Check for updates

Cellular analysis and metagenomic next-generation sequencing of bronchoalveolar lavage fluid in the distinction between pulmonary non-infectious and infectious disease

Yilin Pan[1†], Xue Zhang[1†], Yi Sun[2†], Yingying Zhang[1], Wuping Bao[1], Dongning Yin[1], Pengyu Zhang[1,3*] and Min Zhang[1*]

[1]Department of Pulmonary and Critical Care Medicine, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Laboratory Medicine, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3]Department of Infectious Disease, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: The aim of the current study was to investigate the clinical value of cellular analysis and metagenomic next-generation sequencing (mNGS) of bronchoalveolar lavage fluid (BALF) in differentiating pulmonary non-infectious and infectious diseases in immunocompetent patients.

Methods: The present retrospective study was conducted from December 2017 to March 2020, and included immunocompetent patients with suspected pulmonary infection. High-resolution computed tomography, total cell counts and classification of BALF, conventional microbiological tests (CMTs), laboratory tests and mNGS of BALF were performed. Patients were assigned to pulmonary non-infectious disease (PNID) and pulmonary infectious disease (PID) groups based on final diagnoses. PNID-predictive values were analyzed *via* areas under receiver operating characteristic curves (AUCs). Optimal cutoffs were determined by maximizing the sum of sensitivity and specificity.

Results: A total of 102 patients suspected of pulmonary infection were enrolled in the study, 23 (22.5%) with PNID and 79 (77.5%) with PID. The diagnostic efficiency of BALF mNGS for differentiating PID from PNID was better than that of CMTs. Neutrophil percentage (N%) and the ratio of neutrophils to lymphocytes (N/L) in BALF were significantly lower in the PNID group than in the PID group. The AUCs for distinguishing PNID and PID were 0.739 (95% confidence interval [CI] 0.636−0.825) for BALF N%, 0.727 (95% CI 0.624−0.815) for BALF N/L, and 0.799 (95% CI 0.702−0.876) for BALF mNGS, with respective cutoff values of 6.7%, 0.255, and negative. Joint models of BALF mNGS combined with BALF N/L or BALF N% increased the respective AUCs to 0.872

(95% CI 0.786–0.933) and 0.871 (95% CI 0.784–0.932), which were significantly higher than those for BALF mNGS, BALF N%, and BALF N/L alone.

Conclusions: BALF N% ≤ 6.7% or BALF N/L ≤ 0.255 combined with a negative BALF mNGS result can effectively distinguish PNID from PID in immunocompetent patients with suspected pulmonary infection. BALF mNGS outperforms CMTs for identifying pathogens in immunocompetent patients, and the combination of mNGS and CMTs may be a better diagnostic strategy.

1 Introduction

Pulmonary infection is one of the most common reasons for hospitalization (Rana et al., 2021). However, a considerable number of patients who are initially diagnosed with pulmonary infectious disease (PID) do not exhibit any signs of recovery after treatment with various antibiotics, and their condition ultimately proves to be a pulmonary non-infectious disease (PNID) (Flament et al., 2016; Cottin and Brown, 2019; Fujita et al., 2020). Patients with PID and PNID usually have similar symptoms and physical signs (Allen and Wert, 2018; Raghu and Meyer, 2021), making it difficult to distinguish between the two based on inquiry and physical examination alone, without auxiliary examinations.

The treatment principles of PID and PNID are quite different. Misdiagnosis can lead to a prolonged recovery time, or worse, result in irreparable losses (Cosgrove et al., 2018). Prescribing antibiotics to patients with PNID leads to abuse-grade resistance, and prolongs the patient's condition. Rapid and accurate identification of PID and PNID is therefore crucial.

Clinical parameters such as routine blood test results, C-reactive protein (CRP), procalcitonin (PCT), and erythrocyte sedimentation rate (ESR) are often used to assist determination of the severity of infection (Wacker et al., 2013; Prina et al., 2015; Sproston and Ashworth, 2018; Lapic et al., 2020). They can also be used to predict pulmonary infection, and are relatively easy to obtain, particularly in primary settings. However, their predictive role with respect to infectious diseases remains controversial (Ito and Ishida, 2020; Niu et al., 2021).

Bronchoscopy facilitates the examination of airways, and the acquisition of bronchoalveolar lavage fluid (BALF). Bronchoscopy is a recommended method for the diagnosis of PID (Metlay et al., 2019). Early bronchoscopy in immunocompromised patients with pulmonary infiltrates can reportedly improve outcomes (Gonski et al., 2020; Bourne et al.,

2021; Saksirisampant et al., 2022). Cellular analysis of BALF is simple and safe, and can reflect some underlying features of disease. Kono et al. (Kono et al., 2021) reported that the ratio of lymphocytes to neutrophils (N/L) in BALF can predict the prognosis of acute exacerbation of interstitial lung disease. BALF can also be used for smears, cultures, and metagenomic next-generation sequencing (mNGS) to identify PID pathogens. Compared with conventional microbial detection methods such as smears and cultures, mNGS is evidently superior with regard to diagnostic efficiency (Wilson et al., 2014; Zhou et al., 2016; Takeuchi et al., 2019; Huang et al., 2020). While most studies emphasize the sensitivity of mNGS, however, its excellent specificity is rarely mentioned.

The aim of the current study was to determine whether cellular analysis and mNGS of BALF, and other infection indexes alone or in different combinations, could differentiate between PID and PNID in immunocompetent patients; and therefore enhance the accuracy of clinical diagnoses.

2 Methods

2.1 Participants and study design

Patients who presented with suspected pulmonary infections at Shanghai General Hospital in Shanghai, China, from December 2017 to March 2020 were retrospectively reviewed. The Ethics Committee of Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine approved the protocol (2021KY066). Due to the retrospective nature of the study, the requirement for written informed consent was waived.

The inclusion criteria were (1) suspected pulmonary infection based on at least one compatible symptom such as new-onset fever, cough, or dyspnea, and new-onset high-resolution computed tomography (HRCT) findings on chest

images; (2) the performance of bronchoscopy and bronchoalveolar lavage (BAL), and consent to perform BALF mNGS; (3) BALF and other relevant samples available for standard procedures and BALF mNGS within 48 h after admission; and (4) complete medical data recorded. The exclusion criteria were (1) the presence of an immuno suppressive condition, which was defined as suffering from any of the following: 1) primary immune deficiency diseases; 2) active malignancy excluding patients with early-stage cancers (eg, stage 1 lung cancer); 3) receiving cancer chemotherapy; 4) HIV infection with a CD4 T-lymphocyte count < 200 cells/µL or percentage < 14%; 5) solid organ transplantation; 6) hematopoietic stem cell transplantation; 7) receiving corticosteroid therapy with a dose ≥ 20 mg prednisone or equivalent daily for ≥ 14 d or a cumulative dose > 600 mg of prednisone; 8) receiving biological immune modulators; 9) receiving disease-modifying antirheumatic drugs or other immunosuppressive drugs (eg, cyclosporin, cyclophosphamide, hydroxychloroquine, methotrexate) (Ramirez et al., 2020); (2) BALF samples or detection process failed to pass quality control for mNGS; (3) repeated enrollment of the same patient; and (4) incomplete medical history.

A total of 102 patients were included in the analysis, and categorized into two groups based on final diagnoses: PNID and PID. Samples were subjected to conventional microbiological tests (CMTs), and mNGS of BALF was conducted in a pairwise manner. The CMTs used in this study are detailed in Table S1, and were conducted in accordance with previous studies (Parize et al., 2017; Peng et al., 2021). They including culture, serological diagnosis, antigen detection, PCR, and direct microscopic examination of specimens. The results of CMTs were interpreted in accordance with standard procedures (Patterson et al., 2016; Azoulay et al., 2020). The parameters assessed as infection indices included routine blood tests, high-sensitivity CRP (hs-CRP), PCT, interleukin (IL) 6, endotoxin, and ESR.

2.2 Fiberoptic bronchoscopy

Fiberoptic bronchoscopy was performed within 2 days after the identification of pulmonary infiltrates *via* HRCT. Most patients underwent fiberoptic bronchoscopy in the bronchoscopy unit. Some were performed at the bedside in the intensive care unit. BALF sampling was performed by experienced physicians in accordance with the American Thoracic Society (ATS) operating guidelines. All BALF samples were obtained from the area of lung infiltration. If there were multiple areas of infiltration, the sample was obtained where the infiltration was most severe. All BALF samples were > 20 mL in volume.

2.3 BALF cell count and cell classification

BALF samples were mixed 10–20 times, then approximately 5–10 µL of the sample was loaded into a Neubauer counting plate. After standing for 1 min, the number of cells was counted. The remainder of the sample was centrifuged at 400 *g*/min for 10 mins. After discarding the supernatant, the precipitate was pipetted onto a glass slide, spread evenly, then air-dried naturally. The slides were then Wright Giemsa stained. A site with a uniform distribution of cells was then selected by an experienced cytomorphologist under light microscopy, and at least 200 cells were counted and classified as neutrophils, lymphocytes, eosinophils, basophils, macrophages, ciliated columnar cells, or epithelium. Classification results were recorded as percentages.

2.4 mNGS

2.4.1 Sample processing and nucleic acid extraction

BALF samples were collected in accordance with standard procedures. DNA was extracted from the samples using the TIANamp Micro DNA Kit (DP316; Tiangen Biotech, Beijing, China) in accordance with the manufacturer's protocol.

2.4.2 Construction of DNA libraries

Single-stranded DNA (ssDNA) libraries were constructed after DNA fragmentation, end repair, adapter ligation, denaturation into single strands, and circularization. DNA nanoballs were generated from ssDNA by rolling circle amplification, loaded into the flow cell, and sequenced on a BGISEQ-200 platform (BGI, Beijing, China) and a NextSeq 550 platform (Illumina, California, USA) (Jeon et al., 2014; Miller et al., 2019).

2.4.3 Sequencing and bioinformatic analysis

High-quality sequencing data were generated by removing low-quality and short-length (< 35-bp) reads, followed by a computational subtraction of human sequences mapped to the human reference genome (hg19) *via* Burrows–Wheeler alignment (Li and Durbin, 2009). After removing low-complexity reads, the remaining data were classified *via* simultaneous alignment with four NCBI microbial genome databases (ftp://ftp.ncbi.nlm.nih.gov/genomes/). These databases included whole genome sequences of 4061 viral taxa, 2473 bacterial genomes or scaffolds, and genomic sequences for 199 fungi related to human infection and 135 parasites associated with human diseases.

2.4.4 Criteria for a positive mNGS result

Given the lack of a standard method for interpreting mNGS results and the variety of reporting parameters amongst different sequencing platforms, we used the following criteria in this study, which was derived and revised from prior literature (Qian et al., 2020; Peng et al., 2021), to define clinically significant microbes.

The sequencing results of each sample were categorized into 2 tables, each presenting bacteria, fungi and virus, respectively. The specifically mapped read number (SMRN) of each microbial taxonomy was normalized to SMRN per 20 million (M) of total sequencing reads (SDSMRN, standardized SMRN).

$$SDSMRN = \frac{SMN}{Total\ reads} \times 20\ million$$

A bacterial/fungal species was considered positively detected if: 1) it belonged to the top 10 genera with the highest SDSMRN; 2) it ranked first within its genus; 3) it had a SDSMRN>1; and 4) it was a commonly reported pulmonary infectious pathogen. A virus was considered positively detected if: 1) it was among the top 3 viruses with highest SDSMRN; and 2) it had a SDSMRN > 5.

There were several exceptions for certain pathogens. Because the possibility of Mycobacterium spp., Nocardia spp., etc. contamination and yield rate were low, they were considered positively detected when the SMRN at the species level was more than 3. Given the balance of environmental contamination and the difficulty of DNA extraction, molds, including Aspergillus spp., Rhizopus spp., and Mucor spp., with literature-proven pulmonary pathogenicity, were considered positively when the SMRN at the species level was more than 10.

2.5 Diagnosis of pulmonary infections

The final diagnosis was made by two intensivists with expertise in the management of infection after independently reviewing the medical records including clinical manifestations, laboratory tests, chest HRCT, microbiological tests (including CMTs and BALF mNGS), and treatment responses. Any disagreement between the two intensivists was resolved by in-depth discussion, and another senior intensivist was consulted if a consensus could not be reached.

2.6 Statistical analysis

Data analyses were performed using SPSS version 24.0 (SPSS Inc., Chicago, Illinois, USA) and R version 4.1.1 (Innovative Solutions, St. Louis, MO, USA). Baseline data are presented descriptively. Normality of data distributions was assessed with the Kolmogorov–Smirnov test. Normally distributed data are presented as means ± standard deviation. Non-normally distributed data are presented as medians and interquartile ranges (IQRs). Comparisons between patients with and without infection or organizing pneumonia were performed *via* the Mann-Whitney U test. With the final diagnosis as the gold standard, the conventional method and BALF mNGS were analyzed and compared. Comparative analysis was conducted *via* the Pearson χ^2 test, Fisher's exact test, or the McNemar test for discrete variables where appropriate. Receiver operating characteristic (ROC) analysis was performed to determine the optimal cutoff for BALF cell patterns for infection or organizing pneumonia diagnosis. $p<0.05$ was considered to indicate statistically significant differences.

3 Results

3.1 Demographic and clinical characteristics

A total of 102 patients with suspected pulmonary infection were included in the final analysis. There were 23 patients in the PNID group (22.5%), and 79 patients in the PID group (77.5%) (Table 1). No demographic factors differed significantly in the two groups. Most patients received empiric antibiotic therapy before admission. Conventional infection indicators such as routine blood test results, hs-CRP, PCT, IL-6, endotoxin, and ESR did not differ significantly in the two groups. With respect to HRCT findings there were no significant differences in bilateral lesions, consolidation, ground-glass opacities, solid nodules, tree-in-bud infiltrates, atelectasis, or cavities. Fourteen patients in the PID group had bronchiectasis, and no patients in the PNID group had bronchiectasis (p>0.05).

3.2 Distributions of pulmonary infectious pathogens and non-infectious diseases

Among the 79 patients with PID, bacteria were the most common cause of infection (Table 2). *Mycobacterium tuberculosis* (13.9%) was the most commonly detected pathogen, followed by nontuberculosis mycobacteria (10.1%). Fungal infections occurred in 13.9% of patients, and the most common were *Aspergillus* and *Cryptococcus neoformans*. One patient was infected with cytomegalovirus, and three were infected with *Mycoplasma pneumoniae*. Limited by the sensitivity of current diagnostic techniques and disease progression, there were 26 patients whose specific pathogen was uncertain. The etiologies of PNID are shown in Table 3. The most common was organizing pneumonia (56.5%), followed by lung cancer (17.4%) and eosinophilic pneumonia (13.1%). These three are often mistaken for pneumonia (Rothberg, 2022).

TABLE 1   Demographic data and cytometry and infectious laboratory parameters in noninfectious pulmonary disease and infectious pulmonary disease.

	Non-infectious pulmonary disease (*n* = 23)	Infectious pulmonary disease (*n* = 79)	*p* value
Age (years)¶	57.00 (13.00)	59.00 (16.00)	0.416
Male, *n* (%)	9 (39.13%)	41 (51.90%)	0.166
Empirical use of antibiotics, *n* (%)	21 (91.30%)	68 (86.07%)	0.233
Laboratory findings			
Leukocytes (10^9/L)¶	7.02 (4.32)	8.21 (3.29)	0.086
Neutrophils (10^9/L)¶	4.82 (3.36)	3.52 (2.88)	0.097
Neutrophils (%)§	66.42 ± 9.31	64.39 ± 12.13	0.171
Lymphocytes (10^9/L)¶	1.64 (0.85)	1.44 (0.73)	0.151
Lymphocytes (%)§	24.56 ± 8.28	25.53 ± 10.79	0.211
Neutrophils/lymphocytes¶	2.49 (1.81)	2.73 (2.18)	0.770
Eosinophils (10^9/L)¶	0.15 (0.17)	0.13 (0.13)	0.706
Eosinophils (%)¶	1.90 (2.30)	2.20 (2.20)	0.715
hs-CRP (mg/L)¶	3.5 (18.50)	7 (25.58)	0.521
PCT (ng/mL)¶	0.037 (0.039)	0.037 (0.033)	0.983
IL-6 (pg/mL)¶	3.98 (38.79)	6.60 (12.69)	0.812
Endotoxin (EU/mL)¶	0.027 (0.041)	0.032 (0.055)	0.368
ESR (mm/h)¶	22.5 (72.5)	23.5 (40.0)	0.996
HRCT findings			
Bilateral pulmonary lesions, *n* (%)	16.00 (69.57%)	44 (55.70%)	0.234
Consolidation, *n* (%)	13 (56.52%)	48 (60.76%)	0.715
Ground-glass opacities, *n* (%)	7 (30.43%)	13 (16.46%)	0.235
Solid nodules, *n* (%)	6 (26.09%)	15 (18.99%)	0.654
Bronchiectasis, *n* (%)	0 (0.00%)	14 (17.72%)	0.067
Tree-in-bud infiltrates, *n* (%)	2 (8.70%)	7 (8.86%)	1.000
Atelectasis, *n* (%)	1 (4.35%)	4 (5.06%)	1.000
Cavities, *n* (%)	3 (13.04%)	14 (17.72%)	0.832

hs-CRP, high-sensitivity C-reactive protein; PCT, procalcitonin; IL-6, interleukin 6; ESR, erythrocyte sedimentation rate; HRCT, high-resolution computed tomography.
§ Mean ± standard deviation values.
¶ Median (interquartile range).
Statistical significance is shown by bold font.

3.3 Comparison of BALF mNGS and CMTs

3.3.1 Comparison of diagnostic performance for differentiating PID from PNID

The positivity rates of BALF mNGS and CMTs for pulmonary infectious and noninfectious disease are shown in Figure 1A. The positive predictive value for diagnosing infectious disease *via* BALF mNGS was 97.8%, and the negative predictive value was 38.6%. The positive likelihood ratio was 12.81, and the negative likelihood ratio was 0.46. BALF mNGS increased the sensitivity rate by approximately 27% in comparison with CMTs (55.7% vs. 29.1%, *p*<0.01), whereas there was no significant difference in specificity (95.7% vs. 78.3%, *p*>0.1) (Figure 1A). Additional details are provided in Table S2.

TABLE 2 Distribution of pathogens identified in patients with infectious pulmonary disease.

Pathogen	n (%)*
Bacteria	43 (54.4%)
Mycobacterium tuberculosis	11 (13.9%)
Nontuberculosis mycobacteria	8 (10.1%)
Hemophilus parainfluenzae	5 (6.3%)
Pseudomonas aeruginosa	4 (5.1%)
Klebsiella pneumoniae	4 (5.1%)
Enterococcus faecium	2 (2.5%)
Moraxella catarrhalis	2 (2.5%)
Nocardia	2 (2.5%)
Other bacteria#	5 (6.3%)
Fungi	10 (12.7%)
Aspergillus spp.	4 (5.1%)
Cryptococcus neoformans	4 (5.1%)
Candida spp.	2 (2.5%)
Viruses	1 (1.3%)
Cytomegalovirus	1 (1.3%)
Atypical pathogens	3 (3.8%)
Mycoplasma pneumonia	3 (3.8%)

*Number of patients, with the percentage in parentheses.
#Including Staphylococcus aureus (n = 1); Acinetobacter lwoffii (n = 1); Tropheryma whipplei (n = 1); Streptococcus constellatus (n = 1); Enterobacter cloacae (n = 1).

3.3.2 Diagnosis assisted by BALF mNGS for PID without identifiable etiology by CMTs

Among the 79 patients diagnosed with pulmonary infectious diseases, 44 were confirmed by BALF mNGS and 23 were confirmed by CMTs. Notably 30 (38.0%) were BALF mNGS-positive despite comprehensive CMTs being negative. Bacteria accounted for the highest proportion of confirmed pathogens (74%) (Figure 1B). Antibiotics use was adjusted in these 30 patients based on the results of BALF mNGS, and most patients improved.

TABLE 3 Distribution of non-infectious pulmonary diseases.

Disease	n (%)*
Organizing pneumonia	13 (56.5%)
Eosinophilic pneumonia	3 (13.1%)
Bronchiolitis	2 (8.7%)
Lung cancer	4 (17.4%)
Still's disease	1 (4.3%)

*Number of patients, with the percentage in parentheses.

3.4 Findings of BALF cellular analysis

BAL was performed safely in all patients, irrespective of whether the patient was intubated. N% and N/L in BALF were significantly lower in the PNID group (both $p<0.01$) (Table 4).

3.5 Diagnostic accuracy of single measurements used for PNID prediction

There were significant differences in N% and N/L in BALF between patients with and without infectious disease, indicating that these measurements may predict PNID. Therefore, ROC curves were constructed to evaluate the prognostic value of these measurements for PNID, and AUCs were calculated. In ROC analysis of BALF N% the AUC was 0.739 (95% CI: 0.636–0.825), and the AUC of BALF N/L was 0.727 (95% CI: 0.624–0.815) (Figure 2A). The AUC of BALF N% and BALF N/L combined was 0.742 (95% CI: 0.640–0.828), which was not significantly higher than that of BALF N% alone. The AUC of BALF mNGS for PNID prediction was 0.799 (95% CI: 0.702–0.876) (Figure 2B).

3.6 Diagnostic accuracy of BALF cellular analysis combined with mNGS for PNID prediction

To determine whether combining measurements would improve PNID prediction binary logical regression of BALF N% and BALF mNGS was performed, and it yielded an AUC of 0.872 (95% CI: 0.786–0.933) (Figure 3A), which was significantly higher than those of BALF N% alone and BALF mNGS alone (both $p<0.05$). In ROC analysis of BALF N/L combined with BALF mNGS the AUC was 0.871 (95% CI: 0.784–0.932) (Figure 3B), which was significantly higher than that of BALF mNGS alone and BALF N/L alone (both $p<0.05$).

3.7 Optimal cutoff values for PNID

Optimal BALF N%, BALF N/L, and BALF mNGS cutoff values for predicting PNID were calculated as the values that gave the highest sum of sensitivity and specificity in the immunocompetent patients. The cutoff values for the measures with the highest AUCs were 7% for BALF N%, 0.25 for BALF N/L, and negative for BALF mNGS results. Using the cutoff values, the continuous test variables were converted to dichotomous state variables by defining BALF N%≤6.7% as "1", BALF N% >6.7% as "0", BALF N/L ≤ 0.255 as "1", BALF N/L>0.255 as "0", a negative BALF mNGS result as "1", and a positive BALF mNGS result as "0". Logistic regression was then performed to generate a predictive equation based on the combination of
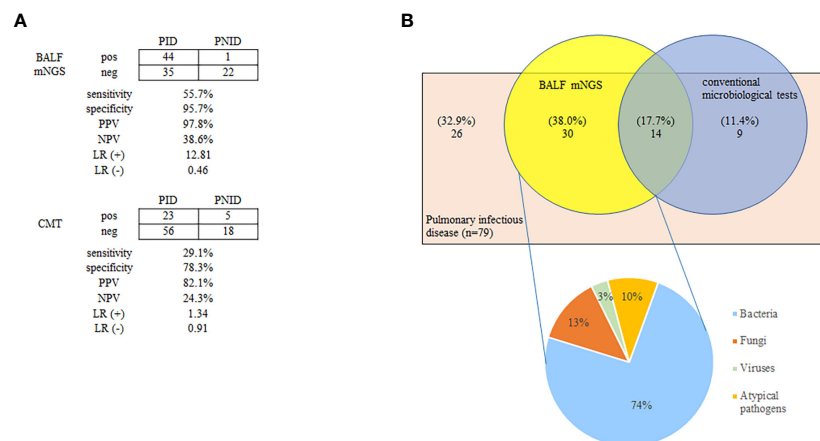
FIGURE 1
Positivity rates of BALF mNGS and CMTs for pulmonary infectious and non-infectious diseases **(A)**, and diagnosis assisted by BALF mNGS for PID without identifiable etiology by CMTs **(B)**. The positive predictive value of BALF mNGS for the diagnosis of infectious disease was 97.8%, and the negative predictive value was 38.6%. BALF mNGS increased the sensitivity rate by approximately 27% in comparison with CMTs (55.7% vs. 29.1%; $p < 0.01$) **(A)**. Thirty patients (38.0%) were BALF mNGS-positive despite comprehensive CMTs being negative. Bacteria constituted the highest proportion of pathogens (74%) **(B)**. BALF, bronchoalveolar lavage fluid; CMT, conventional microbiological test; LR(+), positive likelihood ratio; LR(-), negative likelihood ratio; mNGS, metagenomic next-generation sequencing; NPV, negative predictive value; PPV, positive predictive value; PID, pulmonary infectious disease; PNID, pulmonary non-infectious disease; ROC, receiver operating characteristic.

measurements, followed by multiple logistic regression analyses of BALF N%, BALF N/L, BALF mNGS, BALF N% combined with BALF mNGS (model 1), and BALF N/L combined with BALF mNGS (model 2). The AUC of model 1 was 0.879 (95% CI: 0.793–0.938), the AUC of the model 2 was 0.853 (95% CI: 0.763–0.919), and the AUCs of BALF N%, BALF N/L, and BALF mNGS were 0.737 (95% CI: 0.635–0.824), 0.689 (95% CI: 0.583–0.782), and 0.799 (95% CI: 0.702–0.876), respectively (Figure 4).

4 Discussion

To our knowledge, our research pioneered the use of the combination of BALF cellular analysis and BALF mNGS to distinguish between PNID and PID. The joint models of BALF N%≤6.7% and BALF N/L ≤ 0.255 combined with a negative BALF mNGS result can predict PNID, and may provide a rapid and precise method for physicians to differentiate between non-infectious and infectious diseases. Compared with CMTs, BALF mNGS had higher diagnostic efficiency and effectively filled the gap when CMT evidence was insufficient.

Pulmonary infection is one of the most common types of infection, and a common cause of hospitalization and death in older adults (Global Burden of Disease Study, 2015). Patients with suspected pulmonary infection who require hospitalization tend to have more complicated conditions (Postma et al., 2017). In the current study more than half the patients exhibited a multilobar distribution of pleomorphic lesions on HRCT, and the vast majority had administered antibiotics before admission. Among the patients diagnosed with pulmonary infection, more than 20% had mycobacterial infections. All these factors indicated that it was not simply a case of community-acquired pneumonia. Given the

TABLE 4 Characteristics of BALF cellular analysis in non-infectious pulmonary disease and infectious pulmonary disease.

	Non-infectious pulmonary disease ($n = 20$)	Infectious pulmonary disease ($n = 71$)	p value
BALF karyocytes (10^6/L)¶	280 (330)	280 (621)	0.562
BALF lymphocytes (%)¶	11.65 (47.40)	8.00 (21.00)	0.087
BALF neutrophils (%)¶	4.00 (10.75)	37.50 (72.00)	**0.001**
BALF neutrophils/lymphocytes¶	0.25 (2.17)	3.57 (16.92)	**0.002**

Statistical significance is indicated by bold font.
BALF, bronchoalveolar lavage fluid.
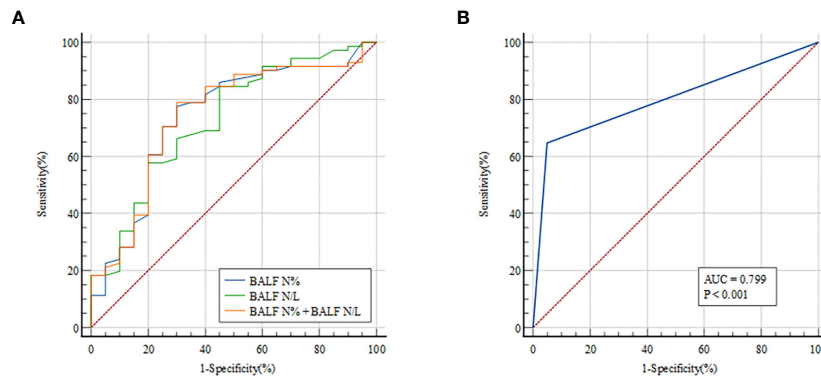¶ Median (interquartile range).

FIGURE 2
ROC curves for predicting PNID with single measurements. AUCs were 0.739 (95% CI 0.636−0.825) for BALF N% **(A)**, 0.727 (95% CI 0.624−0.815) for BALF N/L **(A)**, 0.742 (95% CI 0.640−0.828) for BALF N% combined with BALF N/L **(A)**, and 0.799 (95% CI 0.702−0.876) for BALF mNGS **(B)**. AUC, area under the curve; BALF, bronchoalveolar lavage fluid; mNGS, metagenomic next-generation sequencing; N, neutrophils; N/L, ratio of neutrophils to lymphocytes; PNID, pulmonary non-infectious disease; ROC, receiver operating characteristic.

complexity of hospitalized patients, the diagnoses of PNID and PID were based on clinical manifestations, laboratory tests, chest HRCT, CMTs, BALF mNGS, and treatment responses.

Routine blood tests are the most commonly used diagnostic method, and can reflect the severity of infection to an extent. Physicians in primary hospitals usually diagnose infections based on the results of routine blood tests. In the present study there were no significant differences in routine blood test indexes between the PNID group and the PID group, which is consistent with previous studies (Zhu et al., 2015; Shaddock, 2016). CRP, PCT, and ESR are clinically validated indicators that can be used to assist the diagnosis of infection (Wang et al., 2019; Yin and Mo, 2022). However, many non-infectious conditions such as allergies and autoimmune diseases can dramatically influence CRP and ESR values (Giacomelli

et al., 2018; Davis and van der Hilst, 2018; Aringer, 2020), whereas PCT does not change significantly in some Gram-positive bacterial and fungal infections or local infections (Ito and Ishida, 2020; Xu et al., 2021). In the current study these three indicators were poor for distinguishing between PNID and PID. This may be related to the empirical use of antibiotics before admission. The same results were also evident with respect to emerging indicators for infection assessment such as IL-6 and endotoxin. The principles of treatment for non-infectious diseases and infectious diseases are quite different, and patients with suspected pulmonary infection requiring hospitalization tend to have complex medical conditions. Inappropriate treatment can lead to prolonged disease, antibiotic abuse, and even death. It is therefore important to identify an effective method to distinguish PNID from PID.
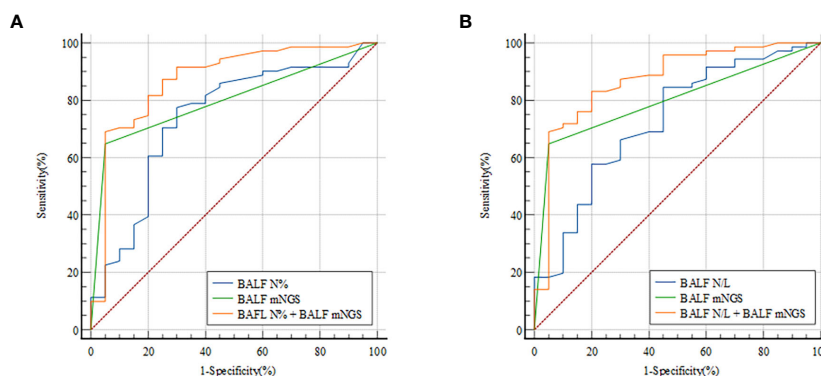
FIGURE 3
ROC curves for predicting PNID with BALF cellular analysis combined with mNGS. AUCs were 0.872 (95% CI 0.786−0.933) for BALF N% combined with BALF mNGS **(A)** and 0.871 (95% CI 0.784−0.932) for BALF N/L combined with BALF mNGS **(B)**. AUC, area under the curve; BALF, bronchoalveolar lavage fluid; mNGS, metagenomic next-generation sequencing; N, neutrophils; N/L, ratio of neutrophils to lymphocytes; PNID, pulmonary non-infectious disease; ROC, receiver operating characteristic.

FIGURE 4
ROC curves of dichotomous state variables of BALF mNGS combined with BALF N% (model 1) or BALF N/L (model 2) for predicting PNID, compared with single variables alone. Dichotomous state variables were separated in accordance with cutoff values of BALF mNGS, BALF N%, and BALF N/L. $AUC_{model\ 1}$ = 0.879 (95% CI 0.793–0.938); $AUC_{model\ 2}$ = 0.853 (95% CI 0.763–0.919); $AUC_{BALF\ N\%}$ = 0.737 (95% CI 0.635–0.824); $AUC_{BALF\ N/L}$ = 0.689 (95% CI 0.583–0.782); $AUC_{BALF\ mNGS}$ = 0.799 (95% CI 0.702–0.876). AUC, area under the curve; BALF, bronchoalveolar lavage fluid; mNGS, metagenomic next-generation sequencing; N, neutrophils; N/L, ratio of neutrophils to lymphocytes; PNID, pulmonary non-infectious disease; ROC, receiver operating characteristic.

BAL is a simple and safe procedure. BALF is obtained from deep bronchi which are relatively sterile, and it contains significantly less background microorganisms than sputum, so it can better reflect the local microenvironment of the lungs, and is more suitable for culture and mNGS. The proportion of lymphocytes in BALF can reportedly be used as a prognostic predictor of acute exacerbation of interstitial lung disease (Takei et al., 2017; Kono et al., 2021). The proportion of eosinophils in BALF can constitute evidence for the diagnosis of eosinophil-associated pneumonia and parasitic infection (Ravin and Loy, 2016; Allen and Wert, 2018). Notably however, few studies have investigated whether cellular analysis of BALF can be used to distinguish between PID and PNID. In the current study BALF N% and BALF N/L could effectively differentiate between PNID and PID, and a possible explanation is that neutrophils in BALF are closer to the pulmonary lesions and more capable of responding to local infection and inflammation of lung. Bronchoscopy is now available in most hospitals, and as a safe and convenient examination BAL is receiving increasing attention from respiratory specialists. We recommend routine BAL to help distinguish between PNID and PID, especially in patients with severe illness.

mNGS is an unbiased detection method that can theoretically detect all pathogens in clinical samples (Chiu and Miller, 2019). In this study, the sensitivity of BALF mNGS for detecting pathogens was 55.7% and the sensitivity of CTMs was 29.1%. Most of the patients with typical symptoms were effectively treated in the outpatient department, while the patients hospitalized in tertiary hospitals for pulmonary infection tended to have more complicated conditions and most of them had been empirically treated with broad-antibiotics. 87.3% of the patients included in our study had received empirical antibiotics before admission. In the context of prior antibiotic application, the sensitivity of CMTs is significantly lower than that of mNGS. Our results are consistent with previous study (Miao et al., 2018). To date most research has focused on the ability of mNGS to detect pathogens, and has emphasized its sensitivity. Conversely mNGS is limited by interference from background microorganisms and contamination, so results may be false-positives, leading to overdiagnosis (Simner et al., 2018). However, this feature means that negative mNGS results are more credible. In the present study negative BALF mNGS results were used to exclude pulmonary infection in order to screen for PNID. BALF mNGS could effectively predict PNID, and its diagnostic efficiency was improved when it was combined with cellular BALF analysis. Therefore, we recommend that BALF mNGS should be conducted when economic considerations permit or the patient is severely ill, to help physicians make correct diagnoses in a short time and formulate appropriate treatment plans.

The current study had some limitations. It was a single-centered study with a relatively small sample size, therefore multicenter, large-scale, prospective studies are needed in the future to validate the results. Immunocompetent patients were recruited rather than immunosuppressed patients, which may be associated with significantly different pathogen profiles, as well as diverse diagnostic performance of mNGS and cellular analysis of BALF. Lastly, the lack of RNA sequencing and insufficient PCR methods hindered evaluation of the diagnostic value of CMTs and mNGS with respect to virus infection.

In conclusion, the current study indicates that BALF N% $\leq$ 6.7% or BALF N/L $\leq$ 0.255 combined with a negative BALF mNGS result can effectively distinguish PNID from PID in immunocompetent patients with suspected pulmonary infection. It may help physicians to determine the optimal treatment, and avoid the abuse of antibiotics. With regard to primary hospitals without access to complex equipment, BAL is a safe, convenient, and economical method that can improve the accuracy of PNID and PID diagnoses. BALF mNGS outperforms CMTs with respect to identifying pathogens in immunocompetent patients, and the combination of mNGS and CMTs may be a better diagnostic strategy.

Data availability statement

The data presented in the study are deposited in CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb), accession number CNP0003771.

Ethics statement

The studies involving human participants were reviewed and approved by The ethics committee of Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine. The ethics committee waived the requirement of written informed consent for participation.

Author contributions

MZ, PZ, and YP conceived of and designed the entire study. YZ, YP, WB, and DY contributed to acquisition of data. YP, XZ, and YS performed statistical analyses. MZ, PZ, and YP were involved in the interpretation of data. YP, XZ, and YS wrote the manuscript, supervised by MZ. All authors contributed to the article and approved the submitted version. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2022.1023978/full#supplementary-material

SUPPLEMENTARY TABLE 1
Conventional microbiological methods used in the study.

SUPPLEMENTARY TABLE 2
Sample details of enrolled cases.

References

Allen, J., and Wert, M. (2018). Eosinophilic pneumonias. *J. Allergy Clin. Immunol. Pract.* 6, 1455–1461. doi: 10.1016/j.jaip.2018.03.011

Aringer, M. (2020). Inflammatory markers in systemic lupus erythematosus. *J. Autoimmun.* 110, 102374. doi: 10.1016/j.jaut.2019.102374

Azoulay, E., Russell, L., Van De Louw, A., Metaxa, V., Bauer, P., Povoa, P., et al. (2020). Diagnosis of severe respiratory infections in immunocompromised patients. *Intensive Care Med.* 46, 298–314. doi: 10.1007/s00134-019-05906-5

Bourne, M. H.Jr., Norton, M. S., Midthun, D. E., Mullon, J. J., Kern, R. M., Utz, J. P., et al. (2021). Utility of transbronchial biopsy in the immunocompromised host with new pulmonary radiographic abnormalities. *Mayo Clin. Proc.* 96, 1500–1509. doi: 10.1016/j.mayocp.2020.09.039

Chiu, C. Y., and Miller, S. A. (2019). Clinical metagenomics. *Nat. Rev. Genet.* 20, 341–355. doi: 10.1038/s41576-019-0113-7

Cosgrove, G. P., Bianchi, P., Danese, S., and Lederer, D. J. (2018). Barriers to timely diagnosis of interstitial lung disease in the real world: the INTENSITY survey. *BMC Pulm Med.* 18, 9. doi: 10.1186/s12890-017-0560-x

Cottin, V., and Brown, K. K. (2019). Interstitial lung disease associated with systemic sclerosis (SSc-ILD). *Respir. Res.* 20, 13. doi: 10.1186/s12931-019-0980-7

Davis, M. D. P., and van der Hilst, J. C. H. (2018). Mimickers of urticaria: Urticarial vasculitis and autoinflammatory diseases. *J. Allergy Clin. Immunol. Pract.* 6, 1162–1170. doi: 10.1016/j.jaip.2018.05.006

Flament, T., Bigot, A., Chaigne, B., Henique, H., Diot, E., and Marchand-Adam, S. (2016). Pulmonary manifestations of sjogren's syndrome. *Eur. Respir. Rev.* 25, 110–123. doi: 10.1183/16000617.0011-2016

Fujita, N., Ando, M., Goto, A., Sakata, M., Ogata, M., Usagawa, Y., et al. (2020). Diffuse Large b-cell lymphoma arising from the lesion of chronic lobar atelectasis. *Tohoku J. Exp. Med.* 250, 129–135. doi: 10.1620/tjem.250.129

Giacomelli, R., Ruscitti, P., and Shoenfeld, Y. (2018). A comprehensive review on adult onset still's di1sease. *J. Autoimmun.* 93, 24–36. doi: 10.1016/j.jaut.2018.07.018

Global Burden of Disease Study (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries 1990-2013: A systematic analysis for the global burden of disease study 2013. *Lancet* 386, 743–800.doi: 10.1016/S0140-6736(15)60692-4

Gonski, K., Cohn, R., Widger, J., and Mcmullan, B. (2020). Utility of bronchoscopy in immunocompromised paediatric patients: Systematic review. *Paediatr. Respir. Rev.* 34, 24–34. doi: 10.1016/j.prrv.2020.02.003

Huang, Z., Li, W., Lee, G. C., Fang, X., Xing, L., Yang, B., et al. (2020). Metagenomic next-generation sequencing of synovial fluid demonstrates high accuracy in prosthetic joint infection diagnostics: mNGS for diagnosing PJI. *Bone Joint Res.* 9, 440–449. doi: 10.1302/2046-3758.97.BJR-2019-0325.R2

Ito, A., and Ishida, T. (2020). Diagnostic markers for community-acquired pneumonia. *Ann. Transl. Med.* 8, 609. doi: 10.21037/atm.2020.02.182

Jeon, Y. J., Zhou, Y., Li, Y., Guo, Q., Chen, J., Quan, S., et al. (2014). The feasibility study of non-invasive fetal trisomy 18 and 21 detection with semiconductor sequencing platform. *PloS One* 9, e110240. doi: 10.1371/journal.pone.0110240

Kono, M., Miyashita, K., Hirama, R., Oshima, Y., Takeda, K., Mochizuka, Y., et al. (2021). Prognostic significance of bronchoalveolar lavage cellular analysis in patients with acute exacerbation of interstitial lung disease. *Respir. Med.* 186, 106534. doi: 10.1016/j.rmed.2021.106534

Lapic, I., Padoan, A., Bozzato, D., and Plebani, M. (2020). Erythrocyte sedimentation rate and c-reactive protein in acute inflammation. *Am. J. Clin. Pathol.* 153, 14–29. doi: 10.1093/ajcp/aqz142

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Metlay, J. P., Waterer, G. W., Long, A. C., Anzueto, A., Brozek, J., Crothers, K., et al. (2019). 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. *Am. J. Respir. Crit. Care Med.* 200, e45–e67.doi: 10.1164/rccm.201908-1581ST

Miao, Q., Ma, Y., Wang, Q., Pan, J., Zhang, Y., Jin, W., et al. (2018). Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clin. Infect. Dis.* 67, S231–S240. doi: 10.1093/cid/ciy693

Miller, S., Naccache, S. N., Samayoa, E., Messacar, K., Arevalo, S., Federman, S., et al. (2019). Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res.* 29, 831–842. doi: 10.1101/gr.238170.118

Niu, D., Huang, Q., Yang, F., Tian, W., Li, C., Ding, L., et al. (2021). Serum biomarkers to differentiate gram-negative, gram-positive and fungal infection in febrile patients. *J. Med. Microbiol.* 70. doi: 10.1099/jmm.0.001360

Parize, P., Muth, E., Richaud, C., Gratigny, M., Pilmis, B., Lamamy, A., et al. (2017). Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. *Clin. Microbiol. Infect.* 23, 574 e1–574.e6. doi: 10.1016/j.cmi.2017.02.006

Patterson, T. F., Thompson, G. R., Denning, D. W., Fishman, J. A., Hadley, S., Herbrecht, R., et al. (2016). Practice guidelines for the diagnosis and management of aspergillosis: 2016 update by the infectious diseases society of America. *Clin. Infect. Dis.* 63, e1–e60.doi: 10.1093/cid/ciw326

Peng, J. M., Du, B., Qin, H. Y., Wang, Q., and Shi, Y. (2021). Metagenomic next-generation sequencing for the diagnosis of suspected pneumonia in immunocompromised patients. *J. Infect.* 82, 22–27. doi: 10.1016/j.jinf.2021.01.029

Postma, D. F., Van Werkhoven, C. H., and Oosterheert, J. J. (2017). Community-acquired pneumonia requiring hospitalization: rational decision making and interpretation of guidelines. *Curr. Opin. Pulm Med.* 23, 204–210. doi: 10.1097/MCP.0000000000000371

Prina, E., Ranzani, O. T., and Torres, A. (2015). Community-acquired pneumonia. *Lancet* 386, 1097–1108. doi: 10.1016/S0140-6736(15)60733-4

Qian, Y. Y., Wang, H. Y., Zhou, Y., Zhang, H. C., Zhu, Y. M., Zhou, X., et al. (2020). Improving pulmonary infection diagnosis with metagenomic next generation sequencing. *Front. Cell Infect. Microbiol.* 10, 567615. doi: 10.3389/fcimb.2020.567615

Raghu, G., and Meyer, K. C. (2021). Cryptogenic organising pneumonia: current understanding of an enigmatic lung disease. *Eur. Respir. Rev.* 30. doi: 10.1183/16000617.0094-2021

Ramirez, J. A., Musher, D. M., Evans, S. E., Dela Cruz, C., Crothers, K. A., Hage, C. A., et al. (2020). Treatment of community-acquired pneumonia in

immunocompromised adults: A consensus statement regarding initial strategies. *Chest* 158, 1896–1911. doi: 10.1016/j.chest.2020.05.598

Rana, J. S., Khan, S. S., Lloyd-Jones, D. M., and Sidney, S. (2021). Changes in mortality in top 10 causes of death from 2011 to 2018. *J. Gen. Intern. Med.* 36, 2517–2518. doi: 10.1007/s11606-020-06070-z

Ravin, K. A., and Loy, M. (2016). The eosinophil in infection. *Clin. Rev. Allergy Immunol.* 50, 214–227. doi: 10.1007/s12016-015-8525-4

Rothberg, M. B. (2022). Community-acquired pneumonia. *Ann. Intern. Med.* 175, ITC49–ITC64. doi: 10.7326/AITC202204190

Saksirisampant, G., Kawamatawong, T., Promsombat, K., Sukkasem, W., Liamsombut, S., Pasomsub, E., et al. (2022). A prospective study of plasma and bronchoalveolar lavage fluid CMV DNA load quantification for the diagnosis and outcome of CMV pneumonitis in immunocompromised hosts. *J. Clin. Virol.* 155, 105243. doi: 10.1016/j.jcv.2022.105243

Shaddock, E. J. (2016). How and when to use common biomarkers in community-acquired pneumonia. *Pneum. (Nathan)* 8, 17.doi: 10.1186/s41479-016-0017-7

Simner, P. J., Miller, S., and Carroll, K. C. (2018). Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin. Infect. Dis.* 66, 778–788.doi: 10.1093/cid/cix881

Sproston, N. R., and Ashworth, J. J. (2018). Role of c-reactive protein at sites of inflammation and infection. *Front. Immunol.* 9, 754.doi: 10.3389/fimmu.2018.00754

Takei, R., Arita, M., Kumagai, S., Ito, Y., Noyama, M., Tokioka, F., et al. (2017). Impact of lymphocyte differential count > 15% in BALF on the mortality of patients with acute exacerbation of chronic fibrosing idiopathic interstitial pneumonia. *BMC Pulm Med.* 17, 67.doi: 10.1186/s12890-017-0412-8

Takeuchi, S., Kawada, J. I., Horiba, K., Okuno, Y., Okumura, T., Suzuki, T., et al. (2019). Metagenomic analysis using next-generation sequencing of pathogens in bronchoalveolar lavage fluid from pediatric patients with respiratory failure. *Sci. Rep.* 9, 12909.doi: 10.1038/s41598-019-49372-x

Wacker, C., Prkno, A., Brunkhorst, F. M., and Schlattmann, P. (2013). Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. *Lancet Infect. Dis.* 13, 426–435. doi: 10.1016/S1473-3099(12)70323-7

Wang, Y., Zhang, S., Li, L., and Xie, J. (2019). The usefulness of serum procalcitonin, c-reactive protein, soluble triggering receptor expressed on myeloid cells 1 and clinical pulmonary infection score for evaluation of severity and prognosis of community-acquired pneumonia in elderly patients. *Arch. Gerontol. Geriatr.* 80, 53–57. doi: 10.1016/j.archger.2018.10.005

Wilson, M. R., Naccache, S. N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., et al. (2014). Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl. J. Med.* 370, 2408–2417. doi: 10.1056/NEJMoa1401268

Xu, J., Yu, Y., Lv, J., Yang, S., Wu, J., Chen, J., et al. (2021). Application of metagenomic next-generation sequencing to diagnose pneumocystis jirovecii pneumonia in kidney transplantation recipients. *Ann. Transplant.* 26, e931059. doi: 10.12659/AOT.931059

Yin, H., and Mo, S. (2022). Value of combined detection of serum amyloid a, c-reactive protein and procalcitonin in differential diagnosis of respiratory tract infection in children of China. *Ann. Med.* 54, 1732–1737. doi: 10.1080/07853890.2022.2064542

Zhou, Y., Wylie, K. M., El Feghaly, R. E., Mihindukulasuriya, K. A., Elward, A., Haslam, D. B., et al. (2016). Metagenomic approach for identification of the pathogens associated with diarrhea in stool specimens. *J. Clin. Microbiol.* 54, 368–375. doi: 10.1128/JCM.01965-15

Zhu, G., Zhu, J., Song, L., Cai, W., and Wang, J. (2015). Combined use of biomarkers for distinguishing between bacterial and viral etiologies in pediatric lower respiratory tract infections. *Infect. Dis. (Lond)* 47, 289–293. doi: 10.3109/00365548.2014.987163