Metagenomic Next-generation Sequencing Indicates More Precise Pathogens in patients with pulmonary infection: a retrospective study

Dengfeng Wu
The Second Affiliated Hospital of Nanchang University
Wei Wang
The Second Affiliated Hospital of Nanchang University
Qiufen Xun
The Second Affiliated Hospital of Nanchang University
Hongluan Wang
Jiangxi Provincial People's Hospital
Jiarong Liu
The Second Affiliated Hospital of Nanchang University
Ziqing Zhong
The Second Affiliated Hospital of Nanchang University
Chao Ouyang
The Second Affiliated Hospital of Nanchang University
Qing Yang (lisaqing2004@126.com)
The Second Affiliated Hospital of Nanchang University

Research Article

Keywords: metagenomic next-generation sequencing (mNGS), bronchoalveolar lavage fluid (BALF), pulmonary infection, comorbidities, diagnostic value

Posted Date: February 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1240151/v1

Abstract

Background: Timely identification of causative pathogens is always an important link for the diagnosis and treat for pulmonary infection. As a novel approach to detect the pathogens, metagenomic next-generation sequencing (mNGS) can directly sequence the nucleic acid of specimens, providing a wide range of microbial profiles. The purpose of this study is to evaluate the diagnostic performance of mNGS of bronchoalveolar lavage fluid (BALF) in patients suspected with pulmonary infection.

Methods: From April 2019 to July 2021, 502 patients with suspected pneumonia undergoing both mNGS of BALF and conventional microbiological tests (CMTs) were classified into different groups based on the comorbidities. The diagnostic performance was compared between mNGS and CMTs. Clinical comprehensive analysis was regarded as reference standard.

Result: The diagnostic accuracy and sensitivity of mNGS were 74.9% (95% confidence interval [CI], 71.7-78.7%) and 72.5% (95% CI, 68.2-76.8%) respectively, outperformed those of CMTs (36.2% for diagnostic accuracy, 25.4% for sensitivity). In most of the pathogens, the detection rate of mNGS was higher than CMTs. Polymicrobial infections most often occurred in immunocompromised group (22.1%), followed by other comorbidities group (22.1% vs 13.3%, p= 0.13) and bronchiectasis group (22.1% vs 9.7%, p= 0.018). Only 2.3% (95%CI, 0.3%-4.4%) patients developed polymicrobial infection in simple pulmonary infection group. Besides, the spectrums of pathogens also varied in different groups. Importantly, the positive predictive values (PPVs) of mNGS were observed discrepant in different pathogens: 94.9% (95%CI 89.1-100%) for *Mycobacterium tuberculosis*, 86.2% (95%CI,72.9-99.6%) for *Chlamydia psittaci*, 86.0% (95%CI, 76.0-96.0%) for *Aspergillus*, and 67.6% (95%CI, 51.1-84.2%) for *Non-mycobacterium tuberculosis*, 67.3% (95%,54.1-80.5%) for *Pneumocystis jeroveci*; as for bacteria, the PPVs also show differences in different types of bacteria.

Conclusion: mNGS of BALF can highly enhance the accuracy and detection rate of pathogens in patients with pulmonary infection. Besides, the comorbidities and the types of pathogens should be taken into consideration when interpreting the report of mNGS.

Introduction

Pulmonary infection remains one of the leading causes of morbidity and mortality in patients worldwide[1]. Accurate antimicrobial treatment can improve the cure rate, reduce broad-spectrum antibiotic usage, and decrease medical costs. Therefore, timely identification of causative pathogens is critical for improving clinical prognosis[2]. Conventional culture has been routinely used to detect the pathogens of pulmonary infection. However, this method is time-consuming and exhibit a low detection rate. Culture-independent techniques (polymerase chain reaction, PCR and serological testing) have been proven useful for broadening the scope of detectable pathogens and increasing the detection rate, but it need clinician to presume the species of pathogens[3, 4].Hence, a rapid and unbiased diagnostic approach is urgently needed in clinic.

Metagenomic next-generation sequencing (mNGS), a nucleic acid sequencing technique, is reported to detect nearly all pathogens in theory[5, 6]. To date, mNGS has been increasingly applied in pulmonary infection and showed better diagnostic performance than culture, especially in immunocompromised patients and severe pneumonia[7-9]. Optimal lower respiratory tract specimen is crucial for the microbial diagnosis. Bronchoalveolar lavage fluid (BALF) is an easy-available clinical specimen, which can avoid contamination of oropharyngeal flora compared with sputum and have a higher sensitivity than blood[10, 11]. However, there are still numerous challenges when applying mNGS of BALF in pulmonary infection, because of the presence of commensals microbes in respiratory tract and the complexity of pulmonary infections[8, 12]. Thus, further investigation of mNGS for BALF is yet demanded in pulmonary infection.

In the present study, we evaluate the value of mNGS in the diagnosis of patients suspected with pulmonary infection compared with that of conventional microbiological tests (CMTs). Moreover, we have assessed the spectrum of pathogens in patients with different comorbidities. Besides, the positive predictive values (PPVs) of different pathogens were calculated to evaluate the reliability of positive results of mNGS.

Methods

Study design and population

Adult patients (age > 18 years old) suspected pulmonary infection from The Second Affiliated Hospital of Nanchang University and Jiangxi Provincial People's Hospital between April 2019 and July 2021 were retrospectively reviewed. Age, gender, comorbidities, results of laboratory tests, the image of lung, therapy of antibiotics and patient outcomes were recorded. Patients suspected with pulmonary infection should possess new-onset shadow on chest images and at least one of the following symptoms:1) cough, dyspnea or other respiratory symptoms; 2) fever; 3) leukocytosis or leukocytopenia. Finally, 502 patients were enrolled. All patients underwent bronchoscopy to obtain BALF. Both mNGS of BALF and CMTs were performed to detect pathogens. Based on the comorbidities, 502 patients were classified into four groups:1) immunocompromised group: patients were viewed as the immunosuppressed host; 2) bronchiectasis group: patients were diagnosed with bronchiectasis, destructive pneumonophthisis patients were also classified into this group because of the structural lesion in lung; 3) other comorbidities group: patients with other comorbidities which can increase the risk of pulmonary infections (including diabetes,chronic obstructive pulmonary disease, interstitial lung disease, bronchial asthma, cerebrovascular disease, lung cancer treated with targeted therapy or surgical, liver cirrhosis, apart from the above two groups)[13, 14]; 4) simple pulmonary infection group: patients without previous underlying diseases. Immunosuppressed host should meet any of the following status: 1) long-term therapy of steroids (> 20 mg/d of prednisone- equivalent and the cumulative dose≥600mg); 2) autoimmune disease treat with immunosuppressive agents or cytotoxic drugs; 3) solid-organ transplantation; 4) recent chemotherapy during the last month; 5) hematological malignancy;6)agranulocytosis; 7)human immunodeficiency virus(HIV) infection [15, 16]. For bronchiectasis patients, the computerized tomography of chest should fulfill at least one of these

followings: 1) the inner diameter of the bronchus/the diameter of concomitant pulmonary artery > 1; 2) the bronchi does not become thinner from center to periphery; 3) the bronchioles could be seen in the range of 1cm from the peripheral pleura[17].

Conventional microbiological tests

Conventional microbiological tests (CMTs) included smear and culture of bacterial and fungal, acid-fast stain, PCR, Grocott´s methenamine silver stain and cryptococcus capsular polysaccharide test. Other methods, like Galactomannan antigen and (1,3)- β-D-glucans for fungi, tuberculin skin test and enzyme-linked immune spot for *Mycobacterium tuberculosis*, were not regarded as etiological confirmed methods.

Clinical comprehensive analysis was regarded as the reference standard

Two experienced clinicians in the management of pulmonary infection independently reviewed the medical records of all patients, along with the results of mNGS. As for the discordant cases, another senior clinician was consulted to reach a consensus. Firstly, clinicians judged whether the patients had infectious or non-infectious diseases. Secondly, the causative pathogens were determined by comprehensive analysis based on clinical manifestation, test results, chest radiology and treatment response.

Sample Processing and DNA Extraction

Once acquired, 2ml BALF were placed in a ribozyme free centrifuge tube at -20℃.Then they were sent to BGI-Hua da (Wuhan, China) with dry ice for sequencing and bioinformatics analysis. The remaining specimens were sent to clinical microbiology lab. 1.5mL microcentrifuge tube with 0.6mL BALF and 250μL 0.5mm glass bead were attached to a horizontal platform on a vortex mixer and agitated vigorously at 2800-3200 rpm for 30 min. Then 7.2μL lysozyme was added for wall-breaking reaction. DNA was extracted using the TIANamp Micro DNA Kit (DP316, TIANGEN BIOTECH) according to the manufacturer's recommendation.

Sequencing and Bioinformatic analysis

DNA sequencing libraries were constructed through DNA-fragmentation，end-repair, adapter-ligation and PCR amplification. RNA sequencing libraries were constructed for patients with suspected viral infection. Agilent 2100 Bioanalyzer was used to assess the quality of the libraries. After qualified libraries were collected, DNA Nanoball (DNB) was made and sequenced by MGISEQ-2000 platform[18].High quality sequencing data were generated by removing low quality, adapter contamination and short (length < 35 base pairs [bp]) reads. The total number of sequencing reads was 20 million sequences per library at least. Then, human host sequences mapped to the human reference genome (hg19) were excluded by computational subtraction using Burrows-Wheeler alignment[19].The remaining reads were classified into four microbial genome databases (bacteria, fungi, viruses and parasites) by simultaneously aligning to Pathogens metagenomics Database (PMDB). Classification reference databases were acquired from

NCBI (ncbi.nlm.nih.gov/genomes/). RefSeq contains the whole genome sequence of 4,945virus, 6,350 bacteria, 1064 fungi related to human infection, and 234 parasites associated with human diseases.

Criteria for a Positive mNGS Result

Given the lack of standard to interpret the reports of mNGS and the variety of different sequencing platforms, the following criteria was applied to define positive result. 1) possessing pulmonary pathogenicity reported in literature; 2) bacteria: >30% relative abundance at the genus level for opportunistic pathogenic bacteria; ≥3 stringently mapped reads at species level for high pathogenicity bacteria; ≥1 stringently mapped read for *Mycobacterium tuberculosis complex* and *Non-tuberculous mycobacteria*[20, 21]; 3) fungi (candida excluded): ≥1 stringently mapped read at species level or ≥10 stringently mapped reads for mold at species level[16]; 4) mycoplasma, chlamydia: ≥1 stringently mapped read at species level; 5) Oral commensals, candida, virus: evaluated by physician based on clinical manifestations and examination results[16].

Statistical analysis

Continuous variables are reported as the median and interquartile range. Categorical variables are expressed as frequencies and percentages. 95% confidence intervals of proportions were calculated by Wilson's method. The McNemar test was used for comparisons of the detection rate of CMTs and mNGS. Chi-square were used to evaluate the statistically difference of non-matched samples.All statistical analyses were performed by SPSS25. All tests were two-tailed, and $p < 0.05$ was considered statistically significant.

Results

Patient characteristics

Baseline characteristics of 502 patients recruited in this study were presented in Table 1. 297 were males and 205 were females. The median age was 58.0 years. 104(20.7%) patients were classified into immunocompromised group. 93(18.5%) patients were classified into bronchiectasis group. 90 patients were classified into othercomorbidities group. 215(42.8%) patients were classified into simple pulmonary infection group. 306(61%) patients had already received therapy of antibiotics before hospitalization (the most common antibiotics were β-lactam and quinolones).

Table 1. Clinical characteristics of 502 patients

Characteristics	Value
Age, years, median (Q1, Q3)	58.0（47.0, 68.0)
Gender, male, n (%)	297 (59.2%)
Antibiotic therapy before hospitalization	306 (61.0%)
Comorbidities	
Immunocompromised group, n (%)	104 (20.7%)
Autoimmune disease	29
Long-term therapy of steroids	25
Recent chemotherapy	21
Solid-organ transplantation	18
Hematological malignancy	7
Agranulocytosis	2
HIV	1
Bronchiectasis group, n (%)	93 (18.5%)
Bronchiectasis patients	91
Destructive pneumonophthisis	2
Other comorbidities group, n (%)	90 (17.9%)
Diabetes	37
Chronic obstructive pulmonary disease	30
Interstitial lung disease	6
Bronchial asthma	4
Cerebrovascular disease	6
Targeted therapy and surgical of lung cancer	4
Liver cirrhosis	3
Simple pulmonary infection group, n (%)	215 (42.8%)

Infection Types

According to the comprehensive analysis of medical records and the results of mNGS, 84.1% (422/502) patients were diagnosed with pulmonary infection, 15.9% (80/502) patients were considered non-

infectious diseases, including malignancies, organizing pneumonia, vasculitis, etc. Among the infectious patients, 355 patients were detected the causative pathogens (including 49 patients with polymicrobial infection, 306 patients with monomicrobial infection), 67 patients were not confirmed the certain pathogens. It was worth noting that polymicrobial infections most often occurred in immunocompromised group, followed by other comorbidities group (22.1% vs 13.3%, p= 0.13) and bronchiectasis group (22.1% vs 9.7%, p= 0.018). However, only 2.3% (95%CI, 0.3%-4.4%) patients were observed polymicrobial infection in simple pulmonary infection group, which is obviously less than other three groups (Table 2).

Table 2. Infection types in different comorbidities

	Pulmonary infection	non-infection	polymicrobial infection	monomicrobial infections	not confirmed pathogen
Total	422(84.1%)	80(15.9%)	49(9.7%)	306(61.0%)	67(13.3%)
Immunocompromised group	98(94.2%)	6(5.8%)	23(22.1%)	69(66.3%)	6(5.8%)
Bronchiectasis group	85(91.4%)	8 (8.6%)	9(9.7%)	60(64.5%)	16(17.2%)
Other comorbidities group	71(78.8%)	19(21.1%)	12(13.3%)	45(50.0%)	14(15.6%)
Simple pulmonary infection group	168(78.1%)	47(21.9%)	5(2.3%)	132(61.4%)	31(14.4%)

Pathogen Spectrum in Different Groups

The spectrum of pathogens varied in patients with different comorbidities. In immunocompromised group, the most common pathogens were *Pneumocystis yersini* (31.0%) and *Aspergillus* (17.2%). In bronchiectasis group, the most common pathogens were *Pseudomonas aeruginosa* (24.1%) and *Non-mycobacterium tuberculosis* (24.1%). In other comorbidities group, the most common pathogens were *Mycobacterium tuberculosis* (20.3%), *Aspergillus* (14.5%) and *Anaerobion* (11.6%). In simple pulmonary infection group, the most common pathogens were *Mycobacterium tuberculosis* (37.3%) and *Chlamydia psittaci* (12.7%). (Figure 1)

Comparison of mNGS and CMTs

In the present study, 306 infectious patients obtained accurate microbiological diagnosis by mNGS. Whereas, only 107 infectious patients were detected the causative pathogens by CMTs. In 37 infectious patients with positive results of mNGS, the detection results were inconsistent with causative pathogens determined by comprehensive analysis of medical records (Figure 2). Diagnostic accuracy was the proportion of correct diagnosed patients (both infectious and non-infectious patients) among the total patients. Sensitivity was the proportion of infectious patients with correct diagnosis among the total

infectious patients. The diagnostic accuracy of mNGS was much higher than that of CMTs (74.9% vs 36.2%, p<0.001). And the sensitivity of mNGS was also higher than that of CMTs (72.5% vs 25.4%, p< 0.001). In addition, 70 of 149 patients with negative reports of mNGS were diagnosed with non-infectious diseases, the negative predictive value (NPV) of mNGS was 47% (95%CI, 38.9%-55.1%) (Table 3).

Table 3. The diagnostic performance between mNGS and CMTs

	Diagnostic accuracy	**Sensitivity**	**NPV**
	(95% CI)	**(95% CI)**	**(95%CI)**
mNGS	74.9(71.7-78.7)	72.5(68.2-76.8)	47.0(38.9-55.1)
CMTs	36.2(32.0-40.5)	25.4(21.2-29.5)	21.1(16.9-25.3)

Among the 355 patients with microbiologically confirmed infection, 406 pathogens were detected (part of patients were diagnosed with polymicrobial infection). 120 pathogens were detected by mNGS and CMTs simultaneously, 239 pathogens were detected by mNGS separately, 12 pathogens were detected by CMTs separately, and 35 pathogens were confirmed by other methods. The pathogens detection rate of mNGS was much higher than that of CMTs (p< 0.001). As is shown in figure 3, in most of pathogens, mNGS also showed better diagnostic performance than CMTs, especially in *Mycoplasma, Chlamydia* and *Viruses* which were exclusively detected by mNGS. As for *Acinetobacter baumanii*, *Nocardia*, *Staphylococcus aureus*, *Escherichia coli* and *Legionella*, there was no statistical difference between the two methods because of the small sample size. But these pathogens also showed the trend that mNGS was superior than CMTs.

The PPV of pathogens in mNGS reports

To evaluate the reliability of positive mNGS reports, we calculated the PPV of mNGS for different pathogens. Bacteria were classified into different groups according to their pathogenicity and the possibility of colonizing in respire tract. *Haemophilus influenzae*, *Streptococcus pneumoniae* and *Moraxella catarrhalis* are the most common pathogens of community acquired pneumonia, which can also colonize in nasopharynx[22]. The summarized PPV of them was 71.9% (95%CI, 60.6-83.2%). *Nocardia, Staphylococcus aureus, Escherichia coli* and *Legionella* were high pathogenicity in respiratory tract. The summarized PPV of them was 82.8%(95%CI,68.1-97.4%). Unlike previous bacteria, *Pseudomonas aeruginosa, Acinetobacter baumanii, Klebsiella pneumoniae* and *Stenotrophomonas maltophilia* can colonize in lower respiratory tract. The summarized PPV of them was 97.7% (95%CI, 93.7-100%). For fungi, the PPVs of *Aspergillus* and *Pneumocystis jeroveci* were 86.0% and 67.3%, respectively. In mycobacteria, the PPVs of *Mycobacterium tuberculosis* and *Non-mycobacterium tuberculosis* were 94.9% and 67.6%, respectively (Figure 4A). It was worth noting that immunocompromised group had a higher PPV of *Pneumocystis jeroveci* than that of other three groups (80.5% vs 12.8% p=0.006) (Figure 4B). Similarly, the PPV of *Non-mycobacterium tuberculosis* in bronchiectasis group was higher than that of other three groups (90.5% vs 30.8% p=0.001) (Figure 4C).

Discussion

In the present study, 502 patients suspected with pulmonary infection were retrospectively analyzed. 422 (84.1%) patients were diagnosed with pulmonary infection and 80 (15.9%) patients were considered non-infectious diseases according to the comprehensive analysis of medical records. Based on the comorbidities, these patients were classified into four groups. The spectrum of pathogens and proportion of polymicrobial infections varied in different groups. Besides, we have systematically compared the diagnostic performance between mNGS of BALF and CMTs, the former was significantly superior to the latter. In addition, it can also be demonstrated that the PPVs of mNGS varied in different types of pathogens, which should be taken attention when interpreting the result of mNGS.

The diagnostic accuracy and sensitivity of mNGS were 74.9% (95%CI, 71.7-78.7%) and 72.5% (95%CI, 68.2-76.8%), respectively, which were much higher than those of CMTs. In a study of 235 patients suspected pneumonia, the sensitivity of BALF mNGS could reach up to 73.33%[21]. In another study with 132 patients, compared to conventional testing, mNGS suggested potentially missed diagnoses in 22 patients involving 48 additional pathogenic microorganisms[23]. This study also demonstrated that the detection rate of mNGS was higher than that of CMTs in most of the pathogens. In a retrospective study of 72 patients, the detection rate of bacteria by mNGS was also higher than conventional methods. As for fungi, the detection rate of mNGS is reported comparable to conventional tests because of the introduction of Galactomannan antigen and (1,3)-β-D-glucans[24]. But these two methods were not regarded as etiological confirmed tests because they cannot indicate the specific species of pathogens. When the two methods were excluded, the detection of mNGS is higher than conventional tests for fungi[25].

Unlike previous study[16, 21], we did not observe a high value of NPV (47.0%, 95%CI [38.9-55.1%]). For example, in a retrospective study, the NPV of different pathogens varied from 73.5% to 100%[16]. Another study demonstrated that the NPV in mNGS of BALF was 85.88%(95%CI,76.25–92.18%)[21]. Here are some explanations for the conflicting results. The reports possessing lot reads of background microbes of respiratory tract (like *Prevotella, Veillonella, Neisseria* etc.) were not regarded as positive results in present study. This interpretation can increase the number of negative results which result in the decreasing of NPV. Besides, rare pathogens (like *Kingella, Tropheryma whipplei*) and virus (like *Human herpes virus*, *Torque teno virus*) were usually not regard as causative pathogens unless they were considered significant by the managing clinicians. In addition, *Mycobacterium tuberculosis* were the most common pathogens in this study. It requires significant cell wall disruption to release nucleic acid[26]. The low biomass in DNA extraction can also decrease the NPV. To the best of our knowledge, this research was the largest sample size investigation to evaluate the diagnostic performance of mNGS for BALF in patients with pneumonia.

Immunocompromised patients had a more complex microbial etiology, with higher detection rate of *Pneumocystis yersini* and *Aspergillus*. *Pseudomonas aeruginosa* and *Non-mycobacterium tuberculosis* were the dominant pathogens in bronchiectasis patient. *Pseudomonas aeruginosa* is the

most common pathogens isolated from respiratory tract specimen in patients with bronchiectasis. And *Non-mycobacterium tuberculosis* has been increasingly detected globally[27]. In this study, 40.0% (14/35) patients with positive acid-fast stain were diagnosed with *Non-mycobacterium tuberculosis* infection by mNGS. Since not all patients with positive acid-fast stain were *mycobacterium tuberculosis* infection, and the culture of *mycobacterium* takes one to three months, mNGS might be an efficient method to distinguish *Non-mycobacterium tuberculosis* infection[28]. As for patients without the above two comorbidities, *mycobacterium tuberculosis* was the most common pathogen. The high detection of this bacteria is reasonable. First, there was a high prevalence of *mycobacterium tuberculosis* infection in China[29]. Besides, most patients enrolled in our study have received empirical anti-infective treatment. Interestingly, a significant number of patients in simple pulmonary infection group were diagnosed with *Chlamydia psittaci* infection, which may be associated with universal poultry production in Jiangxi, China[30].

Another strength of our study is that we have strived to evaluate the PPVs of mNGS regarding to different types of pathogens, which can help clinician to interpret the reports of mNGS. The summarized PPV of *Haemophilus influenzae*, *Streptococcus pneumoniae* and *Moraxella catarrhalis* was 71.9% (95%CI, 60.6-83.2%). As the most common colonization bacteria in nasopharynx[22], these pathogens could be carried into BALF while bronchoscopy. Thus, attention should be paid to distinguishing between colonization and infection in the positive results of these pathogens. The summarized PPV of *Pseudomonas aeruginosa, Acinetobacter baumanii, klebsiella pneumoniae* and *Stenotrophomonas maltophilia* is 97% (95%CI, 93.7-100%). This high value is majorly due to a stricter positive criterion in reports of mNGS: these bacteria were not considered as positive result of mNGS if their relative abundance at the genus level less than 30%. Similar high PPV of these pathogens could be observed in Peng et al's study[16]. It should be noted that the PPV of *Pneumocystis jeroveci* in immunocompromised group was much higher than that in other three groups (80.5% to 18.2%, p=0.006). The positive result of *Pneumocystis jeroveci* in immunocompromised patients is more likely to be causative pathogen. On the contrary, the positive result of *Pneumocystis jeroveci* in immunocompetent patients is more likely to be considered environmental contaminants. Similar feature can be observed in *Non-mycobacterium tuberculosis*. Therefore, the commodities and types of pathogens should be taken into consideration while interpreting the reports of mNGS.

Although the potential of pathogen detection of mNGS has been confirmed in many studies and clinical contexts. There were still limitations for the application of mNGS in clinic. Multiple evidences suggested that the respiratory tract is not a sterile environment. Commensal microbes commonly exist in the respiratory tract of healthy individuals, which means microbes with lots of reads in the reports of mNGS are not always the causative pathogen[8]. Besides, the existence of environmental contaminants, opportunistic pathogen and mismatch of DNA fragment also make challenges for us to interpret the reports of mNGS. Therefore, making a reasonable and accurate interpretation is also an important bottleneck[9]. Cost is another concern for the wide use of mNGS. In China, the current cost is around $600 per sample, which is much higher than that of conventional tests [24, 25]. Currently, the vast proportion of the reads (>90%) sequenced by mNGS are host-derived, which means most of money is

spent on these invalid sequences. Reducing the host reads prior to sequencing by host depletion methods can effectively diminish the total cost [31]. Increasing the number of test samples per run can also cut the cost per sample, but it comes at the expense of the turn-around time [32].Besides, encouraging more high-quality sequencing platforms to participate in market can indirectly promote cost reduction.

There were several limitations in this study. Firstly, we failed to conduct the culture of *Mycobacterium tuberculosis* and *Non-mycobacterium tuberculosis*. Secondly, the patients classified into other comorbidities group had different comorbidities (e.g., diabetes, chronic obstructive pulmonary disease, interstitial lung disease, cerebrovascular disease, lung cancer treated with targeted therapy or surgical, liver cirrhosis), which may be associated with heterogeneity of the spectrum of pathogens. Thirdly, most patients enrolled in present study had received anti-infective therapy before hospitalization, which might result in the underestimation of sensitivity of culture and lead to a higher detection rate of atypical pathogens. At last, the interpretation of the results of mNGS depends on subjective judgment of clinician to a certain extent, which may lead to bias.

Overall, the application of mNGS for BALF can improve the detection of pathogens in patients with pulmonary infection. But the interpretation of reports and the cost of mNGS are still concerns. This study demonstrated that the commodities and the types of pathogens should be taken into consideration when interpreting the reports of mNGS. More investigations are still needed for the extension of mNGS.

Abbreviations

mNGS: Metagenomic next-generation sequencing;

BALF: Bronchoalveolar lavage fluid;

CMTs: conventional microbiological tests

CI: confidence interval

PPV: positive predictive value

NPV: negative predictive value

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by Ethics Committee of Second Affiliated Hospital of Nanchang University (No. 2019-013) and Jiangxi Provincial People's Hospital (No.2021-076). All procedures performed was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Written informed consent was waived due to the retrospective nature of the study.

Consent for publication

All the authors grant consent for publication.

Availability for data and materials

The original data of this study are available from China National GeneBank DataBase(https://db.cngb.org/search/project/CNP0002658/)

Competing interests

The authors have no potential conflicts of interest.

Funding

This work was supported by National Natural Science foundation of China (81860011)

Authors' contributions

Dengfeng Wu, Wei Wang analyzed data, drafted the first version of manuscript and submitted to the publication; Qiufen Xun, Hongluan Wang collected data and helped to analyze data; Jiarong Liu, Ziqing Zhong, Chao Ouyang helped to collect data; Qing Yang was responsible for the entire project, designed the experiment and revised the draft of the manuscript.

Acknowledgments

We would like to acknowledge the patients for cooperating with our investigation.

References

1. Hay S., Collaborators GCoD **Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016**. *Lancet (London, England)* 2017, **390**(10100):1151-1210.
2. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, Cooley LA, Dean NC, Fine MJ, Flanders SA *et al*: **Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America**. *American journal of respiratory and critical care medicine* 2019, **200**(7):e45-e67.
3. Gadsby NJ, Russell CD, McHugh MP, Mark H, Conway Morris A, Laurenson IF, Hill AT, Templeton KE: **Comprehensive Molecular Testing for Respiratory Pathogens in Community-Acquired Pneumonia**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2016, **62**(7):817-823.
4. Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, Reed C, Grijalva CG, Anderson EJ, Courtney DM *et al*: **Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults**.

The New England journal of medicine 2015, **373**(5):415-427.

5. Gu W, Miller S, Chiu CY: **Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection**. *Annual review of pathology* 2019, **14**:319-338.
6. Chiu CY, Miller SA: **Clinical metagenomics**. *Nature reviews Genetics* 2019, **20**(6):341-355.
7. Xie Y, Du J, Jin W, Teng X, Cheng R, Huang P, Xie H, Zhou Z, Tian R, Wang R *et al*: **Next generation sequencing for diagnosis of severe pneumonia: China, 2010-2018**. *The Journal of infection* 2019, **78**(2):158-169.
8. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, Curtis JL: **Bacterial Topography of the Healthy Human Lower Respiratory Tract**. *mBio* 2017, **8**(1).
9. Han D, Li Z, Li R, Tan P, Zhang R, Li J: **mNGS in clinical microbiology laboratories: on the road to maturity**. *Critical reviews in microbiology* 2019, **45**(5-6):668-685.
10. Dubourg G, Abat C, Rolain JM, Raoult D: **Correlation between sputum and bronchoalveolar lavage fluid cultures**. *Journal of clinical microbiology* 2015, **53**(3):994-996.
11. Chen X, Ding S, Lei C, Qin J, Guo T, Yang D, Yang M, Qing J, He W, Song M *et al*: **Blood and Bronchoalveolar Lavage Fluid Metagenomic Next-Generation Sequencing in Pneumonia**. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale* 2020, **2020**:6839103.
12. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB: **The Microbiome and the Respiratory Tract**. *Annual review of physiology* 2016, **78**:481-504.
13. Wu HP, Chu CM, Lin CY, Yu CC, Hua CC, Yu TJ, Liu YC: **Liver Cirrhosis and Diabetes Mellitus Are Risk Factors for Staphylococcus aureus Infection in Patients with Healthcare-Associated or Hospital-Acquired Pneumonia**. *Pulmonary medicine* 2016, **2016**:4706150.
14. Martino R, Foley N, Bhogal S, Diamant N, Speechley M, Teasell R: **Dysphagia after stroke: incidence, diagnosis, and pulmonary complications**. *Stroke* 2005, **36**(12):2756-2763.
15. Ramirez JA, Musher DM, Evans SE, Dela Cruz C, Crothers KA, Hage CA, Aliberti S, Anzueto A, Arancibia F, Arnold F *et al*: **Treatment of Community-Acquired Pneumonia in Immunocompromised Adults: A Consensus Statement Regarding Initial Strategies**. *Chest* 2020, **158**(5):1896-1911.
16. Peng JM, Du B, Qin HY, Wang Q, Shi Y: **Metagenomic next-generation sequencing for the diagnosis of suspected pneumonia in immunocompromised patients**. *The Journal of infection* 2021, **82**(4):22-27.
17. Bonavita J, Naidich DP: **Imaging of bronchiectasis**. *Clinics in chest medicine* 2012, **33**(2):233-248.
18. Korostin D, Kulemin N, Naumov V, Belova V, Kwon D, Gorbachev A: **Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing**. *PloS one* 2020, **15**(3):e0230301.
19. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics (Oxford, England)* 2009, **25**(14):1754-1760.
20. Langelier C, Zinter MS, Kalantar K, Yanik GA, Christenson S, O'Donovan B, White C, Wilson M, Sapru A, Dvorak CC *et al*: **Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic**

Cellular Transplant Patients. *American journal of respiratory and critical care medicine* 2018, **197**(4):524-528.

21. Chen Y, Feng W, Ye K, Guo L, Xia H, Guan Y, Chai L, Shi W, Zhai C, Wang J *et al*: **Application of Metagenomic Next-Generation Sequencing in the Diagnosis of Pulmonary Infectious Pathogens From Bronchoalveolar Lavage Samples**. *Frontiers in cellular and infection microbiology* 2021, **11**:541092.
22. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt BJ, Hales BJ, Walker ML, Hollams E *et al*: **The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development**. *Cell host & microbe* 2015, **17**(5):704-715.
23. Zhan Y, Xu T, He F, Guan WJ, Li Z, Li S, Xie M, Li X, Chen R, Cheng L *et al*: **Clinical Evaluation of a Metagenomics-Based Assay for Pneumonia Management**. *Frontiers in microbiology* 2021, **12**:751073.
24. Fang X, Mei Q, Fan X, Zhu C, Yang T, Zhang L, Geng S, Pan A: **Diagnostic Value of Metagenomic Next-Generation Sequencing for the Detection of Pathogens in Bronchoalveolar Lavage Fluid in Ventilator-Associated Pneumonia Patients**. *Frontiers in microbiology* 2020, **11**:599756.
25. Miao Q, Ma Y, Wang Q, Pan J, Zhang Y, Jin W, Yao Y, Su Y, Huang Y, Wang M *et al*: **Microbiological Diagnostic Performance of Metagenomic Next-generation Sequencing When Applied to Clinical Practice**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018, **67**(suppl_2):S231-s240.
26. Simner PJ, Miller S, Carroll KC: **Understanding the Promises and Hurdles of Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Infectious Diseases**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018, **66**(5):778-788.
27. To K, Cao R, Yegiazaryan A, Owens J, Venketaraman V: **General Overview of Nontuberculous Mycobacteria Opportunistic Pathogens: Mycobacterium avium and Mycobacterium abscessus**. *Journal of clinical medicine* 2020, **9**(8).
28. Shi CL, Han P, Tang PJ, Chen MM, Ye ZJ, Wu MY, Shen J, Wu HY, Tan ZQ, Yu X *et al*: **Clinical metagenomic sequencing for diagnosis of pulmonary tuberculosis**. *The Journal of infection* 2020, **81**(4):567-574.
29. WHO.**Global tuberculosis report 2021**. https://www.who.int/publications/i/item/9789240037021 Accessed 14 Oct 2021
30. Li H, Tan M, Zhang F, Ji H, Zeng Y, Yang Q, Tan J, Huang J, Su Q, Huang Y *et al*: **Diversity of Avian leukosis virus subgroup J in local chickens, Jiangxi, China**. *Scientific reports* 2021, **11**(1):4797.
31. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K: **Improving saliva shotgun metagenomics by chemical host DNA depletion**. *Microbiome* 2018, **6**(1):42.
32. Rutanga JP, Van Puyvelde S, Heroes AS, Muvunyi CM, Jacobs J, Deborggraeve S: **16S metagenomics for diagnosis of bloodstream infections: opportunities and pitfalls**. *Expert review of molecular diagnostics* 2018, **18**(8):749-759.

Figures

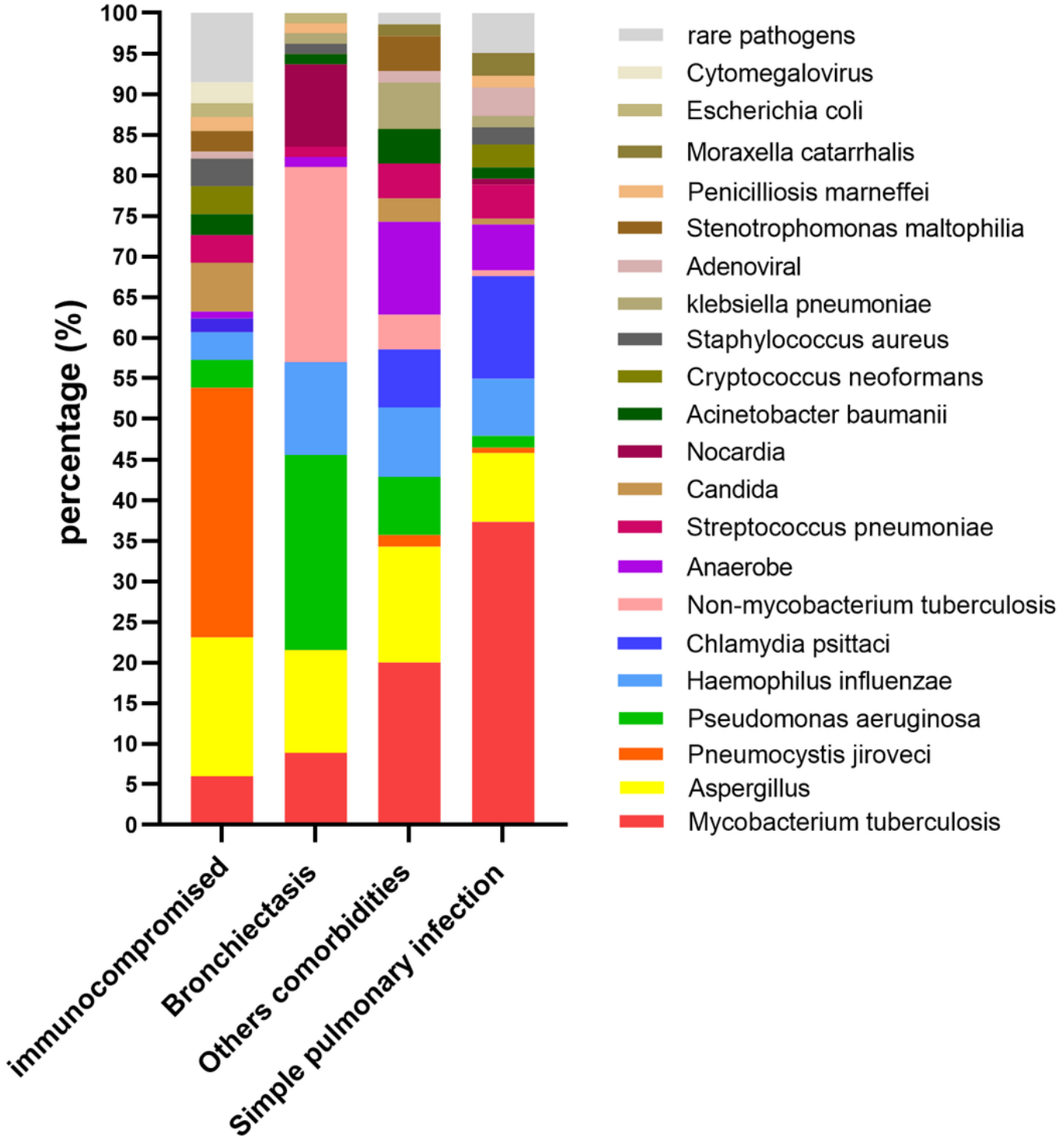

Figure 1

The percentage of different pathogens(y-axis) in patients with different comorbidities(x-axis).

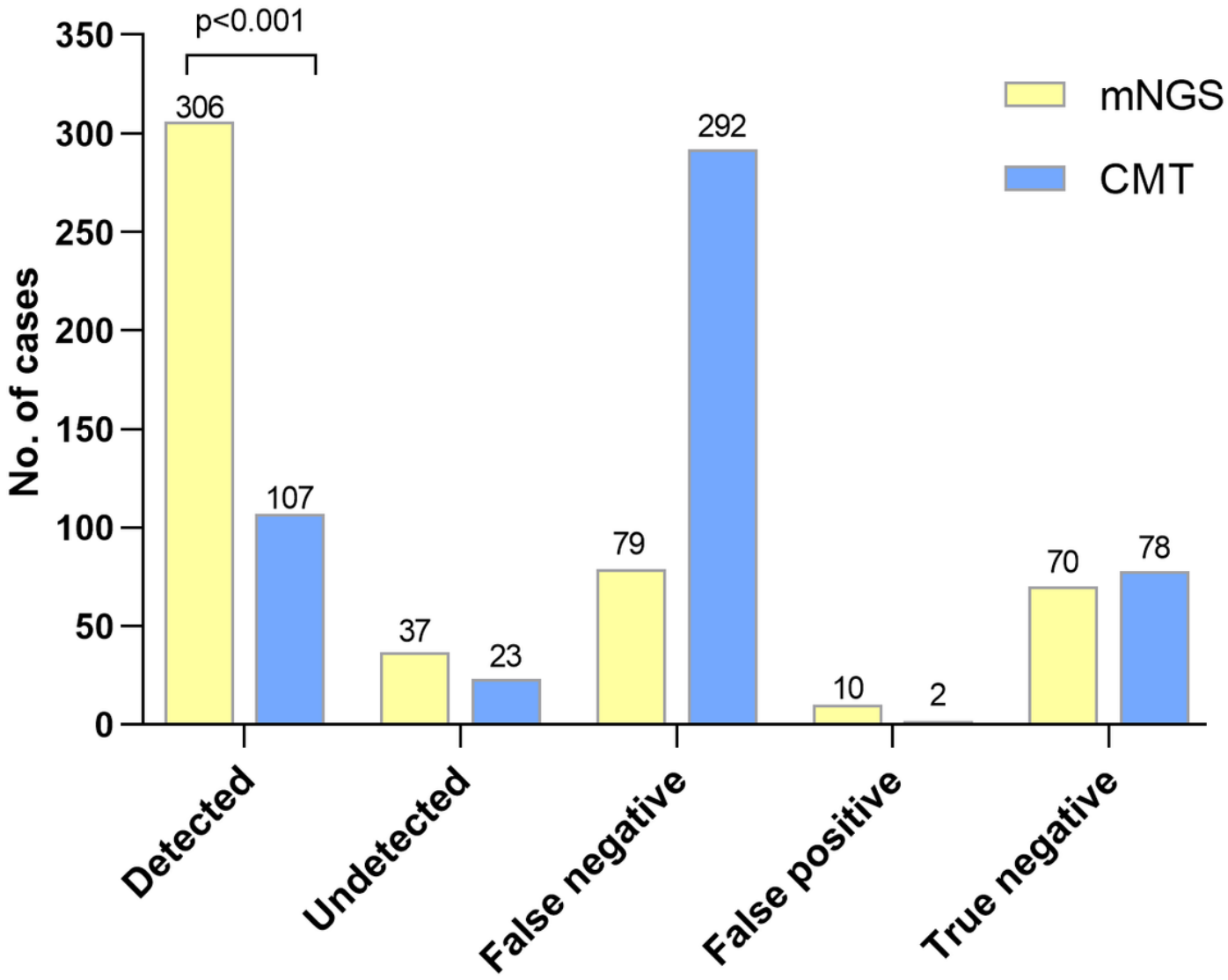

Figure 2

The diagnostic performance between mNGS and CMTs. Detection group:

the causative pathogens were detected by mNGS or CMTs in infectious patients. Misdetection group: positive results by mNGS or CMTs were not corresponded with the reference standard in infectious patients. False negative: negative results but in non-infectious patients. False positive: positive results but in non-infectious patients. True negative: negative result by mNGS or CMTs in non-infectious patients. Patients were diagnosed with infectious or non-infectious diseases and the causative pathogens of infectious patients were determined according to comprehensive analysis of medical records. Causative pathogens detected by mNGS were more than by CMTs ($p<0.001$).

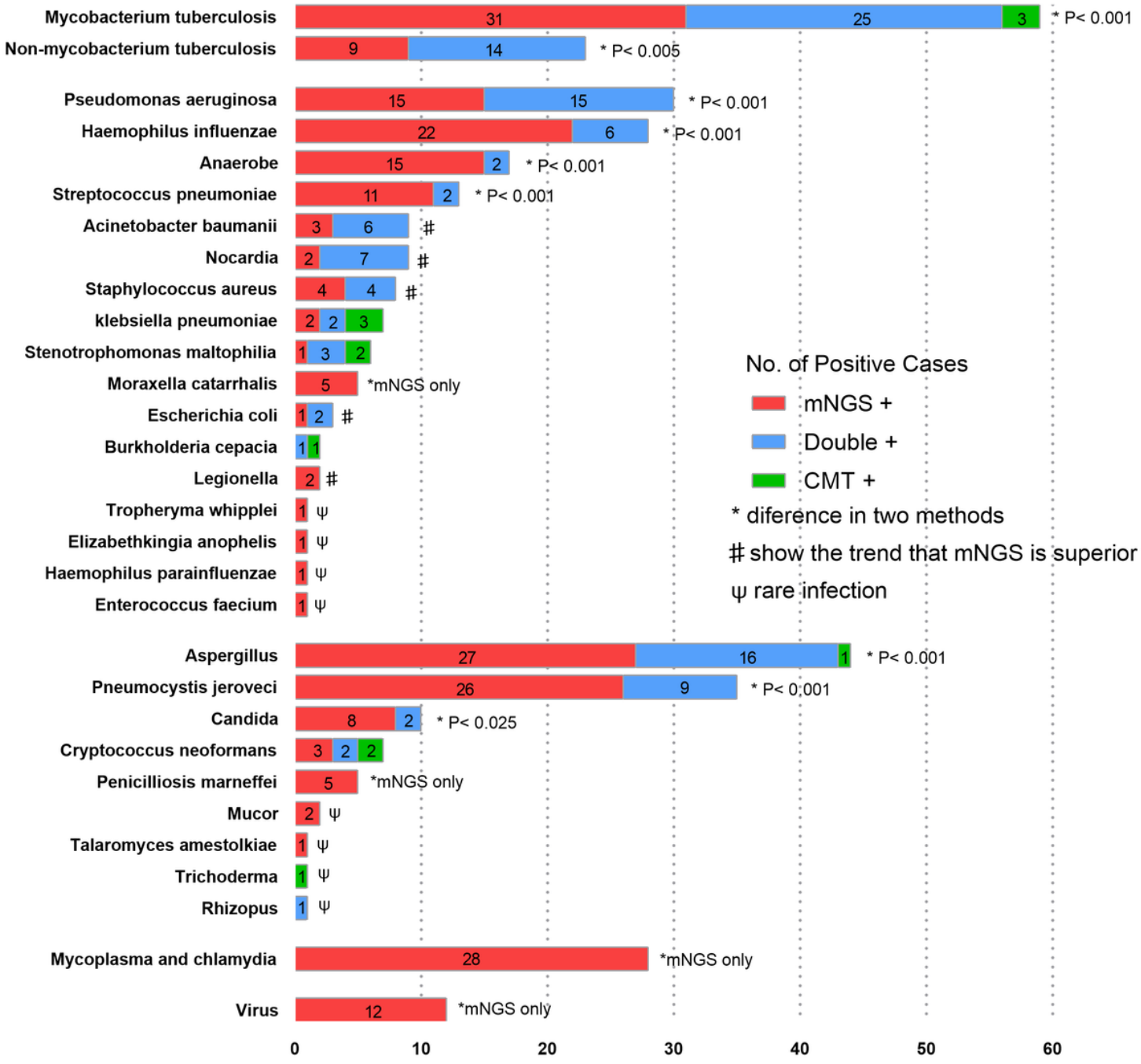

Figure 3

The overlap of positivity between mNGS and CMTs for different pathogens. Numbers in the histogram represents the corresponding cases of pathogens. In *Mycobacterium tuberculosis*, *Non-mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Anaerobe*, *Streptococcus pneumoniae*, *Aspergillus* and *Pneumocystis yersini*, the detection rates of mNGS were higher than those of CMTs. As for *Acinetobacter baumanii*, *Nocardia*, *Staphylococcus aureus* and *Escherichia coli*, there were no statistical differences, but mNGS tended to outperform CMTs.

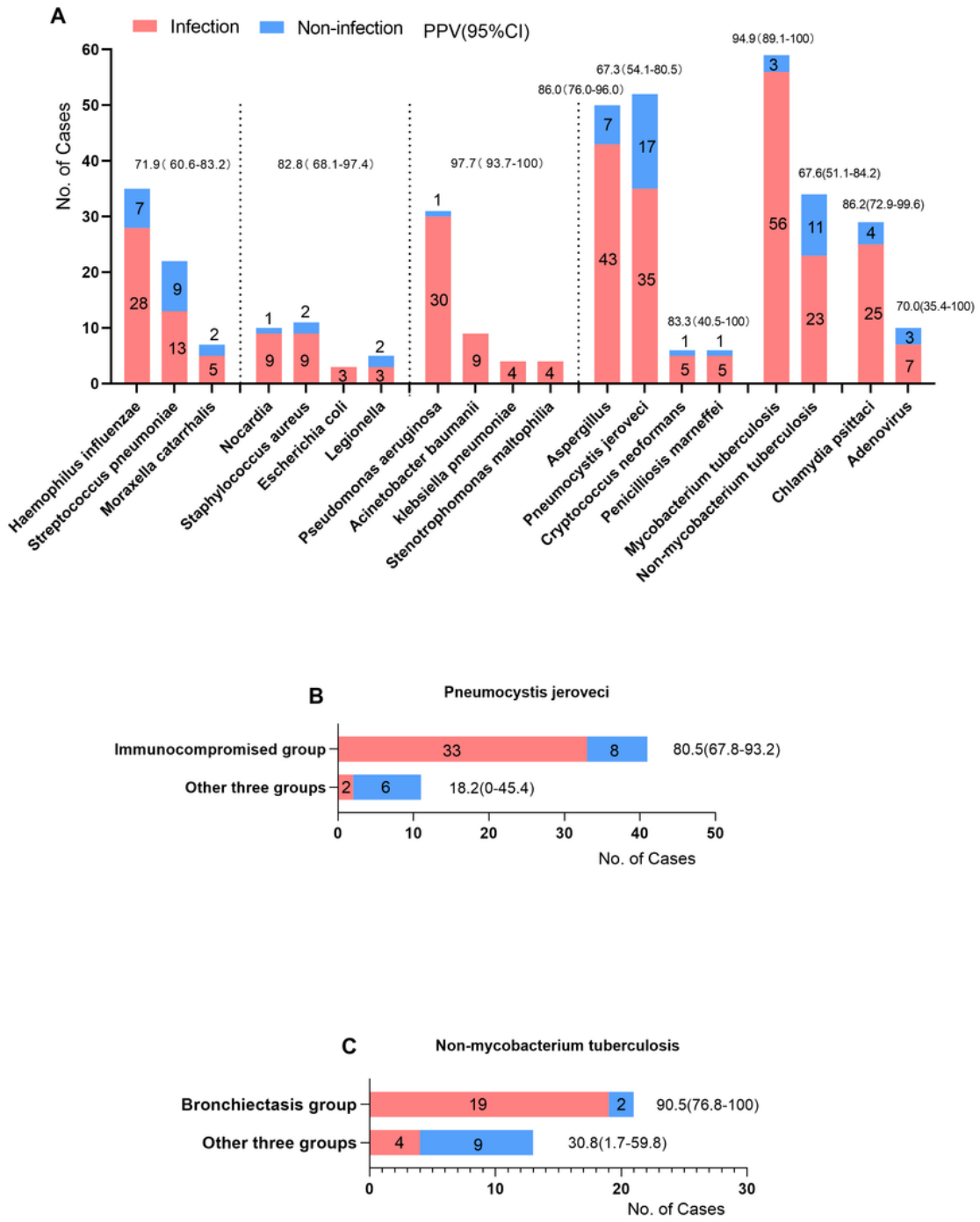

Figure 4

(A) The cases of positive results of mNGS (y-axis) in different pathogens (x-axis). Some were the causative pathogens (infection cases), others were considered environmental contaminants or colonized microbes (non-infection cases). The PPV was the proportion of infection cases among the cases of positive results. The PPVs and 95%CI were listed on the top of the bars. (B) The PPV of *Pneumocystis*

jeroveci in immunocompromised group compared with other three groups. (C) The PPV of *Non-mycobacterium tuberculosis* in bronchiectasis group compared with other three groups.