Annals of Clinical Microbiology and Antimicrobials

RESEARCH

Open Access

Host DNA depletion assisted metagenomic sequencing of bronchoalveolar lavage fluids for diagnosis of pulmonary tuberculosis

Jinfeng Yuan[1†], Liping Ma[2†], Juan Du[3†], Hailin Sun[4†], Shanshan Li[1], Gang Zhou[5], Guanhua Rao[5], Fengshuo Sun[5], Wangyang Chen[5], Hui Miao[5], Dan Tian[6], Changhao Cheng[3], Yan Wang[4], Liang Li[7*], Lifeng Li[5*] and Yu Pang[1*]

Abstract

Metagenomic next-generation sequencing (mNGS) has greatly improved our understanding of pathogens in infectious diseases such as pulmonary tuberculosis (PTB). However, high human DNA background (> 95%) impedes the detection sensitivity of mNGS in identifying intracellular *Mycobacterium tuberculosis* (MTB), posing a pressing challenge for MTB diagnosis. Therefore, there is an urgent need to improve MTB diagnosis performance in PTB patients. In this study, we optimized mNGS method for diagnosis of PTB. This led to the development of the host DNA depletion assisted mNGS (HDA-mNGS) technique, which we compared with conventional mNGS and the host DNA depletion-assisted Nanopore sequencing (HDA-Nanopore) in diagnostic performance. We collected 105 bronchoalveolar lavage fluid (BALF) samples from suspected PTB patients across three medical centers to assess the clinical performance of these methods. The results of our study showed that HDA-mNGS had the highest sensitivity (72.0%) and accuracy (74.5%) in PTB detection. This was significantly higher compared to mNGS (51.2%, 58.2%) and HDA-Nanopore (58.5%, 62.2%). Furthermore, HDA-mNGS provided an increased coverage of the MTB genome by up to 16-fold. Antibiotic resistance gene analysis indicated that HDA-mNGS could provide increased depth to the detection of Antimicrobial resistance (AMR) locus more effectively. These findings indicate that HDA-mNGS can significantly improve the clinical performance of PTB diagnosis for BALF samples, offering great potential in managing antibiotic resistance in PTB patients.

Keywords Pulmonary tuberculosis, Drug resistance, Host DNA depletion, Metagenomic next-generation sequencing, Bronchoalveolar lavage fluid

†Jinfeng Yuan, Liping Ma, Juan Du and Hailin Sun contributed equally to this work.

*Correspondence:
Liang Li
Liliang69@vip.sina.com
Lifeng Li
lf.li@genskey.com
Yu Pang
pangyupound@163.com
[1] Department of Bacteriology and Immunology, Beijing Chest Hospital, Capital Medical University/Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China
[2] Department of Tuberculosis, Beijing Chest Hospital, Capital Medical University/Beijing Tuberculosis & Thoracic Tumor Research Institute, Beijing, China
[3] Department of Tuberculosis, Wuhan Pulmonary Hospital, Wuhan, China
[4] Department of Tuberculosis, Ordos Second People's Hospital, Ordos, China
[5] Genskey Medical Technology Co., Ltd, A212 Innovation Building, Changping Life Garden, Beijing, China
[6] Department of Tuberculosis Prevention, Wuhan Pulmonary Hospital, Wuhan, Hubei, China
[7] Clinical Center On Tuberculosis Control, Beijing Chest Hospital, Capital Medical University, Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China

Introduction

In 2023, there were over 10.8 million reported cases of tuberculosis (TB) and 1.25 million TB-related deaths. TB is returning as the leading cause of death from a single infectious disease worldwide, surpassing COVID-19 [1]. Traditional diagnostic strategies used by most microbiology laboratories rely on tests based on *Mycobacterium tuberculosis* (MTB) culture growth and isolation, as well as costly and poorly performing MTB detection tests based on MTB-specific antibodies or antigens, and molecular methods based on MTB complex-associated nucleic acids detection, such as PCR or GeneXpert MTB/RIF analysis. However, these current molecular diagnostic methods only target a few pathogen-related genes [2] and cannot provide comprehensive detection and recommendations for clinicians. Therefore, there is a urgent need for more rapid, reliable and affordable high-throughput tools to address the clinical challenges in MTB detection [3].

Diagnosis of infectious disease requires assays that can provide a comprehensive analysis of pathogen DNA or RNA, which could be resolved via the introduction of metagenomic next-generation sequencing (mNGS) technologies. Consequently, the use of mNGS methods has greatly broadened our clinical knowledge of effects of pathogenic microorganisms on the human microbiome and microbiota and enhanced our understanding of pathogenic antimicrobial resistance mechanisms [2]. Currently, metagenomic sequencing based on next-generation and third-generation sequencing platforms are widely used in clinical infectious disease diagnosis and surveillance, due to their demonstrated performance in successfully detecting MTB in various types of specimens [4]. However, the slow growth of MTB and its facultative intracellular growth characteristics make it challenging to recover intact MTB from host cells, which is necessary to maximize MTB DNA yield while minimizing host DNA contamination. Therefore, we aim to overcome this obstacle to improve the performance of conventional mNGS and maximize the full clinical potential of these technologies.

To tackle this problem, saponin-based host DNA depletion-assisted (HDA) metagenomic sequencing was developed to effectively remove large quantities of human DNA from samples, thus reducing high host-related metagenomic sequencing output read rates that mask genuine pathogenic signals by increasing base-calling error rates [5, 6]. Particularly, HDA-mNGS has been applied to clinical *Mycobacterium* spp., detection, resulting in improved diagnostic performance and data quality when used to detect mycobacteria in sputum samples [6]. For patients unable to provide adequate sputum samples, BALF samples can serve as alternative specimens to diagnose active TB [7, 8]. In a recent study, it was demonstrated that BALF had the same detection performance as GeneXpert and higher sensitivity to MTB detection compared to sputum samples [9]. Although BALF offers the potential to improve sensitivity for the diagnosis of pulmonary TB, successful mNGS-based detection of MTB in BALF samples has been limited due to the aforementioned challenges related to host DNA contamination coupled with extremely low amounts of bacterial DNA in BALF samples.

In this study, we focused on the BALF samples to leverage the benefit of bronchoscopy and tried to refine metagenomics methods to generate HDA-mNGS. We then compared the performance of these optimized approaches with conventional mNGS and HDA-Nanopore, with the ultimate goal of enhancing the diagnostic performance of PTB in BALF samples.

Methods

Study patients

We retrospectively reviewed 105 bronchoalveolar lavage fluid (BALF) samples obtained from patients with suspected PTB who sought care at Beijing Chest Hospital, Wuhan Pulmonary Hospital, and Ordos Second People's Hospital between June 2021 and April 2022, and followed their progress for 3 months. The inclusion criteria required patients to be suspected of tuberculosis based on definitive imaging findings and clinical symptoms of tuberculosis. Additionally, all patients were required to have valid results for microbiological culture, Xpert MTB/RIF, mNGS, HDA-mNGS, and HDA-nanopore pathogen detection. During the diagnostic process, some patients also underwent sputum smear and T-SPOT tests as part of the diagnostic criteria when necessary. Exclusion criteria included incomplete basic information or pathogen detection data, immunosuppressed status, lack of signed informed consent, or prior anti-tuberculosis treatment.

Based on the final diagnosis, these patients were divided into 2 groups: PTB and non-PTB (Additional file 1: Fig. S1). The attending physicians made the final clinical diagnosis at least 3 months after collecting the samples using the China Clinical Treatment Guide for Tuberculosis [10] and other clinical criteria. Patients diagnosed with PTB were classified into the definite PTB group and the clinically diagnosed PTB group. Patients with positive results for MTBC from microbiological culture or Xpert MTB/RIF testing were categorized as the definite PTB group. Additionally, according to the guidelines, patients diagnosed with tuberculous bronchitis or tuberculous pleuritis based on imaging and clinical evaluation were also included in the definite PTB group. For suspected patients lacking microbiological evidence,

those exhibiting clinical symptoms consistent with PTB and positive results from tests such as the tuberculin skin test, interferon-gamma release assay, or *M. tuberculosis* antibody testing were classified into the clinically diagnosed PTB group. The non-PTB category comprised patients with malignancies, non-infectious inflammatory diseases, non-tuberculous infections, and other conditions.

BALF collection

BALF used for MTB detection was collected bronchoscopically by a cardiothoracic surgeon after informed written consent was obtained from patients or caregivers. For each patient, sterile normal saline was injected into a subsegment of the lung followed by suction and collection of BALF for further analysis.

Mycobacterial culture

One ml of BALF collected from each patient and 0.8 ml of MGIT additive (Becton Dickinson [BD], NJ, USA) were mixed, and the suspension was cultured for 6 weeks according to the BACTEC MGIT 960 (Becton Dickinson [BD], NJ, USA) instructions. Positive samples were defined based on the growth of MTBC, as indicated by an increase in fluorescence within the MGIT tubes, signaling oxygen consumption by the bacteria. For successfully cultured samples, Mycobacterium species identification was performed using an mpt64 antigen detection test with standard reagents (SD Bioline TB Ag MPT64 Rapid Test, Abbott, Illinois, USA).

GeneXpert MTB/RIF assay

Each BALF sample was evenly mixed with a sample reagent buffer (containing sodium hydroxide and isopropanol) in a 2:1 ratio and vortexed for 15–30 s. Next, each sample was incubated at room temperature for 15 min and then the treated sample and additional reagents were added to the Xpert-Cartridge. After 2 h, the GeneXpert system generated a TB test report and an MTB drug resistance report.

Metagenomic sequencing pipeline

Sample pre-processing and DNA extraction

For HDA-mNGS pipeline, Sputasol (Oxoid) was added to each BALF sample followed by incubation at room temperature for 2–5 min. For HDA-Nanopore pipeline, BALF samples were first treated with sputasol on a vortexer for 10 min at 42 °C. Subsequently, the treated samples from HDA-mNGS and HDA-Nanopore were centrifuged, and then the microbial cells were resuspended and lysed according to the procedures described previously [5]. The conventional mNGS workflow did not include the host depletion process, and samples were directly used for nucleic acid extraction. DNA was extracted for mNGS, HDA-mNGS and HDA-Nanopore assays using the TIANamp Micro DNA Kit (TIANGEN Biotech Co., Ltd., Beijing, China) and DNA quality control was determined by measuring DNA concentration using a Qubit 4.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

Library construction and sequencing

For MGI sequencing, DNA libraries were constructed using the VAHTS® Universal Plus DNA Library Prep Kit for MGI (Vazyme). Sequencing libraries were pooled and sequenced on a MGISEQ-2000 sequencer using a single-ended 50-cycle kit. The Rapid PCR Barcoding Kit (SQK-RPB004, ONT) was used for nanopore sequencing according to the manufacturer's protocol. Briefly, 6 µL of extracted DNA (no more than 10 ng) was fragmented and ligated with unique barcodes. PCR amplification was performed using a 100 µL mix consisting of barcoded DNA, PrimeSTAR® GXL DNA Polymerase (TaKaRa) and other PCR amplification reagents on a PCR thermocycler for 35 cycles. DNA libraries were pooled in equal mass and the sequencing library pool was quantified using Qubit 4.0 fluorometer (Thermo Fisher). The DNA was then sequenced using R9.4.1 flow cells on a GridION X5 device (Oxford Nanopore Technologies). Negative control (NC) samples were included in each run to monitor background microbial DNA contamination.

Bioinformatic analyses

mNGS and HDA-mNGS*:* Raw data underwent trimming to remove adapter sequences and low-quality bases, and sequences with read lengths of < 35 bp were filtered out. Data for downstream analysis were obtained based on quality control with Q30 > 85%. High-quality reads that aligned with human host DNA sequences were removed using Bowtie 2 [11]. The remaining sequence data were aligned with a custom-built microorganism genome database. The pathogen genomes in the database were sourced from the RefSeq website (https://www.ncbi.nlm.nih.gov/refseq/) and include representative genome sequences of 4,491 bacterial species, 556 fungal species, 175 parasite species, 114 mycobacterial species, 157 mycoplasma/chlamydia species, 5735 viral species, and 172 archaea species. Positive criteria for the inclusion of the mNGS and HDA-mNGS sequence data in bioinformatic analysis were as follows: (1) mNGS: unique mapped reads ≥ 1, HDA-mNGS: unique mapped reads ≥ 3; (2) $RP20M \geq RP20M_{maximum}{}^{*}5/100{,}000$, where RP20M represents the number of unique mapped reads per 20 million reads and $RP20M_{maximum}$ represents the maximum number of RP20M reads obtained for the same batch; (3) $RP20M/RP20M_{NC} \geq 2$ or $RP20M_{NC} = 0$,

where $RP20M_{NC}$ represents the RP20M value obtained for the negative control for the same batch. The establishment of the positive reporting threshold was based on internal testing results of the NGS product used.

HDA-Nanopore*:* Reads with lengths < 200 bp and quality values < 7 were removed using NanoPlot and NanoFilt [12]. High-quality reads were aligned with human DNA sequences then host DNA sequences were removed using Minimap2 [13]. Subsequently, the remaining sequences were aligned with the microorganism genome database using Minimap2. A sample was considered positive for microbial DNA if it contained a number of unique mapped reads that was ≥ 3. Conversely, the sample was considered negative for microbial DNA.

Statistical analysis

In our study, McNemar test were used to compare the sensitivity and specificity of tuberculosis diagnostic methods, the Spearman test was applied to explore the correlation between NGS reads and Xpert CT values, and the Wilcoxon test was employed to assess the difference in pathogen reads detected by HDA-mNGS and conventional mNGS [14, 15]. And 95% confidence intervals of diagnostic performance were calculated using an online tool (http://www.vassarstats.net/clin1.html). GraphPad Prism 9 (GraphPad Software Inc., San Diego, CA) and R software (version 4.1.1) was used to perform statistical analyses and generate graphs.

Results

Patient characteristics

Between June 1, 2021, and May 1, 2022, a total of 105 patients with clinical or radiological suspicion of active pulmonary TB were screened from three hospitals for review and prospective enrolment in this study. Tuberculous tests, including culture, Xpert, mNGS, HDA-mNGS and HDA-Nanopore, were conducted on BALF samples from 105 patients (Fig. 1). Out of 105 samples, 7 patients who met our exclusion criteria were excluded for analysis. The average age of the remaining 98 patients (57.1% male) was 46 (Table 1). Of these patients, 28.6% (28 patients) had underlying pulmonary disease, such as chronic obstructive pulmonary disease (COPD), bronchiectasis, and lung cancer, while 24.5% (24 patients) had underlying extrapulmonary disease. Acid-fast bacillus (AFB) smear by microscopy was positive in 17.4% (17 patients) of the patients.

Of the 98 patients, 83.7% (82 patients) were finally diagnosed with active pulmonary TB infection, including 50 patients with definitive results of PTB pathogen detection and 32 patients with clinical diagnoses (Additional file 1: Fig. S1). The remaining 16.3% (16 patients) were diagnosed with non-TB disease, including one case of nontuberculous mycobacteria (NTM), three cases of common bacterial infection, two cases of viral infection, three cases of fungal infection, one case of acute exacerbation of COPD (AECOPD), one case of diffuse panbronchiolitis (DPB), one case of *Mycoplasma pneumoniae* infection, and two cases of malignant lung disease (Additional file 2: Table S1).

HDA-mNGS increases both microbial and TB reads in BALF samples

After quality control and exclusion of non-biologically significant signals, 97.2% of the 98 conventional mNGS samples were identified as sequences of human origin. Following the removal of host reads, only 0.24% of the sequencing data remained, attributed to microbial genomes based on the microorganism genome database (unpublished). As a contrast, following the host DNA depletion procedure, both HDA-mNGS and HDA-Nanopore exhibited marked improvements. The average percentage of host reads decreased to 39.7% and 27.4%, respectively, while the average percentage of microbial reads increased significantly to 18.8% and 61.1%, respectively. This underscores the enhanced efficacy of the host DNA depletion method (Fig. 2A, B).

We next sought to determine whether the increase of microbial reads could proportionally yield an increase in TB reads. To this end, we utilize the Ct values from Xpert tests to determine the correlation between Ct values and the read numbers obtained through metagenomic methods. Among the 39 Xpert-positive samples, Ct values were obtained for 37. A negative correlation was identified between the Ct value and the normalized read number of MTBC using three metagenomic methods (Spearman correlation index: vs. mNGS $R^2 = 0.47$, $p < 0.001$; vs. HDA-mNGS $R^2 = 0.54$, $p < 0.001$; vs. HDA-Nanopore $R^2 = 0.58$, $p < 0.001$) (Fig. 2C).

Following the cutoffs from a previous study [14], samples were categorized into three groups based on Ct values (Medium: Ct < 22, Low: Ct 22–28, Very low: Ct > 28). The MTBC detection performance exhibited a gradual decline with the transition of Ct values from the medium to the very low group (Fig. 2D). In the three groups HDA-mNGS outperformed mNGS in the MTBC reads (Wilcoxon test, $p < 0.001$), indicating superior pathogen detection performance. In Ct < 22 and Ct 22–28 groups, HDA-mNGS detected a high level of MTBC reads in all the 27 samples, whereas mNGS obtained reads in only 19 samples. In situation with poor sample quality (Ct > 28), HDA-mNGS still performed robustly, detecting reads in 9 out of 11 samples, compared to mNGS (4 out of 11) and HDA-Nanopore (6 out of 11). Furthermore, HDA methods (HDA-mNGS and HDA-Nanopore) significantly increased MTBC genome coverage in the medium and

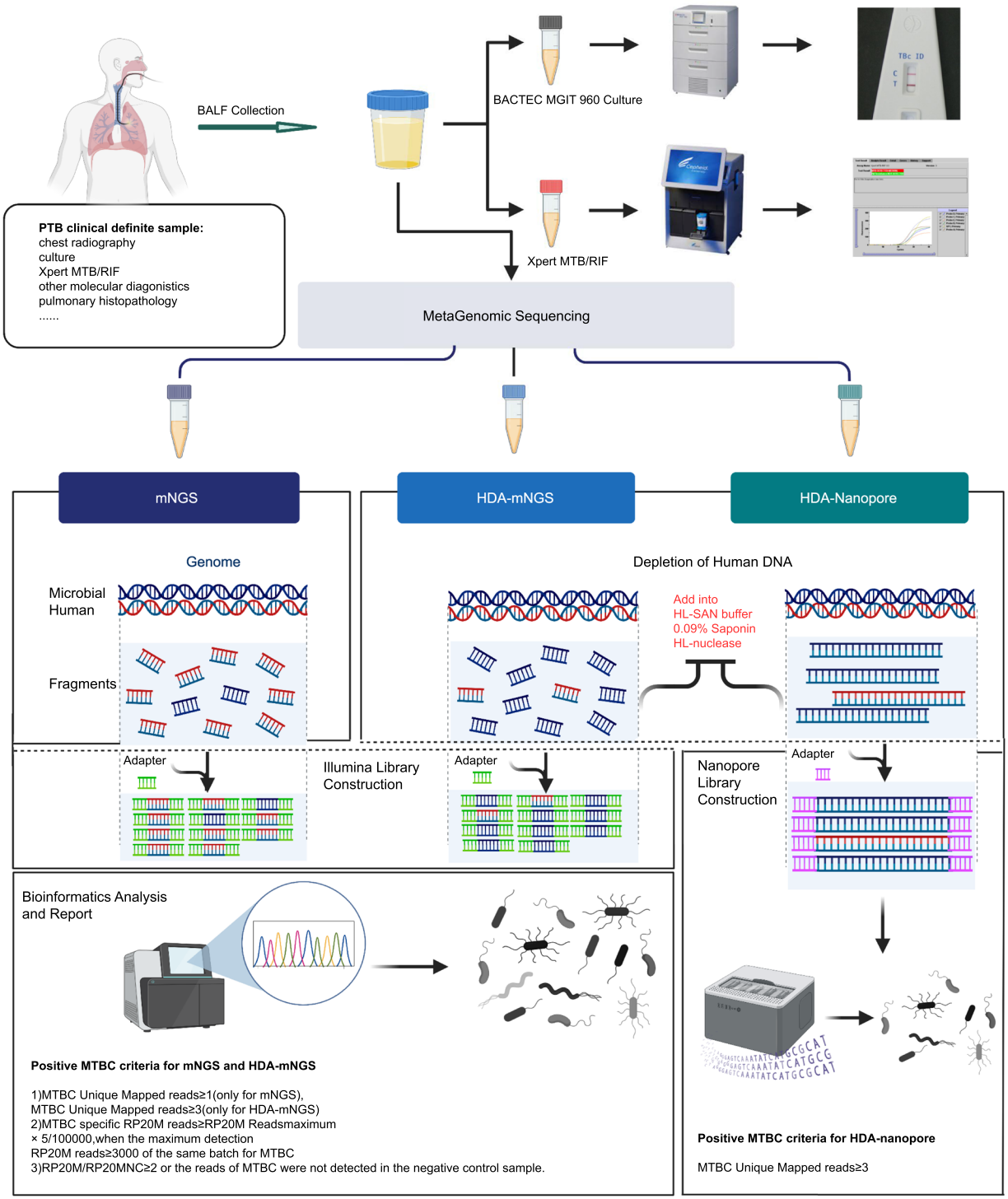


Fig. 1 Flowchart showing steps of MTB-detection methods used in this study. BALF samples collected from patients were each aliquoted into five tubes that were subjected to separate tests that included MGIT 960 culture (one sample), Xpert (one sample) and three mNGS tests (three samples) that included conventional mNGS (one sample), HDA-mNGS (one sample) and HDA-Nanopore (one sample). The conventional mNGS library was constructed without depletion of human DNA, whereas MGI and Nanopore libraries were constructed after human DNA was removed from the HDA-mNGS and HDA-Nanopore pipelines. Sequencing, bioinformatic analysis and reporting PTB diagnostic results were interpreted based on standard TB cut-offs. The mean sequence number outputs obtained using conventional mNGS, HDA-mNGS and HDA-Nanopore were 56.84M, 28.16M and 103.74K. Created in BioRender [16]

Table 1 Demographic and Clinical Characteristics of the 98 Patients

Characteristics	Value
Age, mean (range), years	46.36 (13–75)
Distribution—no. (%)	
13–18 yr	6 (6.12)
19–40 yr	31 (31.63)
41–60 yr	32 (32.65)
> 60 yr	29 (29.59)
Sex, male, no. (%)	56 (57.14)
Laboratory parameters	
WBC, mean(range), 10^9/L	6.53 (2.64–17.02)
Neutrophil (range), 10^9/L	4.31 (1.02–13.29)
Lymphocyte(range), 10^9/L	1.57 (0.62–383)
CRP (range), mg/L	18.69 (0.02–333.65)
PCT (range), ng/L	0.07 (0.02–0.56)
Underlying pulmonary disease	
COPD, no. (%)	10 (10.20)
Bronchiectasis, no. (%)	20 (20.41)
Lung cancer, no. (%)	4 (4.08)
Previous history of tuberculosis, no. (%)	13 (13.27)
Underlying extrapulmonary disease	
Diabetes, no (%)	19 (19.39)
Hypertension, no (%)	14 (14.29)
Institution—no. (%)	
Beijing Chest Hospital	37 (37.76)
Wuhan Pulmonary Hospital	46 (46.94)
Ordos Second People's Hospital	15 (15.31)

WBC, White Blood Cell; CRP, C-Reactive Protein; PCT, Procalcitonin; COPD, Chronic obstructive pulmonary disease. Numbers in the brackets indicate the percentage of the total number of patients or the range of the corresponding parameters

low groups, and with a nonsignificant increase in the very low group.

HDA-mNGS improves PTB diagnostic performance in BALF samples

We evaluated the diagnostic performance of Culture, Xpert, mNGS, HDA-mNGS, and HDA-Nanopore for 98 BALF samples using sensitivity, specificity and accuracy based on the clinical final diagnosis (Fig. 3A, Additional file 1: Fig. S1).

Overall, as is expected, the application of HDA-mNGS has demonstrably enhanced the diagnostic efficacy of PTB in the present study. HDA-mNGS accurately identified 59 instances as positive for the MTBC, attaining a sensitivity of 72.0% (95%CI, 60.8–81.0%). According to McNemar test, the sensitivity of HDA-mNGS was significantly higher than the other four methods (Fig. 3B). Furthermore, HDA-mNGS exhibited the highest accuracy among the evaluated techniques, reaching 74.5% (95%CI, 65.9–83.1%). The performance of HDA-Nanopore ranked second in both sensitivity (58.5%) and accuracy (62.2%). Conventional mNGS and Xpert exhibited similar diagnostic performance, with sensitivities of 51.2% versus 49.4% and accuracy of 58.3% and 56.0% (Fig. 3C). Notably, as the currently golden standard for PTB pathogen detection, mycobacterial culture yielded positive results in only 25 samples. The culture method exhibited limitations, with 52 false-negative outcomes.

In the subset of definitively diagnosed PTB cases (n = 50) characterized by clear clinical manifestations, HDA-mNGS displayed a correct detection of 86.0% (43/50), surpassing the performance of conventional mNGS, which exhibited a rate of 54.0% (27/50). In the clinically diagnosed PTB cases, both mycobacterial culture and Xpert failed to detect MTBC. HDA-mNGS and conventional mNGS demonstrated comparable positive detection rate in detecting MTBC. However, they exhibited variations in the specific positive samples identified. In summary, the study highlights the significant improvement in diagnostic accuracy provided by HDA-mNGS for PTB.

Interestingly, we have investigated the co-infection among 98 specimens using metagenomic sequencing methods including conventional mNGS, HDA-mNGS, and HDA-Nanopore. Positive co-infection cases were defined as those detected by any of two methods. Out of the 82 PTB positive patients, 56 were found to have co-infections such as bacteria, fungi, and viruses, and HDA-mNGS demonstrated the best detection performance in co-infection analysis due to its highest rate (100.0%) of pathogen identification (Additional file 3: Table S2).

HDA-mNGS increases the coverage of MTB genome and drug resistance gene

Multiple TB drug resistance-inducing mutations can be detected simultaneously by metagenomic sequencing but are highly dependent on sequence coverage. As reported previously, the detection rate of antimicrobial resistance genes (ARGs) using mNGS was extremely low due to limited coverage and depth of the MTB genome sequence [3]. In our study, 43 MTB positive cases using mNGS had a mean coverage rate of ≥ 1X across the MTB reference genome (H37Rv), reaching 95,267 bp (2.16%, ranging from 50 bp (0.001%) to 1,441,635 bp (32.7%) (Fig. 4A). In contrast, HDA-mNGS detected 61 MTB-positive cases with a mean coverage rate of > 1X across the MTB reference genome, reaching 1,528,168 bp (34.6%, ranging from 146 bp (0.003%) to 4,334,162 bp (98.2%)) (Fig. 4A). The sequences identified using conventional mNGS and HDA-mNGS methods varied across the different samples.

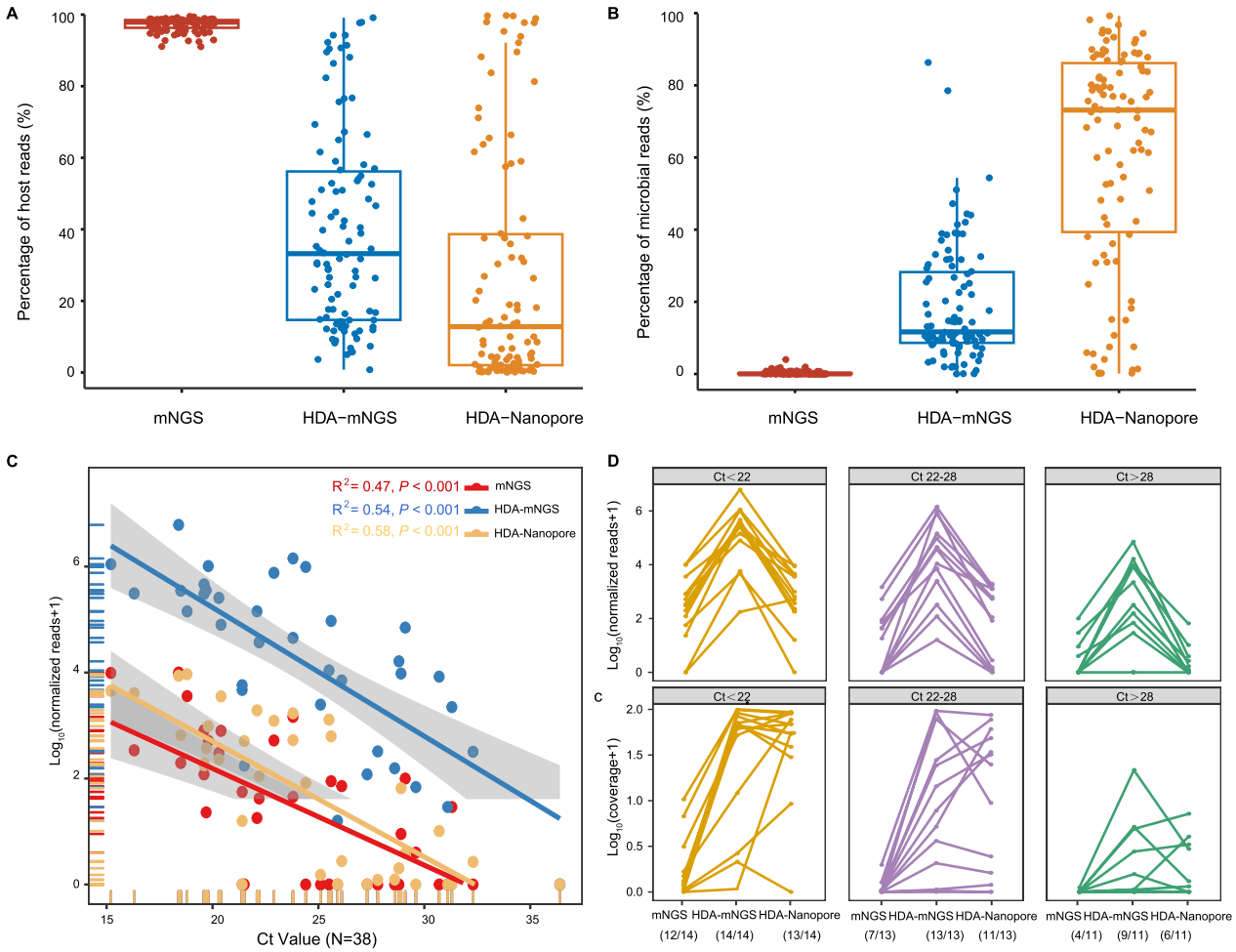


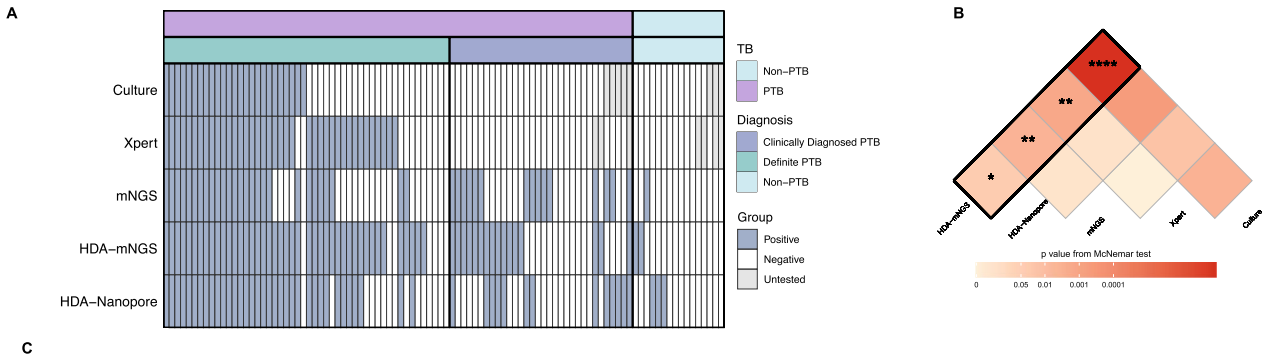
Fig. 2 HDA-mNGS increases both microbial and TB reads in Xpert positive samples. **A** Boxplot of host reads rates from three methods suggest that HDA steps reduced human DNA percentage rates to thereby dramatically increase the Pathogen Reads rate, as confirmed using the Wilcoxon test. **B** Metagenomics sequencing methods that incorporate steps to substantially remove host DNA were associated with increased Pathogen Reads rates. **C** The Ct value of Xpert indicated a negative correlation with the number of normalised reads obtained by metagenomic sequencing. **D** According to Ct value cutoffs (< 22, 22–28, > 28), the detection performance of metage-nomic sequencing gradually decreased with decreasing MTB DNA content. Moreover, HDA-mNGS results were more robust as compared to results obtained using mNGS and HDA-Nanopore: Coverage of the M. tuberculosis genome obtained using metagenomic sequencing, based on Ct values for different patient groups (non-PTB versus PTB groups). The results indicated that HDA-mNGS provided excellent coverage depth that exceeded the coverage depths obtained using the other two methods

We explored the depth of 184 antimicrobial resistance locus published by WHO within the positive samples classified by mNGS and HAD-mNGS. The depth analysis showed that 15 of the 61 cases obtained from PTB cases had coverage depths of at least 3X, while 9 within these cases had had coverage depths of 10X or more (Fig. 4B).

In contrast, conventional mNGS provided only little sequence information compared to HDA-mNGS, making it impossible to determine the AMR gene coverage (Additional file 4: Table S3). Taken together, these results suggest that HDA-mNGS provided a greatly improved coverage of the MTB genome sequence and relatively higher AMR gene coverage in positive PTB cases.

Discussion

Infections caused by MTB have posed a consistent and formidable threat to global human health across centuries. The urgent need to accurately detect and effectively manage diseases associated with this notorious human pathogen has prompted us to enhance the diagnostic performance. In this study, we present a modified diagnostic approach, HDA-mNGS, which enhances the detection



C

Methods	TP	FN	FP	TN	Sensitivity (95%CI)	Specficity (95%CI)	Accuracy(95%CI)
960 Culture	25	52	0	13	32.5% (22.5%~44.2%)	100.0% (22.5%~44.2%)	42.2% (32.0%~52.4%)
Xpert MTB/RIF	39	40	0	12	49.4% (38.0%~60.8%)	100.0% (38.0%~60.8%)	56.0% (45.8%~66.2%)
mNGS	42	40	1	15	51.2% (40.0%~62.3%)	93.8% (40.0%~62.3%)	58.2% (48.4%~67.9%)
HDA-mNGS	59	23	2	14	72.0% (60.8%~81.0%)	87.5% (60.8%~81.0%)	74.5% (65.9%~83.1%)
HDA-Nanopore	48	34	3	13	58.5% (47.1%~69.1%)	81.3% (47.1%~69.1%)	62.2% (52.6%~71.8%)

Fig. 3 HDA-mNGS improves PTB diagnostic performance for BALF samples. **A** MTB detection rate obtained using five different methods. np, number of patients. **B** Heatmap of *P* value analyzed using McNemar test between HDA-mNGS with other methods (culture, Xpert, mNGS and HDA-Nanopore), * indicates the significance. (**C**) Comparative diagnostic performance of five different methods

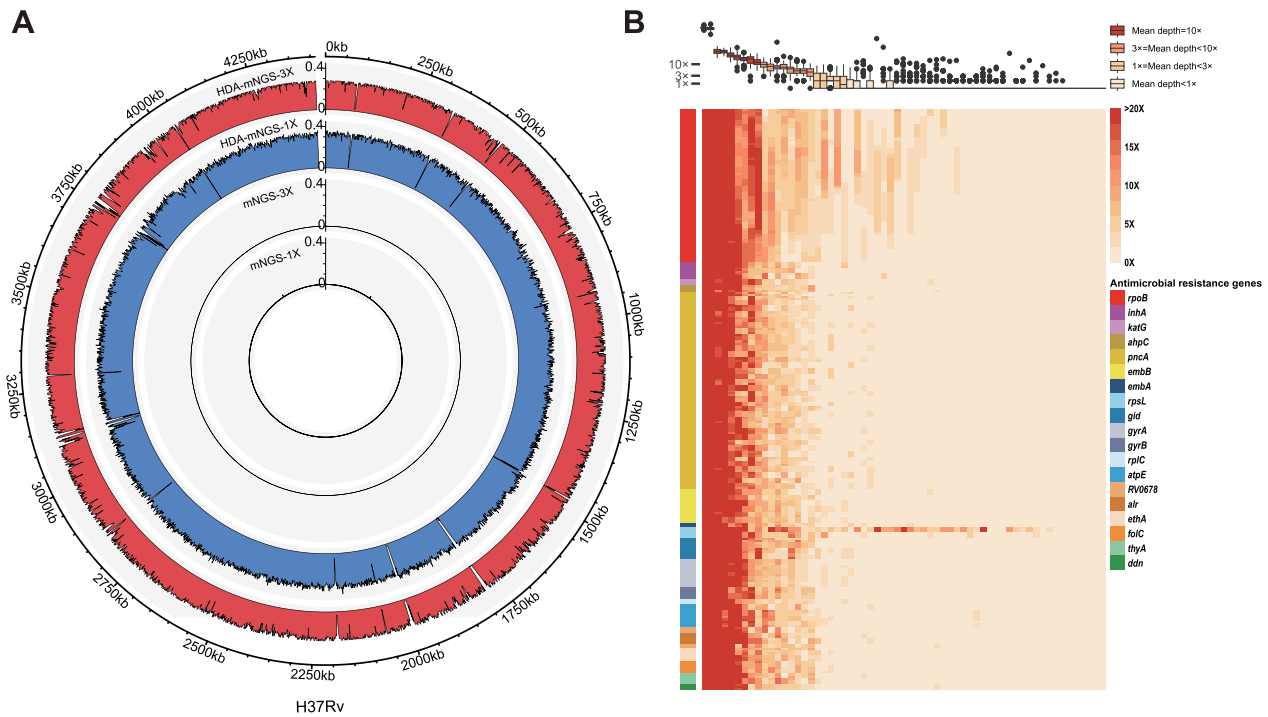


Fig. 4 Sequencing depth of MTB genome sequences obtained using mNGS and HDA-mNGS. **A** Coverage depth distribution of MTB genome. From inside to outside shows sequence depths of conventional mNGS of 1X and 3X and HDA-mNGS of 1X and 3X. The vertical text on the right indicates percent values of samples, whereby the circled numbers represent positions within the M. tuberculosis H37Rv sequence. **B** Heatmap of depth for each sample in 184 antimicrobial resistance locus

of pulmonary TB in BALF samples. By combining host DNA depletion and conventional metagenomics, this method demonstrated superior clinical performance in a cohort of 98 enrolled patients with suspected pulmonary TB infection. Moreover, HDA-mNGS offers increased coverage of the MTB genome and drug resistance genes, underscoring its potential for effective antibiotic resistance management in PTB patients.

Sputum smear examinations are the only testing method used by most TB laboratories in China, particularly in rural aeras. However, the lack of additional testing methods such as nucleic acid analysis and culture results in delayed diagnosis and treatment, prolonged infectiousness, and continued transmission of MTB [17]. In our study, only 20.73% (17/82) PTB patients were sputum smear positive. Our findings suggest that conventional mNGS had similar diagnostic efficiency to Xpert for TB diagnosis in BALF samples, which was consistent with previous studies [3, 7, 9]. However, the rate of confirmed pulmonary TB was less than 52%, which is lower than the 80% bacteriological detection rate in several high-income countries. By combining host depletion and mNGS, we were able to increase the MTB detection rate to 71.95% in 82 patients with PTB.

BALF samples typically contain high levels of human nucleic acid that could lower the accuracy of detecting low-abundance pathogens [18]. In our study, more than 95% of the raw NGS reads were from human DNA, highlighting the need to reduce irrelevant human nucleic acid to enhance the relative proportion of microorganism-derived sequences [19]. Pre-extraction methods have been proposed to deplete the host cell using its fragility compare to viral capsids and microbial cell walls [20]. This method has a potential risk of increasing exogenous background contamination due to the use of additional reagents or procedures. In addition, this technique may lead to a decrease in microbial reads due to the elimination of intact or intracellular microbes such as MTB [18, 21]. It is intriguing that we have not observed the decrease of MTB sequences in our tested BALF samples but improved its detection after host DNA depletion (Fig. 2), compared to conventional mNGS. One of the reasons is that during the extraction of BALF samples, some MTB cells were subjected to osmotic shock, which caused them to be expelled from the human cells and then detected by mNGS method.

We further conducted a systematic comparison of five methods for detecting MTB by culture, Xpert, mNGS, HDA-mNGS, and HDA-Nanopore. We revealed that HDA-mNGS provided superior sensitivity and robust results, with a PTB diagnostic yield of up to 71.95% (59/82) for all suspected PTB cases and 84.62% (44/52) for definite PTB cases. Moreover, the depletion of host DNA significantly improved the sensitivity and robustness of the HDA-mNGS method compared to Nanopore-based sequencing method and conventional mNGS methods. We have previously developed a conventional mNGS method that demonstrated comparable sensitivity to Xpert and culture in TB detection [3]. These findings highlight the clinical utility of HDA-mNGS in PTB diagnosis, particularly for BALF samples. Furthermore, this approach can be adapted to establish modified mNGS methods that are suitable for a wider range of clinical scenarios. In addition, we have uncovered increased co-infections using HDA-mNGS assay compared to conventional mNGS and HDA-Nanopore assays, further proving that HDA-mNGS is potentially useful in PTB differential diagnosis for BALF samples.

Our results highlight the potential use of mNGS-based methods for detecting TB drug resistance gene in inferring antimicrobial susceptibility [22]. However, conventional mNGS has limited effectiveness in detecting pathogen AMRs due to the low abundance of these sequences in respirators or body fluids [23] and thus the use of mNGS for this purpose has remained limited [5, 24–27]. Meantime, our results demonstrated limited coverage and depth of MTB genomic sequencing of conventional mNGS, resulting in very low effectiveness of AMR gene detection, consistent with the results of a previously reported study [3]. However, by combining host DNA depletion with mNGS, we observed a significant improvement in the coverage and depth of MTB genomic sequencing compared to that of conventional mNGS, thus leading to better detection rate of AMR gene sequences in PTB positive samples.

Furthermore, we investigated the potential use of HDA-mNGS for identifying mutations that confer resistance to rifampicin. Our analysis of the AMR locus revealed that 24.59% of tested samples using HDA-mNGS had sequence depths higher than 10X, with significant variation in coverage observed between different samples. It is worth noting that the detection of AMR-inducing mutations requires larger sequencing depth or other state-of-the-art technique [28–30]. Several studies have shown that bacterial whole-genome data, combined with machine learning models, can be used for drug resistance prediction [31–35]. In the future, similar approaches could potentially be applied to mNGS data. In this context, the increased microbial sequencing depth achieved through the host DNA removal process has the potential to enhance the accuracy of drug resistance prediction. Recently, targeted NGS methods has been proposed that may enhance the ability of mNGS to detect AMR gene sequences by selectively enriching genomic regions that encompass AMR marker sequences [36, 37]. This method may

provide an alternative to direct identification of AMR-inducing mutations that are currently used for paucibacillary specimens and may be useful for detecting AMRs present in host DNA-depleted templates.

Although this study yielded interesting and promising results, there are still several limitations. First, six false positive results occurred in the test, due to errors in one sample analyzed via conventional mNGS, two samples analyzed by HDA-mNGS, and three samples analyzed by HDA-Nanopore. These false-positive results may have been caused by insufficiently stringent threshold settings, highlighting the need for more stringent thresholds to improve the specificity for respiratory tract pathogen detection. Second, false positive MTB detection results in HDA-Nanopore may have been due to high error rates of nanopore-based mNGS (up to 15%), which may have led to incorrect barcode assignments [38]. Third, although mNGS improves pathogen detection compared to 16S rRNA-based microbial detection methods, it comes with higher costs and relies on more complex experimental platforms. Currently, targeted NGS technology, as a more cost-effective alternative to mNGS, has also been applied to tuberculosis pathogen detection and has been proven to exhibit good performance [39]. However, mNGS is not limited by targeted design and can diagnose *M. tuberculosis* while simultaneously detecting a wide range of pathogens, including fungi, viruses, as well as emerging or rare pathogens. This gives mNGS irreplaceable value in the differential diagnosis of complex infections. Fourth, the small sample size of this investigation limits its generalizability, and a larger multi-center study is needed to validate these findings. Finally, it is worth noting that negative drug resistance analysis in this cohort may be attributed to the limited number of drug-resistant cases enrolled. Therefore, expanding the enrollment of suspected patients and investigating the epidemiology of PTB in a larger cohort would be a fascinating avenue for future research.

In summary, our findings demonstrated that HDA-mNGS has the potential to significantly improve PTB diagnosis and identify drug resistance in BALF samples.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12941-025-00782-y.

Additional file 1: Fig. S1. Overview of the study design.

Additional file 2: Table S1. Basic information of 98 samples.

Additional file 3: Table S2. Co-infection spectrum of PTB and non-PTB.

Additional file 4: Table S3. *rpoB* gene coverage analysis for HDA-mNGS and mNGS.

Acknowledgements
We thank all patients and their families for their participation in the study. We also thank Beijing Chest Hospital, Wuhan Pulmonary Hospital, and Ordos Second People's Hospital for their support.

Author contributions
Liang Li, Lifeng Li, and Yu Pang designed and conceived the experiments; Jinfeng Yuan, Liping Ma, Juan Du, and Hailin Sun identified the patients and provided patient samples. Jinfeng Yuan, Shanshan Li, Guanhua Rao, Fengshuo Sun, Hui Miao and Gang Zhou analyzed the metagenomic sequencing data and wrote the manuscript. Dan Tian, Changhao Cheng, Yan Wang and Wangyang Chen provided guidance for metagenomic sequencing and data analyses. All authors discussed these conclusions and reviewed the manuscript.

Funding
The study was supported by the Capital's Funds for Health Improvement and Research (2024-4-1042 and 2024-1-1041).

Data availability
Most of the generated or analyzed data are available in the current study. If necessary, all genomic data are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest
The authors declare that the study has no competing interests. Wangyang Chen, Hui Miao, Gang Zhou, Guanhua Rao, Fengshuo Sun and Lifeng Li are affiliated with Genskey Medical Technology Co., Ltd.

Ethics approval and consent to participate
This study was approved by the Ethics Committee of Beijing Chest Hospital affiliated with Capital Medical University (approval No. KY-2022–022). Informed consent was obtained from all participants in accordance with the Declaration of Helsinki. As the study was retrospective and utilized anonymized data, the committee waived the requirement for informed consent from patients.

Received: 4 December 2024 Accepted: 5 February 2025
Published online: 17 February 2025

References

1. WHO. Global Tuberculosis Report; 2022. https://www.who.int/publications/i/item/9789240061729.
2. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20(6):341–55.
3. Shi C-L, Han P, Tang P-J, et al. Clinical metagenomic sequencing for diagnosis of pulmonary tuberculosis. J Infect. 2020;81(4):567–74.
4. Hall MB, Rabodoarivelo MS, Koch A, et al. Evaluation of nanopore sequencing for *Mycobacterium tuberculosis* drug susceptibility testing and outbreak investigation: a genomic analysis. Lancet Microbe. 2023;4(2):e84–92.
5. Charalampous T, Kay GL, Richardson H, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783–92.
6. Kok NA, Peker N, Schuele L, et al. Host DNA depletion can increase the sensitivity of *Mycobacterium* spp. detection through shotgun metagenomics in sputum. Front Microbiol. 2022;13:949328.
7. Liu X, Hou XF, Gao L, et al. Indicators for prediction of *Mycobacterium tuberculosis* positivity detected with bronchoalveolar lavage fluid. Infect Dis Poverty. 2018;7(1):22.
8. Jin X, Li J, Shao M, et al. Improving suspected pulmonary infection diagnosis by bronchoalveolar lavage fluid metagenomic next-generation sequencing: a multicenter retrospective study. Microbiol Spectrum. 2022;10(4):e0247321.

9. Liu X, Chen Y, Ouyang H, et al. Tuberculosis diagnosis by metagenomic next-generation sequencing on bronchoalveolar lavage fluid: a cross-sectional analysis. Int J Infect Dis. 2021;104:50–7.
10. Association C. China clinical treatment guide for tuberculosis. Beijing: People's Medical Publishing House; 2005.
11. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
12. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C, Berger B. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666–9.
13. Li H, Birol I. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
14. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. BMC Med Res Methodol. 2013;13:91.
15. Winters R, Winters A, Amedee RG. Statistics: a brief overview. Ochsner J. 2010;10(3):213–6.
16. Sun F. 2025. https://BioRender.com/a33e822
17. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-generation sequencing for infectious disease diagnosis and management: a report of the association for molecular pathology. J Mol Diagn. 2015;17(6):623–34.
18. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. Annu Rev Pathol. 2019;14:319–38.
19. Yang J, Yang F, Ren L, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol. 2011;49(10):3463–9.
20. Marotz C, Zuniga C, Zaramela L, Knight R, Zengler K. Host DNA depletion in saliva samples for improved shotgun metagenomics. Methods Mol Biol. 2021;2327:87–92.
21. Diao Z, Han D, Zhang R, Li J. Metagenomics next-generation sequencing tests take the stage in the diagnosis of lower respiratory tract infections. J Adv Res. 2022;38:201–12.
22. Liu H, Zhang Y, Yang J, Liu Y, Chen J. Application of mNGS in the etiological analysis of lower respiratory tract infections and the prediction of drug resistance. Microbiol Spectrum. 2022;10(1):e0250221.
23. Serpa PH, Deng X, Abdelghany M, et al. Metagenomic prediction of antimicrobial resistance in critically ill patients with lower respiratory tract infections. Genome Med. 2022;14(1):74.
24. Yang L, Haidar G, Zia H, et al. Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: a feasibility and clinical validity study. Respir Res. 2019;20(1):1–12.
25. Charalampous T, Alcolea-Medina A, Snell LB, 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):1–16.
26. Chao L, Li J, Zhang Y, Pu H, Yan X. Application of next generation sequencing-based rapid detection platform for microbiological diagnosis and drug resistance prediction in acute lower respiratory infection. Ann Transl Med. 2020;8(24):1644.
27. Quan J, Langelier C, Kuchta A, et al. FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. Nucleic Acids Res. 2019;47(14):e83.
28. Gweon HS, Shaw LP, Swann J, et al. The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environ Microbiome. 2019;14(1):1–15.
29. Hu X, Zhao Y, Han P, et al. Novel clinical mNGS-based machine learning model for rapid antimicrobial susceptibility testing of *Acinetobacter baumannii*. J Clin Microbiol. 2023;61:e0180522.
30. Sanabria AM, Janice J, Hjerde E, Simonsen GS, Hanssen AM. Shotgun-metagenomics based prediction of antibiotic resistance and virulence determinants in *Staphylococcus aureus* from periprosthetic tissue on blood culture bottles. Sci Rep. 2021;11(1):20848.
31. Hu X, Zhao Y, Han P, et al. Novel clinical mNGS-based machine learning model for rapid antimicrobial susceptibility testing of *Acinetobacter baumannii*. J Clin Microbiol. 2023;61(5):e0180522.
32. Liu B, Gao J, Liu XF, et al. Direct prediction of carbapenem resistance in *Pseudomonas aeruginosa* by whole genome sequencing and metagenomic sequencing. J Clin Microbiol. 2023;61(11):e0061723.
33. Sun L, Chen W, Li H, et al. Phenotypic and genotypic analysis of KPC-51 and KPC-52, two novel KPC-2 variants conferring resistance to ceftazidime/avibactam in the KPC-producing *Klebsiella pneumoniae* ST11 clone background. J Antimicrob Chemother. 2020;75(10):3072–4.
34. Tian Y, Zhang D, Chen F, Rao G, Zhang Y. Machine learning-based colistin resistance marker screening and phenotype prediction in *Escherichia coli* from whole genome sequencing data. J Infect. 2024;88(2):191–3.
35. Wang S, Wang L, Jin J, et al. Genomic epidemiology and characterization of carbapenem-resistant *Klebsiella pneumoniae* in ICU Inpatients in Henan Province, China: a multicenter cross-sectional study. Microbiol Spectr. 2023;11(3):e0419722.
36. Cabibbe AM, Spitaleri A, Battaglia S, et al. Application of targeted next-generation sequencing assay on a portable sequencing platform for culture-free detection of drug-resistant tuberculosis from clinical samples. J Clin Microbiol. 2020;58(10):10–1128.
37. Tafess K, Ng TTL, Lao HY, et al. Targeted-sequencing workflows for comprehensive drug resistance profiling of *Mycobacterium tuberculosis* cultures using two commercial sequencing platforms: comparison of analytical and diagnostic performance, turnaround time, and cost. Clin Chem. 2020;66(6):809–20.
38. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 2018;19(1):90.
39. Schwab TC, Perrig L, Göller PC, et al. Targeted next-generation sequencing to diagnose drug-resistant tuberculosis: a systematic review and meta-analysis. Lancet Infect Dis. 2024;24:1162–76.

Publisher's Note