for serum samples has limited sensitivity in diagnose IPA in non-neutropenic hosts. Instead, detection of GM in BAL fluid samples of patients at risk of IPA has an excellent diagnostic accuracy, because GM is released earlier and at higher concentrations in the BAL fluid in comparison with serum. Recently, mNGS has been playing an increasingly important role in the diagnosis of fungal diseases. Ding *et al* demonstrated that the sensitivity of mNGS was significantly higher than culture in detecting *Candida albicans*. In this study, BALF was collected directly from the lower respiratory tract, which avoided contamination from the oral microbiota. *Aspergillus* in BALF were detected by mNGS, culture, smear and GM test. The results showed that mNGS had comparable specificity and PPV to culture in *Aspergillus* detection. Although the sensitivity of mNGS lower than GM test, mNGS showed more sensitive than smear and more specific than serum and BALF GM test. These results indicated that mNGS acts as a useful non-culture based test for detecting *Aspergillus*, suggesting as a complement to routine microbiological test in diagnosis of patients at risk of IPA.

Our study discovered that CAP patients with *Aspergillus* were more likely to coinfect with bacterial or (and) viral infection due to the immunocompromised status and corticosteroid treatment. Patients with influenza who were immunocompromised were reported to have incidence of IPA as high as 32%, whereas in the non-immunocompromised influenza case group, incidence was 14%. With the emergence of the COVID-19 pandemic and corticosteroids use in patients with severe COVID-19, COVID-19-associated pulmonary aspergillosis (CAPA) has been reported in a number of case reports. In a meta-analysis, the incidence and mortality of CAPA in the ICU were estimated to be 10.2% and 54.9%, respectively.
Due to the vast microbial reads in the mNGS data that can be used to analyse microbial signature across numerous diseases, our study attempted to explore the difference in structure and composition of microbiome between IPA patients and non-IPA CAP patients by analysing the mNGS data. Our present study comprised the lung microbiota of 24 IPA patients and 85 non-IPA CAP patients. The relationship of different microbiota and clinical data for all 109 CAP patients were investigated. For alpha diversity, there were no significant differences in the Shannon index and Simpson index of microbiota in IPA patients and non-IPA CAP patients, demonstrating that the with-in sample diversity of communities were similar between IPA patients and non-IPA patients. For beta diversity, our results displayed a dramatical difference in the weighted UniFrac distance (p=2.1×10^{-6}) and Bray-Curtis distance (p=0.033) between IPA and non-IPA CAP patients, indicating a potential effect of *Aspergillus* spp on lung microbiota structure in patients diagnosed with IPA. Similar studies have shown significant differences in beta diversity of microbiota between IPA and controls.\textsuperscript{11} This may be that the causative agents in the controls were complex, which can be any CAP pathogens (such as bacteria, virus, fungi, parasite and even atypical pathogens) except *Aspergillus*. Thus the lung microbial heterogeneity in the control was higher than IPA. Meanwhile, these studies demonstrated that *Firmicutes*, *Proteobacteria*, *Actinobacteria* and *Bacteroidetes* were the most abundant phyla in the IPA cohort.\textsuperscript{11} However, beyond the bacteria, our study found that the relative abundance of *Ascomycota* was also high in the level of phylum, for the metagenomics can harvest almost all the nucleic acid of microbiota, including bacteria, fungi, virus and parasite. The significantly differential species, particularly *S. salivarius*, *P. timonensis* and *H. betahemorrhagicus* 5, were strongly related with clinical data. *S. salivarius* and *P. timonensis* were positively related with laboratory biomarkers such as T lymphocytes and negatively related with clinical outcomes such as hospital mortality. Whereas, *H. betahemorrhagicus* 5 was negatively related with laboratory biomarkers such as T lymphocytes and positively related with clinical outcomes such as hospital mortality. Previous study demonstrated that *S. salivarius* is a probiotic bacterium that can inhibit pneumococcal adherence to pharyngeal epithelial cells.\textsuperscript{33} Also, another study proven that *S. salivarius* isolates had a superior ability to inhibit the growth of *Haemophilus influenzae*, *Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Streptococcus pyogenes* and *Staphylococcus aureus*.\textsuperscript{32} All these studies coincided our study that *S. salivarius* was positively correlated with the better clinical outcome of the patients. The possible mechanism of *S. salivarius* positively correlated with the improved outcome of the patients may be that they can inhibit the potential pathogens growth and activate the host immune system. For *P. timonensis*, previous study showed that *P. timonensis* could dramatically induce the expression of membrane-associated mucins in the three-dimensional endometrial epithelial cell model and formed the biofilm in the epithelial surface, which in turn influence the adhesion of other bacteria.\textsuperscript{34} The reason for the positive correlation with the better clinical outcome of the patients may be the colonisation resistance of *P. timonensis* in the lung and the activation of the host immune system. In clinical settings, *H. betahemorrhagicus* 5 can build life-long latent infection in the healthy people. However, once immunocompromised, *H. betahemorrhagicus* 5 can be reactivated, for example, solid organ transplant recipients and HIV-positive patients.\textsuperscript{34} Thus, the reactivation of the *H. betahemorrhagicus* 5 indicated the weakened immune system and the poor clinical outcomes of IPA patients.

Our study has several limitations. First, the retrospective study was designed to analyse collected data and limited to relatively small sample size, which could predispose to introducing bias from incomplete information. Second, this was the observation study. Though we found some interesting correlation between lung microbiota and clinical data, causal relationship needs to be resolved by prospective cohort study or mechanism experiment, such as animal experiment, which will be the next step of our study. Otherwise, the study identified patients with probable *Aspergillus* infection strictly based on new EORTC/MSG 2019 diagnostic criteria so that a few cases could not satisfy the criteria for a probable IPA diagnosis due to absence of classic host factors.

Contributors Study concept and design: YL, ZA and HX. Acquisition of data: ZA, ML, HL and MD. Analysis and interpretation of data: ZA and HX. Drafting of the manuscript: ZA, HX and YL. Critical revision of the manuscript for important intellectual content: HX, ZA and YL. Statistical analysis: ZA, HL and HX. Administrative, technical or material support: HX, HL and YL. All authors contributed to the article and approved the submitted version.

Funding Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau (2022GRCQ010)).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. This study was approved by the ethical committee of the First Affiliated Hospital of Chongqing Medical Hospital (approved number 2021-301). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.
REFERENCES


