Society of Medical Education & Research

2026, Volume 6, Issue 1, Page No: 160-171
ISSN: 3108-4826

Journal of Medical Sciences and Interdisciplinary Research

A Combined Targeted and Metagenomic Nanopore Sequencing Method for Fast and Accurate Identification of Lower Respiratory Tract Infections

Charlotte Dupuis[1]*, Hugo Perrin[1], Elise Morel[2], Thomas Martin[3]

[1] Department of Respiratory Medicine and Metagenomic Diagnostics, Faculty of Medicine, University of Lille, Lille, France.
[2] Department of Nanopore Sequencing and Infectious Disease Diagnosis, Faculty of Medicine, University of Montpellier, Montpellier, France.
[3] Department of Lower Respiratory Tract Infection and Molecular Testing, Faculty of Medicine, University of Nantes, Nantes, France.

***E-mail** ✉ charlotte.dupuis@univ-lille.fr

Abstract

Nanopore-based metagenomic methods have proven useful for identifying bacterial causes of infections. Yet, thorough assessments of their effectiveness in everyday clinical practice for simultaneously detecting bacteria, fungi, and viruses remain limited. We created a paired sequencing strategy called the Nanopore Meta-Panel process (NanoMP). For each respiratory sample, it ran an untargeted metagenomic workflow (Meta) alongside a targeted enrichment workflow (Panel). This prospective multicenter trial included 450 respiratory samples. A machine learning-driven filtering framework, refined with ROC analysis from orthogonal testing on 21 key species, enhanced detection reliability. AMR gene detection relied on the CARD database combined with SNP profiling. In the Meta workflow, sensitivity reached 82.9% for bacteria, 88.7% for fungi (excluding Aspergillus species), and 75.0% for the Mycobacterium tuberculosis complex. The Panel workflow, through targeted amplification, substantially raised sensitivity for viruses (over 80.0% versus 39.4%) and Aspergillus (81.8% versus 42.3%). Across both processes, NanoMP delivered 80.2% sensitivity and 98.8% specificity against the composite reference standard. It also reliably pinpointed key AMR determinants like blaKPC-2, blaOXA-23, and mecA, correctly linking them to their host organisms even in polymicrobial cases. The NanoMP strategy merges unbiased metagenomics with targeted enrichment on the same specimen via parallel Nanopore runs. This enabled comprehensive detection of bacterial, viral, and fungal pathogens in lower respiratory samples and showed strong real-world clinical utility.

Keywords: Nanopore sequencing, Metagenomics, Target amplification, Lower respiratory infections, Machine learning

Introduction

Lower respiratory tract infections (LRTIs) — including community-acquired pneumonia (CAP), hospital-acquired pneumonia (HAP), and ventilator-associated pneumonia (VAP) — contribute to more than 25% of pneumonia-related hospital deaths [1]. Pinpointing responsible pathogens (bacteria, fungi, viruses, and atypicals) is difficult because conventional testing methods have limitations.

Metagenomic next-generation sequencing (mNGS) has moved swiftly into clinical laboratories [2, 3]. Nanopore sequencing stands out due to fast library preparation, real-time basecalling and analysis, rapid turnaround, and long reads, which are especially beneficial for AMR gene characterization [4]. Most published studies, however, have concentrated on bacteria and fungi [4-6], while viral detection performance has been weaker [7]. Large-scale studies addressing bacteria, fungi, and viruses together are scarce. Challenges like background contamination and overwhelming host DNA content hinder routine use, leading to inconsistent results across studies that apply different filtering criteria [5, 8]. Reliable thresholds

Access this article online https://smerpub.com/

Received: 14 January 2026; **Accepted:** 21 April 2026

How to cite this article: Dupuis C, Perrin H, Morel E, Martin T. A Combined Targeted and Metagenomic Nanopore Sequencing Method for Fast and Accurate Identification of Lower Respiratory Tract Infections. J Med Sci Interdiscip Res. 2026;6(1):160-71. https://doi.org/10.51847/J0drDH5J9D

therefore demand evaluation of diverse sequencing parameters.

To better align with clinical demands, we developed the NanoMP approach: an integrated dual-barcode system that performs unbiased metagenomic detection (Meta process) and pathogen-specific amplification (Panel process) on one sample. Because HSV1 (Human alphaherpesvirus 1), EBV (Human gammaherpesvirus 4), and CMV (Human betaherpesvirus 5) frequently occur in healthy carriers [9-11], these common DNA viruses, plus Aspergillus fumigatus and representative RNA viruses such as rhinovirus (Rhv), were added to the enrichment panel for validation.

We tested NanoMP performance in a prospective multicenter study—currently the largest of its kind—involving 450 specimens from 418 patients. The workflow uses dual barcodes for demultiplexing and combines full-spectrum metagenomics (with host DNA removal) and targeted amplification for improved detection of priority pathogens. We developed a new filtering algorithm based on machine learning and ROC analysis using multiple quality-control features. We also characterized AMR genes and their specific variants.

Materials and Methods

NanoMP approach

Our custom NanoMP pipeline includes two complementary arms. The Meta arm uses saponin-mediated selective lysis to deplete human DNA, followed by microbial nucleic acid extraction and library construction to survey the entire pathogen community. The Panel arm involves total nucleic acid extraction, specific target amplification, and library preparation (**Figure 1**). Host-depleted metagenomic data are available in the NCBI SRA under Bioproject PRJNA818777.

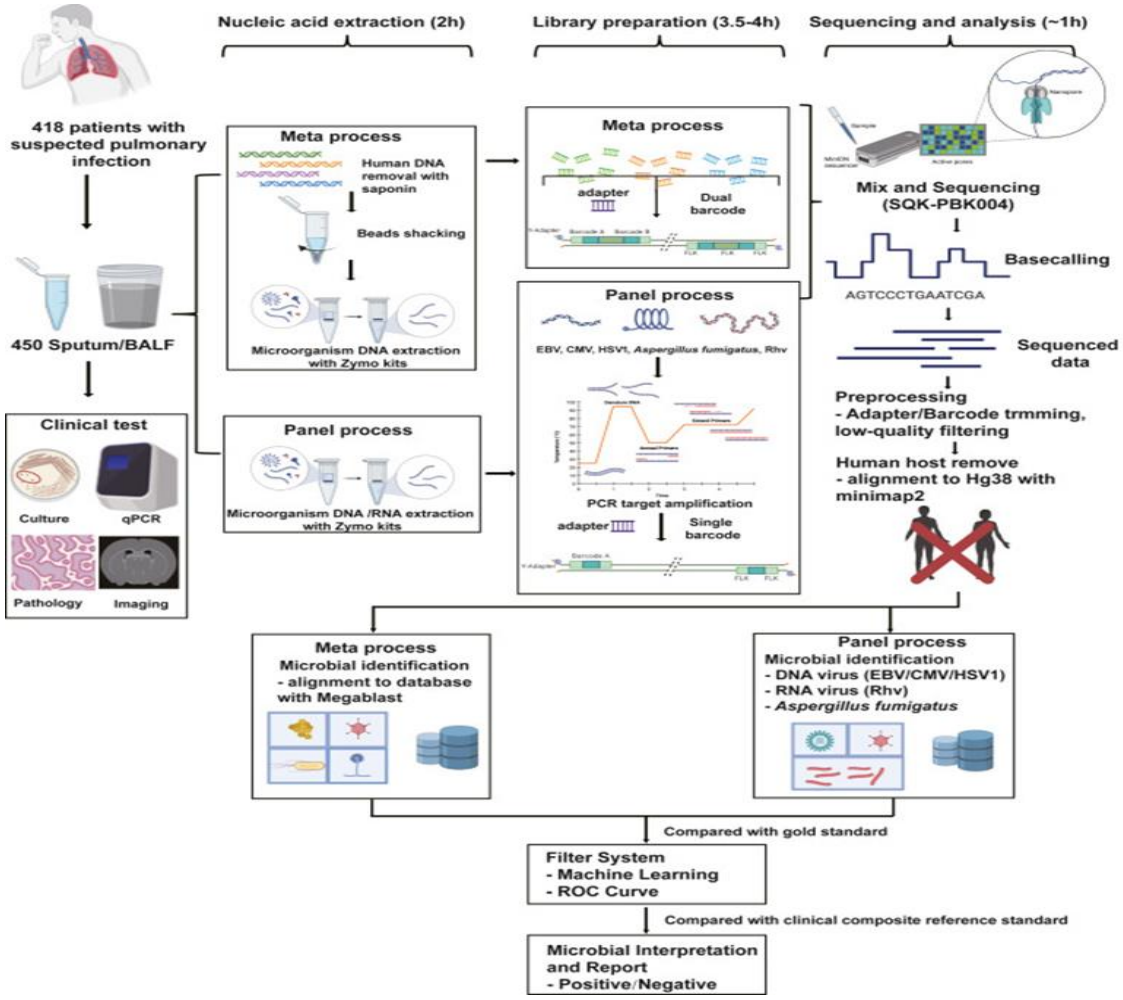


Figure 1. Overview of the study design.

We collected 450 sputum or bronchoalveolar lavage fluid (BALF) specimens from 418 individuals suspected of having pulmonary infections. Standard microbiological testing included culture, qPCR, ELISA, GeneXpert MTB/RIF, and histological examination. Each sample underwent parallel processing via two workflows: (1) the Meta workflow, which involved human genome depletion, microbial cell wall disruption, DNA isolation, dual-barcode attachment, and library construction; and (2) the Panel workflow, which included total DNA/RNA isolation, targeted amplification of selected organisms (such as EBV, CMV, HSV1, Aspergillus fumigatus, and Rhv), single-barcode attachment, and library construction. We pooled libraries from both workflows for sequencing. We generated final pathogen calls using established filtering criteria and compared them against the gold standard. Clinical reporting of pathogens incorporated physician judgment. Abbreviations: BALF = bronchoalveolar lavage fluid; ELISA = enzyme-linked immunosorbent assay; qPCR = quantitative polymerase chain reaction; EBV = Epstein–Barr virus (human gammaherpesvirus 4); CMV = human cytomegalovirus (human betaherpesvirus 5); HSV1 = human alphaherpesvirus 1; Rhv = rhinovirus.

Dual-barcode strategy for Nanopore sequencing

The commercial SQK-PBK004 kit provides only 12 barcodes, which was insufficient for the dual-process design (17 samples per flow cell, consisting of 6 samples processed in parallel for both workflows plus 5 negative controls). To address this, a custom dual-barcode system was developed specifically for the Meta workflow (**Figure 1**). Building on the standard barcode ligation approach used in commercial kits (as applied in the Panel workflow with Barcode A), an additional barcode (Barcode B) was incorporated into the flanking region (FLK) to generate dual barcodes. This allowed clear differentiation between Meta and Panel libraries.

Sample collection

The remaining 450 specimens consisted of bronchoalveolar lavage fluid (BALF) samples (n = 348) and sputum samples (n = 102). These were retrieved from the clinical laboratories of Peking University People's Hospital, China–Japan Friendship Hospital, and Beijing Chao-yang Hospital between January 28 and June 28, 2021. The study focused exclusively on the presence of LRTIs and did not consider gender. We obtained patient demographic information, laboratory results, imaging and pathology findings, treatment details, and clinical outcomes for the 418 individuals from electronic medical records.

Limit of detection

The limit of detection (LoD) for the assay was determined by spiking serial tenfold dilutions (ranging from 10 to 10^4 CFU/mL or 10 to 10^4 copies/mL) of Pseudomonas aeruginosa, Staphylococcus aureus, Aspergillus fumigatus, CMV, and Rhv into BALF samples containing normal respiratory flora (NRF) that tested negative for these organisms by qPCR. Background commensal levels were quantified using 16S qPCR as previously reported [4] and categorized as high (Ct = 25) or low (Ct = 28). Each dilution level was tested in ten replicates across the different background conditions. LoD was established as the lowest concentration at which the spiked pathogen was detected in at least 8 out of 10 mock samples.

Computational algorithm for pathogen identification

A specialized pathogen detection algorithm was created for Nanopore data to improve and validate diagnostic accuracy. Using alignments and results from culture and/or qPCR as reference, a panel of 21 representative species — including Gram-positive and Gram-negative bacteria, fungi, DNA viruses, RNA viruses, and Mycoplasma — was chosen. These organisms reflect commonly encountered and clinically significant pathogens. Whenever feasible, validation relied on gold-standard methods such as approved IVD kits or thoroughly validated laboratory-developed tests.

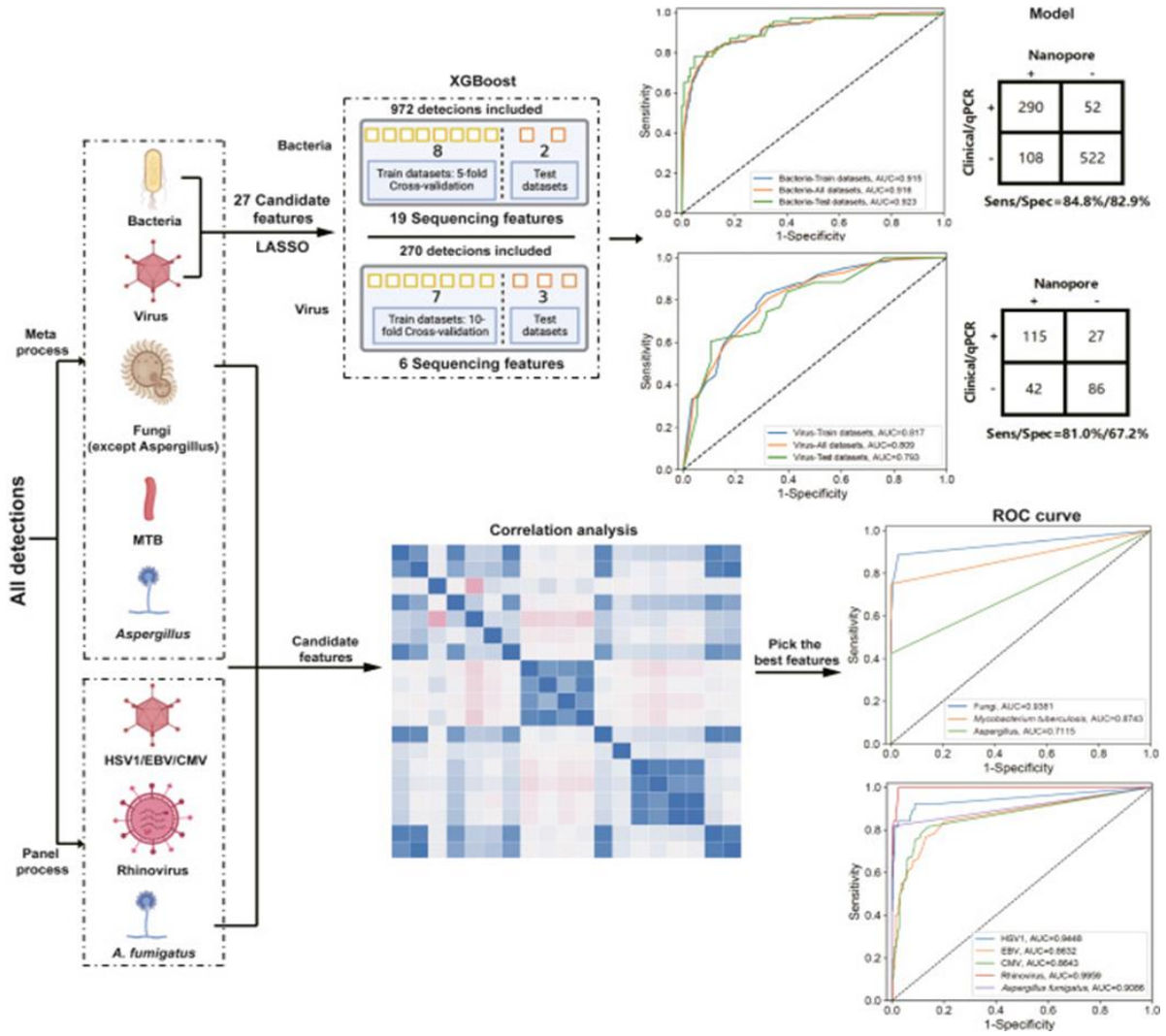


Figure 2. Process for developing the filtering framework.

For bacterial and viral categories in the Meta workflow, candidate variables were first identified through LASSO regression. The dataset incorporating these selected variables was split into training and testing sets. Hyperparameters for the XGBoost classifier were optimized using stratified cross-validation. Model performance was then evaluated with ROC curves using the best hyperparameters. For fungal groups (excluding Aspergillus), the MTB group, and Aspergillus in the Meta workflow, as well as for HSV1, EBV, CMV, Rhinovirus, and Aspergillus fumigatus in the Panel workflow, the most informative metrics were chosen according to Spearman's correlation. Optimal cut-off values were determined by identifying the point yielding the highest Youden's index on the ROC curves. Abbreviations: LASSO = least absolute shrinkage and selection operator; MTB = Mycobacterium tuberculosis; HSV1= human alphaherpesvirus 1; EBV = Epstein–Barr virus (human gammaherpesvirus 4); CMV = human cytomegalovirus (human betaherpesvirus 5).

Clinical judgment

Every organism identified through the Nanopore platform was cross-checked in PubMed and the CAP-CHINA database to evaluate its potential role in causing pneumonia. Sequencing outcomes were classified into four levels (Defined, Probable, Possible, and Unlikely), with specific standards outlined. The composite reference standard combined all microbiological results (culture and additional assays), pathology reports, and clinical expert reviews. Defined and Probable classifications counted as positive for diagnosis, while Possible and Unlikely were treated as negative. A sample

was considered to match the clinical diagnosis if any pathogen met the positive criteria.

Antimicrobial resistance gene identification

A single nucleotide polymorphism matrix (SNPM) reference database was constructed using sequences derived from homology models in the Comprehensive Antibiotic Resistance Database (CARD, v3.1.2).

Sequencing reads were aligned to CARD reference sequences with BLAST (v2.10.1+) to identify AMR gene families. For each read, we selected the most common AMR subtype among matching alignments and linked it to the corresponding reference in the SNPM. Subtype refinement used minimap2 and Samtools to detect allelic differences based on the SNPM. The BLAST scoring matrix was then restricted to these alleles, enabling site-specific scoring across subtypes. When scores for different subtypes differed significantly ($P < 0.05$), reads sharing a specific reference were assigned to the subtype with the highest median score; otherwise, they were assigned to the most prevalent subtype among those with the top score. Additional CARD annotations were applied to the identified AMR genes.

Statistical analysis

Sample size determination was performed with PASS software (v15.0; NCSS Corporation, USA). Group comparisons for continuous variables used the Student t-test or Mann–Whitney U test. In contrast, categorical variables were analyzed with the chi-square test, continuity correction, or Fisher's exact test as appropriate. A P-value below 0.05 was considered statistically significant. All analyses were conducted using SPSS software (v25.0; IBM Corporation, USA) and SciPy (v1.7.3, Python 3.7).

Results and Discussion

Patient characteristics

This prospective investigation enrolled 450 specimens (348 BALF and 102 sputum samples) collected from 418 patients. Participants were categorized into LRTI (n = 249) and non-LRTI (n = 169) groups according to established definitions for CAP [12], HAP [13], and suspected infection as determined by clinical specialists. As detailed in **Table 1**, patients in the LRTI group showed higher rates of aspiration history, invasive mechanical ventilation, and immunosuppression. Symptoms such as fever, sputum production, and shortness of breath were more common in the LRTI group. The non-LRTI group exhibited a greater proportion of malignancy cases but fewer individuals with hematological disorders (**Table 1**).

Table 1. The characteristics of patients in the LRTIs and non-LRTIs groups (n = 418).

Empty cell	Non-LRTIs (n = 169)	LRTIs (n = 249)	Median difference[a]/Odds ratio (95% CI)	P
Age, median (IQR), years	60.0 (53.5–66.5)	63.0 (51.0–71.0)	3.0 (−1.0 to 5.0)[a]	0.061
Sex				0.008
Male, n (%)	93 (55.0)	169 (67.9)	1.2 (1.1–1.4)	
Female, n (%)	76 (45.0)	80 (32.1)	0.7 (0.5–1.0)	
Infection types				
CAP	0	159 (63.9)	–	< 0.001
HAP	0	66 (26.5)	–	< 0.001
VAP	0	6 (2.4)	–	0.01
TB	0	9 (3.6)	–	0.013
AECOPD	0	7 (2.8)	–	0.045
Lung abscess	0	2 (0.8)	–	0.517
History of aspiration, n (%)	4 (2.4)	23 (9.2)	3.9 (1.4–11.1)	0.009
Invasive mechanical ventilation, n (%)	3 (1.8)	85 (34.1)	19.2 (6.2–59.8)	0.000
Immunosuppression, n (%)	32 (18.9)	99 (39.8)	2.1 (1.5–3.0)	0.000
Symptoms, n (%)				
Cough	92 (54.4)	147 (59.0)	1.1 (0.9–1.3)	0.351
Fever	25 (14.2)	111 (44.6)	3.0 (2.0–4.4)	< 0.001
Sputum	59 (34.9)	122 (49.0)	1.4 (1.1–1.8)	0.004
Dyspnea	28 (16.6)	90 (36.1)	2.2 (1.5–3.2)	< 0.001

Hemoptysis	16 (9.5)	15 (6.0)	0.6 (0.3–1.3)	0.187
Any comorbidity, n (%)				
Diabetes	29 (17.2)	61 (24.5)	1.4 (1.0–2.1)	0.073
Hypertension	49 (29.0)	69 (27.7)	1.0 (0.7–1.3)	0.775
Cardiovascular disease	23 (13.6)	56 (22.5)	1.7 (1.1–2.6)	0.023
Malignancy	31 (18.3)	25 (10.0)	0.5 (0.3–0.9)	0.014
Hematological diseases	10 (5.9)	41 (16.5)	2.8 (1.4–5.4)	0.001
Hospital stays, median (IQR), days	13.0 (9.0–16.0)	18.0 (12.0–27.0)	5.0 (3.0–7.0)[a]	< 0.001
Antibiotic use, n (%)	71 (42.0)	232 (93.2)	2.2 (1.9–2.7)	< 0.001
WBC, median (IQR), 10^9/L	6.23 (4.84–7.97)	6.91 (5.15–10.37)	0.7 (0.1–1.7)[a]	0.004
< 4, n (%)	16 (9.8)	35 (14.1)	1.5 (0.8–2.6)	
4–10, n (%)	131 (79.9)	147 (59.3)	0.8 (0.7–0.9)	< 0.001
> 10, n (%)	17 (10.4)	66 (26.6)	2.6 (1.6–4.3)	
Percentage of neutrophils	61.7 (54.2–71.5)	72.0 (61.4–83.3)	10.3 (6.4–14.7)[a]	< 0.001
< 40%, n (%)	4 (2.5)	13 (5.2)	2.2 (0.7–6.6)	
40–75%, n (%)	131 (80.4)	124 (50.0)	0.6 (0.6–0.7)	< 0.001
> 75%, n (%)	28 (17.2)	111 (44.8)	2.7 (1.9–3.9)	
Percentage of lymphocytes	26.5 (18.4–34.1)	16.5 (8.3–26.6)	−10.1 (−13.4 to 6.7)[a]	< 0.001
< 20%, n (%)	48 (29.4)	149 (60.6)	2.1 (1.6–2.7)	
20–50%, n (%)	112 (68.7)	92 (37.4)	0.6 (0.5–0.7)	< 0.001
> 50%, n (%)	3 (1.8)	5 (2.0)	1.1 (0.3–4.7)	
CRP, median (IQR), mg/L	2.5 (2.5–18.7)	25.0 (5.5–85.8)	22.5 (16.1–33.2)[a]	< 0.001
Procalcitonin, median (IQR), ng/mL	0.05 (0.05–0.1)	0.1 (0.059–0.41)	0.05 (0.00–0.06)[a]	< 0.001

Abbreviations: LRTI = lower respiratory tract infection; CAP = community-acquired pneumonia; HAP = hospital-acquired pneumonia; VAP = ventilator-associated pneumonia; TB = pulmonary tuberculosis; AECOPD = acute exacerbation of chronic obstructive pulmonary disease; WBC = white blood cell count; CRP = C-reactive protein; IQR = interquartile range. Age and length of hospital stay were compared using the Mann–Whitney U test; all other variables were analyzed with the chi-square test, continuity correction, or Fisher's exact test. Laboratory results reflect the most recent measurements obtained within seven days before sample collection.

[a] Age, hospital stay duration, WBC count, neutrophil percentage, lymphocyte percentage, CRP, and procalcitonin are presented as median differences with 95% confidence intervals; other variables are expressed as odds ratios with 95% confidence intervals.

Clinical laboratory findings indicated that individuals in the LRTI group had higher white blood cell counts, higher C-reactive protein concentrations, elevated procalcitonin levels, greater antibiotic use, and longer hospital stays (**Table 1**).

Limit of detection

LoD assessment utilized uninfected normal respiratory flora (NRF) BALF specimens spiked in ten replicates with microorganisms (P. aeruginosa, S. aureus, A. fumigatus, CMV, Rhv) across serial tenfold dilutions of defined concentrations. Variations in commensal bacterial background levels influenced LoD values for bacterial targets. The Panel workflow achieved better (lower) LoDs for A. fumigatus and CMV relative to the Meta workflow.

NanoMP characteristics

Every sequencing run included six patient samples plus five negative controls: host depletion (NC1), DNA extraction (NC2), and library preparation (NC3) controls for the Meta workflow; and DNA/RNA extraction (NC2) plus library preparation (NC3) controls for the Panel workflow. The complete workflow required roughly 6.5–7.0 hours (**Figure 1**), encompassing nucleic acid extraction for both processes (2 hours), library construction for both processes (3.5–4.0 hours), and sequencing plus data analysis (approximately 1 hour).

Establishment of the filtering system and performance of the Meta process

A total of 972 species-level records for bacteria and 270 for viruses were utilized to develop and validate XGBoost classification models. In the bacteria category, the XGBoost model achieved an area under the curve (AUC) of 0.915 (95% confidence interval [CI], 0.894–0.916) on the training set and 0.923 (95% CI, 0.897–

0.931) on the test set. For the virus category, AUC values were 0.817 (95% CI, 0.772–0.811) in training and 0.793 (95% CI, 0.726–0.805) in testing (**Figure 2**).

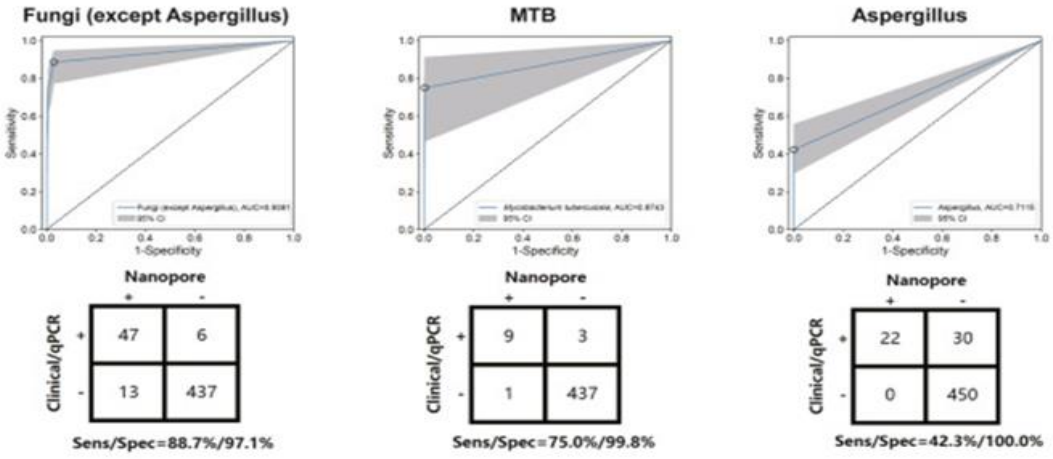
Using the finalized models, overall AUC scores reached 0.916 for bacteria and 0.809 for viruses (**Figure 2**). When benchmarked against clinical and orthogonal validation results, the classifiers demonstrated 84.8% sensitivity and 82.9% specificity for bacteria, and 81.0% sensitivity and 67.2% specificity for viruses (**Figure 2**).

For the fungal group (excluding Aspergillus), MTB group, and Aspergillus group in the Meta workflow, we selected Reads and Ratio_run parameters via correlation analysis. Thresholds were established from ROC curves, yielding AUC values of 0.94, 0.87, and 0.71, respectively (**Figure 2**).

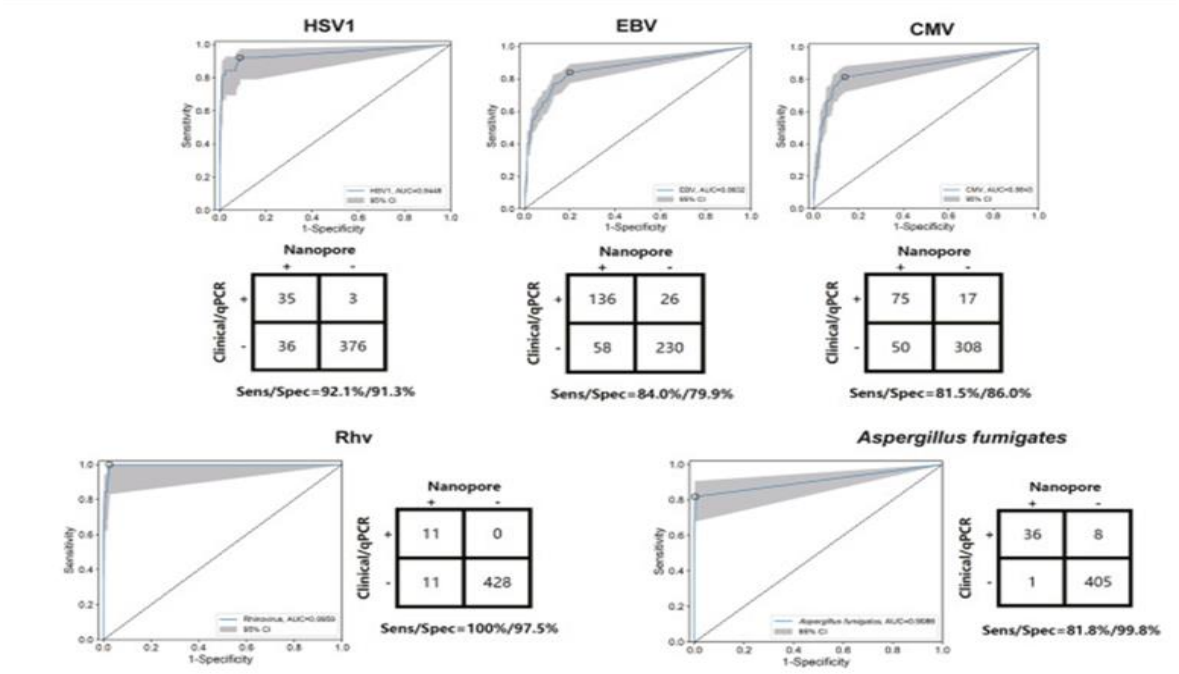
To benchmark performance, ROC analysis using only Reads and Ratio_run metrics produced 71.6% sensitivity and 69.3% specificity for bacteria, and 48.6% sensitivity and 77.1% specificity for viruses. The XGBoost approach provided superior sensitivity for bacteria and improved specificity across both categories (**Figure 2**).

Applying the XGBoost classifiers yielded sensitivity exceeding 80% for both bacterial and viral groups, confirming the value of machine learning for threshold setting. After accounting for organisms that were missed, the bacteria group showed excellent results with 82.9% sensitivity and 82.9% specificity (**Figure 3a**). In contrast, the virus group had more missed detections, yielding 39.4% sensitivity (**Figure 3a**). ROC-based evaluation for other groups revealed 88.7% sensitivity and 97.1% specificity for fungi (excluding Aspergillus), 75.0% sensitivity and 99.8% specificity for MTB, and 42.3% sensitivity for Aspergillus (**Figure 3b**). Species-specific performance metrics in the Meta process are detailed in **Table 2**.

(a)



(b)



(c)

Figure 3. Evaluation of Nanopore sequencing performance using the gold standard and the established scoring criteria. (a) Results for bacterial and viral groups after incorporating organisms that were initially missed by detection. (b) Optimal performance metrics and corresponding contingency tables for the fungal group (excluding Aspergillus), MTB group, and Aspergillus group within the Meta workflow, derived from ROC analysis with 95% confidence intervals. (c) Optimal performance metrics and contingency tables for HSV1, EBV, CMV, Rhinovirus, and Aspergillus fumigatus in the Panel workflow, obtained via ROC analysis with 95% confidence intervals.

Abbreviations: MTB = Mycobacterium tuberculosis; HSV1 = human alphaherpesvirus 1; EBV = Epstein–Barr virus (human gammaherpesvirus 4); CMV = human cytomegalovirus (human betaherpesvirus 5); ROC = receiver operating characteristic.

Table 2. The detection performance of each species.

Group	Species	Meta		Panel		Meta + Panel	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Viral group	Human herpesvirus 4	33.3%	95.1%	84.0%	79.9%	93.2%	79.9%
	Human alphaherpesvirus 1	65.8%	99.0%	92.1%	91.3%	94.7%	91.3%
	Human herpesvirus 5	39.1%	93.3%	81.5%	86.0%	91.3%	86.0%
	Rhinovirus	NA	NA	100.0%	97.5%	NA	NA
Bacteria group	*Streptococcus anginosus*	59.0%	99.3%	NA	NA	NA	NA
	Stenotrophomonas maltophilia	74.4%	97.3%	NA	NA	NA	NA
	Enterococcus faecium	100.0%	96.7%	NA	NA	NA	NA
	Tropheryma whipplei	81.8%	100.0%	NA	NA	NA	NA
	Mycoplasma hominis	100.0%	99.8%	NA	NA	NA	NA
	Enterococcus faecalis	100.0%	97.5%	NA	NA	NA	NA
	Corynebacterium striatum	95.1%	96.1%	NA	NA	NA	NA
	Klebsiella pneumoniae	83.3%	97.3%	NA	NA	NA	NA
	Streptococcus pneumoniae	75.0%	95.9%	NA	NA	NA	NA
	Staphylococcus aureus	90.9%	98.1%	NA	NA	NA	NA

	Pseudomonas aeruginosa	81.8%	96.6%	NA	NA	NA	NA
	Acinetobacter baumannii	82.4%	99.3%	NA	NA	NA	NA
MTB group	*Mycobacterium tuberculosis*	75.0%	99.8%	NA	NA	NA	NA
Fungi group (except *Aspergillus*)	*Candida albicans*	93.8%	97.5%	NA	NA	NA	NA
	Pneumocystis jirovecii	40.0%	99.3%	NA	NA	NA	NA
***Aspergillus* group**	*Aspergillus fumigatus*	34.1%	100.0%	81.8%	99.8%	95.5%	99.8%
	Aspergillus niger	87.5%	100.0%	NA	NA	NA	NA

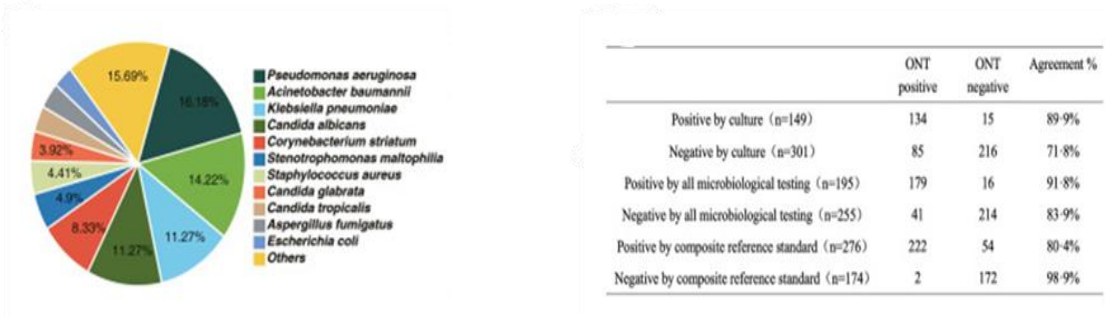
Abbreviation: NA= not applicable.

Enhancement of detection sensitivity for selected pathogens in the panel workflow

Using correlation testing and ROC curve assessment of the most informative indicators for EBV, CMV, HSV1, A. fumigatus, and Rhv within the Panel workflow, the resulting AUC scores were 0.86 for EBV, 0.86 for CMV, 0.94 for HSV1, 0.91 for A. fumigatus, and 0.99 for Rhv (**Figure 2**). Achieved sensitivity and specificity stood at 92.1% and 91.3% for HSV1, 84.0% and 79.9% for EBV, 81.5% and 86.0% for CMV, 81.8% and 99.8% for A. fumigatus, and 100% and 97.5% for Rhv (**Figure 3c**, **Table 2**).

Diagnostic performance of the Nanopore platform according to clinical assessment

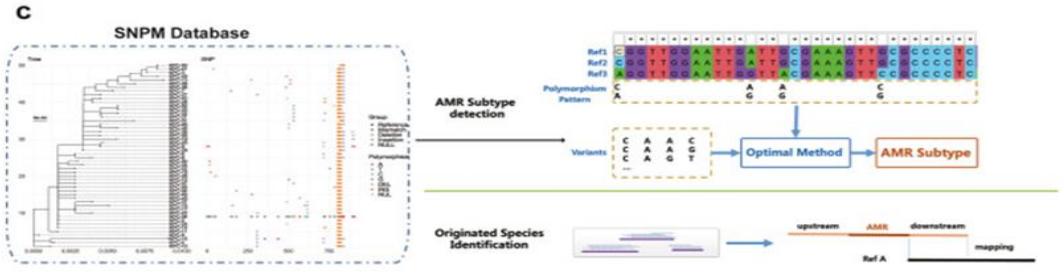
Out of the 450 enrolled specimens, culture methods alone identified positive results in 149 cases (33.1%) (**Figure 4a**). This figure increased to 195 samples (43.3%) when additional microbiological assays were included alongside culture (**Figure 4b**). Application of the Nanopore method flagged 224 samples (49.8%) as containing Definite or Probable causative agents. In comparison with culture-only outcomes, Nanopore sequencing provided 89.9% clinical sensitivity and 71.8% clinical specificity. When the extra 45 samples positive solely by other lab tests were factored in, clinical sensitivity reached 91.8% and specificity 83.9% versus the complete microbiological testing panel. Against the overall composite reference standard, the Nanopore approach delivered 80.2% clinical sensitivity paired with 98.8% clinical specificity (**Figure 4b**).



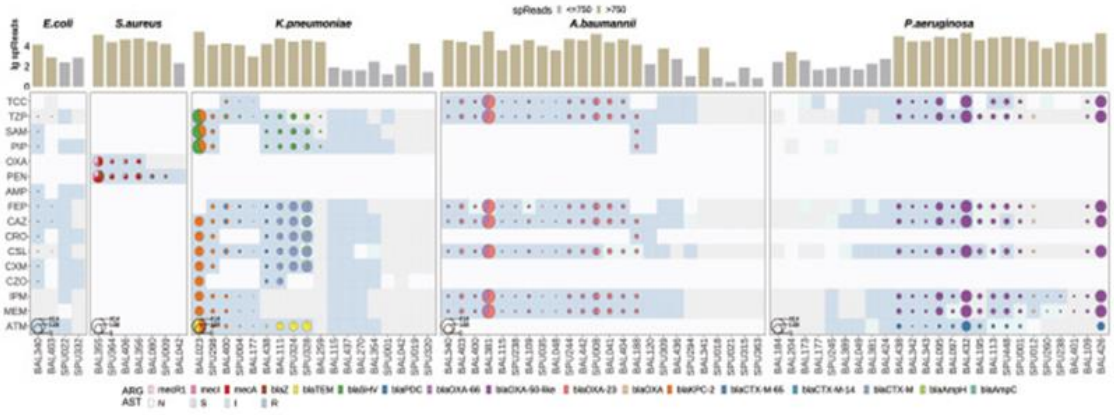
(a)

	ONT positive	ONT negative	Agreement %
Positive by culture (n=149)	134	15	89·9%
Negative by culture (n=301)	85	216	71·8%
Positive by all microbiological testing (n=195)	179	16	91·8%
Negative by all microbiological testing (n=255)	41	214	83·9%
Positive by composite reference standard (n=276)	222	54	80·4%
Negative by composite reference standard (n=174)	2	172	98·9%

(b)



(c)



(d)

Figure 4. Overview of pathogen identification and AMR gene detection performance using Nanopore sequencing evaluated through clinical assessment. (a) Breakdown of results from standard culture testing. (b) Diagnostic accuracy of Nanopore sequencing compared with culture alone, combined microbiological assays, and the composite reference standard. (c) Process for detecting AMR genes and linking them to their source organisms. (d) Results of AMR gene detection via Nanopore sequencing.

Abbreviations: SNPM = single nucleotide polymorphism matrix; AMR = antimicrobial resistance; TCC = Ticarcillin/Clavulanic acid; TZP = Piperacillin/Tazobactam; SAM = Ampicillin/Sulbactam; PIP = Piperacillin; OXA = Oxacillin; PEN = Penicillin G; AMP = Ampicillin; FEP = Cefepime; CAZ = Ceftazidime; CRO = Ceftriaxone; CSL = Cefoperazone/Sulbactam; CXM = Cefuroxime; CZO = Cefazolin; IPM = Imipenem; MEM = Meropenem; ATM = Aztreonam

Antimicrobial resistance gene detection

Prior investigations have employed Nanopore sequencing to detect genes conferring antibiotic resistance [4, 14-16]. In this work, an SNPM database was built to identify AMR genes along with their specific subtypes, with the advantage of long reads enabling direct linkage to the responsible microorganisms (**Figure 4c).** Focus was placed on frequently isolated and clinically burdensome cultured pathogens — P. aeruginosa, A. baumannii, K. pneumoniae, and S. aureus (**Figure 4a**) — for AMR analysis. As illustrated in **Figure 4d**, blaOXA variants were found in 16 P. aeruginosa samples (15 blaOXA-50-like and 1 blaOXA), 15 A. baumannii samples (12 blaOXA-66, 12 blaOXA-23, and 4 blaOXA), four mecA-positive S. aureus samples, and five blaKPC-2-positive K. Concordance with clinical antimicrobial susceptibility testing reached 78.6% for carbapenem-resistant P. aeruginosa, 83.3% for A. baumannii, 62.5% for K. pneumoniae, and 100% for methicillin-resistant S. aureus. These outcomes highlight the strong capability of Nanopore sequencing to detect and correctly attribute resistance genes to their host bacteria in real clinical specimens.

This prospective multicenter investigation analyzed 450 respiratory specimens to assess the practical value of the Nanopore platform in identifying infectious agents. A dual-barcoding NanoMP strategy was implemented to encompass both DNA and RNA pathogens. This included host DNA removal in the Meta workflow and targeted amplification of priority pathogens in the Panel workflow. A robust filtering framework incorporating machine learning and ROC analysis was created and applied to clinical samples. Performance for pathogen detection and AMR gene characterization using the SNPM database was evaluated against clinical outcomes. Earlier research from our group noted that Nanopore sequencing generated fewer reads than Illumina sequencing when applied to lung biopsy tissues [17], resulting in suboptimal detection rates. Another report indicated that primer-based enrichment could boost viral detection on the Nanopore system [7]. To address these constraints, the current study combined targeted amplification with unbiased metagenomic sequencing. Given the frequent occurrence of certain viruses and fungi, EBV, CMV, HSV1, A. fumigatus, and Rhv were chosen to test the approach. In the Meta workflow, strong results were achieved for bacteria (82.9% sensitivity) and

fungi excluding Aspergillus (88.7% sensitivity) thanks to effective host DNA depletion during sample processing. However, viral detection was lower (39.4% sensitivity), likely due to host depletion and potential loss of fragile, small, lightweight viral particles during centrifugation steps. Aspergillus detection was also limited (42.3% sensitivity), possibly because of incomplete cell wall breakdown during extraction given their robust structure. In the Panel workflow, PCR amplification markedly enriched the selected targets. RNA viruses, for instance, reached 100% sensitivity post-amplification. These findings confirm that targeted amplification substantially improves detection of challenging organisms. The relatively low number of respiratory RNA virus cases may reflect widespread mask usage during the COVID-19 period (specimens collected from January 28 to June 28, 2021). Future panel expansions could incorporate additional difficult-to-detect pathogens and RNA viruses.

In metagenomic next-generation sequencing, defining appropriate thresholds is essential for accurate pathogen identification. Various studies have incorporated metrics such as reads per million, environmental contamination markers, library amplification effects, and mapping characteristics (unique versus multiple) to build filtering systems [5, 8, 18-20]. However, cut-off determination methods have varied widely across publications. Developing standardized approaches remains important. Here, all sequencing-derived features were gathered and analyzed with the XGBoost machine learning algorithm to rank their contribution. This resulted in sensitivity and specificity exceeding 80% across taxa, suggesting the model offers a versatile tool adaptable to other mNGS applications. Previous Nanopore studies primarily addressed bacteria only [6] or bacteria plus fungi [5], without comprehensive coverage of viruses. Thus, this large multicenter cohort of 450 samples evaluating bacteria, viruses, fungi, and associated AMR genes provides a more representative view of real-world clinical performance.

For AMR gene detection, three carbapenemase-producing isolates (in P. aeruginosa, A. baumannii, and K. pneumoniae) showed discrepancies with phenotypic susceptibility results. Nonetheless, samples carrying resistance genes had more host-organism reads than those lacking such genes. The method successfully differentiated multiple AMR genes and assigned them to their correct source organisms even in mixed-infection cases. Spearman analysis showed strong positive correlations between organism abundance and linked AMR gene reads, warranting further study.

To ensure consistency throughout the extended study period, we retained the original basecalling software version (Guppy v3.2.10) despite the availability of newer releases (e.g., Guppy v6.3.8). Comparative analysis of 17 samples processed with both versions revealed no meaningful differences in pathogen detection or final clinical interpretation . This demonstrates the stability and reliability of the overall workflow.

Several limitations should be noted. The Panel workflow included only a limited set of pathogens for validation; future work should include a broader range of clinically challenging organisms. Most species lacked commercially available IVD kits for gold-standard confirmation, and orthogonal validation was restricted to 21 representative species. AMR analysis focused solely on acquired resistance genes and did not examine chromosomal mutations. Additionally, detected organisms were not always proven to be the definitive cause of disease, and many potential pathogens appear in respiratory samples from asymptomatic individuals.

Conclusion

In conclusion, Nanopore sequencing effectively broadens the diagnostic capabilities of traditional methods for lower respiratory tract infections. The NanoMP approach successfully increased sensitivity for specific pathogens through targeted amplification. The machine learning-based framework for establishing detection thresholds shows promise for application across various sample types and in-house mNGS platforms. Overall, the study confirms strong clinical performance of Nanopore sequencing for LRTI diagnosis and demonstrates reliable identification and host attribution of key AMR genes leveraging long-read sequencing.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Corrado RE, Lee D, Lucero DE, Varma JK, Vora NM. Burden of adult community-acquired, health-care-associated, hospital-acquired, and ventilator-associated pneumonia: New York City, 2010 to 2014. Chest. 2017;152(5):930-42.
2. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20(6):341-55.
3. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Tabernero L, Wren BW, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434-9.
4. Charalampous T, Kay GL, Richardson H, Alcolea-Medina A, Pickard D, Gharbia S, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783-92.
5. Gu W, Deng X, Lee M, Sucu YD, Arevalo S, Stryke D, et al. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. Nat Med. 2021;27(1):115-24.
6. Mu SR, Hu L, Zhang Y, Wang Y, Chen X, Li J, et al. Prospective evaluation of a rapid clinical metagenomics test for bacterial pneumonia. Front Cell Infect Microbiol. 2021;11:684965.
7. Deng X, Achari A, Federman S, Yu G, Somasekar S, Bártolo I, et al. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. Nat Microbiol. 2020;5(3):443-54.
8. Zinter MS, Dvorak CC, Mayday MY, Iwanaga K, Ly NP, McGarry ME, et al. Pulmonary metagenomic sequencing suggests missed infections in immunocompromised children. Clin Infect Dis. 2019;68(11):1847-55.
9. Shen JH, Huang KY, Chao-Yu C, Chen CJ, Lin TY, Huang YC. Seroprevalence of herpes simplex virus type 1 and 2 in Taiwan and risk factor analysis, 2007. PLoS One. 2015;10(8):e0134178.
10. Xiong G, Zhang B, Huang MY, Zhou Y, Ye S, Chen Y, et al. Epstein-Barr virus (EBV) infection in Chinese children: a retrospective study of age-specific prevalence. PLoS One. 2014;9(6):e99857.
11. Griffiths P, Baraniak I, Reeves M. The pathogenesis of human cytomegalovirus. J Pathol. 2015;235(2):288-97.
12. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, 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. Am J Respir Crit Care Med. 2019;200(7):e45-67.
13. Kalil AC, Metersky ML, Klompas M, Muscedere J, Sweeney DA, Palmer LB, et al. Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. Clin Infect Dis. 2016;63(5):e61-111.
14. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwansa J, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol. 2015;33(3):296-300.
15. Brinda K, Callendrello A, Ma KC, MacFadden DR, Charalampous T, Lee RS, et al. Rapid inference of antibiotic resistance and susceptibility by genomic neighbour typing. Nat Microbiol. 2020;5(3):455-64.
16. Charalampous T, Alcolea-Medina A, Snell LB, Williams TGS, Batra R, Alder C, et al. Evaluating the potential for respiratory metagenomics to improve treatment of secondary infection and detection of nosocomial transmission on expanded COVID-19 intensive care units. Genome Med. 2021;13(1):182.
17. Guo Y, Li H, Chen H, Liu Y, Wang S, Zhang L, et al. Metagenomic next-generation sequencing to identify pathogens and cancer in lung biopsy tissue. eBioMedicine. 2021;73:103639.
18. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res. 2019;29(5):831-42.
19. Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat Microbiol. 2019;4(4):663-74.
20. Jing C, Chen H, Liang Y, Zhong L, Li Y, Wang Z, et al. Clinical evaluation of an improved metagenomic next-generation sequencing test for the diagnosis of bloodstream infections. Clin Chem. 2021;67(8):1133-43.