Metagenomics next-generation sequencing tests take the stage in the diagnosis of lower respiratory tract infections

Zhenli Diaoa,b,c,1, Dongsheng Han d,1, Rui Zhang a,c,* , Jinming Lia,b,c,⇑

a National Center for Clinical Laboratories, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing Hospital/ National Center of Gerontology, PR China
b Peking University Fifth School of Clinical Medicine, Beijing Hospital, Beijing, PR China
c Beijing Engineering Research Center of Laboratory Medicine, Beijing, PR China
d Key Laboratory of Clinical In Vitro Diagnostic Techniques of Zhejiang Province, Department of Clinical Laboratory, First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, PR China

HIGHLIGHTS
• The applications of mNGS for LRIs span a wide range of areas including LRI diagnosis, airway microbiome analyses, human host response analyses, and prediction of drug resistance.
• The workflow of mNGS used in clinical practice involves the wet-lab pipeline and dry-lab pipeline, the complex workflow poses challenges for its extensive use.
• mNGS will become an important tool in the field of infectious disease diagnosis in the next decade.

GRAPHICAL ABSTRACT

The typical workflow of mNGS in clinical laboratory.

ARTICLE INFO

Article history:
Received 3 February 2021
Revised 13 September 2021
Accepted 25 September 2021
Available online 29 September 2021

ABSTRACT

Background: Metagenomic next-generation sequencing (mNGS) has changed the diagnosis landscape of lower respiratory tract infections (LRIs). With the development of newer sequencing assays, it is now possible to assess all microorganisms in a sample using a single mNGS analysis. The applications of mNGS for LRIs span a wide range of areas including LRI diagnosis, airway microbiome analyses, human host response analyses, and prediction of drug resistance. mNGS is currently in an exciting transitional period;

Abbreviations: ARGs, antibiotic resistance genes; CAP, Community-acquired pneumonia; CSF, Cerebrospinal fluid; DNase, Deoxyribonuclease; DASH, Depletion of Abundant Sequences by Hybridization; DNBs, DNA nanoballs; Fil, 5-lM filtration; IQC, Internal quality control; IQR, Interquartile range; LDTs, Laboratory-developed tests; LRIs, Lower respiratory tract infections; mNGS, Metagenomic next-generation sequencing; Mol, MolYsis Basic; MTB, M. tuberculosis; NCBI, National Center for Biotechnology Information; NEB, NEBNext Microbiome DNA Enrichment Kit; NPA, nasopharyngeal aspirate; PMA, Propidium monoazide; PT, Proficiency testing; QIA, QIAamp DNA Microbiome Kit; RMB, renminbi; RoC, Receiver-operating curve; RT-PCR, Reverse-transcription PCR; RVP, respiratory virus panel; SMRT, single-molecule real-time sequencing; TATs, Typical turnaround times; WGS, Whole-genome sequencing.

Peer review under responsibility of Cairo University.
* Corresponding authors at: National Center for Clinical Laboratories, Beijing Hospital, No.1 Dahua Road, Dongdan, Beijing 100730, PR China.
E-mail addresses: ruzhang@nccl.org.cn (R. Zhang), jml@nccl.org.cn (J. Li).
1 They contributed equally to the manuscript.

https://doi.org/10.1016/j.jare.2021.09.012
2090-1232/© 2022 The Authors. Published by Elsevier B.V. on behalf of Cairo University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Introduction

Lower respiratory tract infection (LRI), including community-acquired pneumonia (CAP), hospital-acquired pneumonia, bronchitis, bronchiolitis, and tracheitis, is the fifth-leading cause of death, which has been reported to cause 2.74 million deaths (95% uncertainty interval 2.50 million to 2.86 million) in 2015 [1]. LRIs are caused by a wide array of pathogens, such as bacteria, viruses, mycoplasma, and fungi, all of which present indistinguishable clinical presentations. The etiologies of up to 62% of CAP remain undiagnosed despite comprehensive diagnostic work-up [2]. Without a definitive microbiological diagnosis, patients with severe LRIs are often treated with empirical broad-spectrum antibiotics to relieve their symptoms during the initial treatment [3]. Clinicians should adjust or stop such empirical treatment once pathogens are identified. However, such therapies are often continued if the patient is responding well or if no contributory pathogens have been detected, which leads to the abuse of broad-spectrum antibiotics. Furthermore, in the absence of a microbial etiology, clinicians may mistakenly classify the symptoms into a noninfectious inflammatory condition and prescribe empiric corticosteroids for treatment, which may result in reinfection [4].

Rapid and accurate identification of pathogens enables tailored treatments, reduces the abuse of broad-spectrum antibiotics, and prompts the eventual recovery of patients. Culture, as the gold standard for microbiological identification, is time-consuming with low sensitivity, especially for fastidious organisms [5,6]. Although culture-independent techniques, such as immunological assays and nucleic acid testing using PCR, are rapid and accurate, they require prior knowledge or assumptions regarding the types of pathogenic microorganisms. Metagenomic next-generation sequencing (mNGS) may serve as a new tool to overcome the shortcomings of conventional diagnostic methods. The chief advantage of mNGS lies in its unbiased sampling, which enables the simultaneous identification of all potentially infectious agents in samples and avoids defining the targets for diagnosis beforehand [7]. Therefore, it has obvious advantages in the diagnosis of unexplained and co-infectious LRIs. mNGS is a useful technique to detect novel or rare microorganisms and is also efficient in improving the analytical sensitivity for the identification of fastidious microorganisms and diagnosis of pulmonary co-infections. The outcomes of mNGS are less likely influenced by prior antibiotic exposure than culture-dependent methods [8]. In addition to pathogen identification, mNGS also provides additional genomic information necessary for airway microbiome analyses, human host response analyses, and prediction of drug resistance, all of which facilitates clinical management of patients with LRIs (Fig. 1) [3,9]. However, the biological variation in sampling (timing of sampling, host DNA level, contamination, etc.) and the technical variation in methodology (ununified nucleic acid extraction methods, incomplete databases, differentiated bioinformatics tools, and unstandardized interpretation standards) limit its widespread use in a clinical setting. In this review, we illustrate the potential applications and challenges of integrating mNGS into the management of patients with LRIs in all aspects and provide a comprehensive understanding of mNGS for both clinicians and researchers.

Applications of mNGS in lower respiratory infections

Diagnosis of lower respiratory tract infections

mNGS, as a culture-independent, unbiased, and hypothesis-free approach, has emerged as a diagnostic method for respiratory tract infections in recent years (Table 1). Current molecular tests for LRI diagnoses are usually pathogen-specific that clinicians select relevant tests according to the symptoms of patients, which poses a challenge when novel or unexpected pathogens emerge. In contrast, mNGS can provide a comprehensive view of pathogens in a given sample, which enables the detection of novel and rare causative pathogens in the diagnosis of unexplained pneumonia. For example, in early December 2019, severe unexplained pneumonia emerged in Wuhan, Hubei Province, China. On February 3, 2020, a novel coronavirus (SARS-CoV-2) was identified to cause pneumonia, which was determined using RNA based mNGS [10]. Compared to the time taken to identify SARS (five months), mNGS shortened the time taken considerably to five days for the accurate identification of the gene sequence of the virus [10]. In addition, mNGS could provide clues for identifying rare pathogens and reducing the delay in the diagnosis of unexplained pneumonia. For example, humans infected by *Chlamydia psittaci* could present various degrees of severity of pneumonia, which is responsible for less than 5% of the cases of CAP [11]. The diagnosis of *C. psittaci* infection is challenging in clinical, as the traditional culture-based methods are time-consuming and have low yields, and serology tests may cross-react with other Chlamydiaeae species. Although PCR-based methods are more rapid, sensitive, and specific, they are only performed if the clinicians request for the relevant tests. The wide detection range and lack of requirement for an assumption for the suspected causative organism make mNGS an effective tool to diagnose *C. psittaci* pneumonia [12].

Conventional culture methodology has a low detection rate for pathogens that are difficult to culture or require long culture periods. mNGS, as a culture-independent detection method, is promising for the detection of these fastidious organisms in shorter feedback times. Miao et al. investigated a cohort of 561 patients with acute or chronic infections to assess mNGS performance in real-life clinical practice [8]. They demonstrated that mNGS has 50.7% sensitivity and 85.7% specificity for diagnosing infectious diseases. Moreover, the analytical performance of mNGS outperformed that of the culture, especially for fastidious organisms, such as *Mycobacterium tuberculosis*, viruses, anaerobes, and fungi. mNGS has been reported to achieve 100% specificity in the evaluation of fungi from lung biopsy tissues when compared to histopathology methods [5,8]. *M. tuberculosis* (MTB) requires prolonged culture
time and its detection rate is low. A recent study by Shi et al. demonstrated that mNGS showed 47.92% sensitivity for the detection of MTB, which is consist with X-pert (45.83%) and culture (46.81%) [13]. However, mNGS required merely three days to identify 67.23% of cases of MTB infections, whereas 49.58% of MTB infection cases detected using conventional methods required over 90 days [13]. Currently, the average typical turnaround times (TATs) for most mNGS platforms from specimen receipt to the final results is 48 h [3, 7, 8]. Nanopore sequencing technology has been reported to even reduce TAT to 6 h [3]. Compared with the conventional culture-based methods that the average TAT of pathogen culture is ≥ 3 days for bacteria, 7 days for fungi, and 45 days for mycobacteria, 2-day TAT for mNGS is acceptable for clinical laboratories [8, 14]. mNGS improves the detection conditions for fastidious organisms, accelerates clinical decision-making process and promotes rational antibiotic therapy.

Compared to conventional tests, mNGS has a broader spectrum for pathogen detection in a single test, which streamlines clinical testing for pulmonary co-infection diagnosis. A retrospective study evaluated 55 enrolled patients with mixed pulmonary infections to explore the analytical performance of mNGS [7]. They found that mNGS had a higher sensitivity for diagnosing mixed pulmonary infection than conventional tests (97.2% vs. 13.9%; P < 0.01); however, the specificity was lower (63.2% vs. 94.7%; P = 0.07). Babičker et al. used mNGS to estimate the RNA respiratory virus infection status in 75 individuals who were examined [15]. In this study, mNGS showed 100% concordance (n = 45, 60%) with reverse-transcription PCR (RT-PCR) for detecting SARS-CoV-2, which also identified both co-infections (n = 1, 2.2%) and alternative viral infections (n = 4, 13.3%) that were missed during routine clinical workup. A correlation was observed between SARS-CoV-2 read recovery using mNGS and the threshold cycle value obtained using RT-PCR.

The prior use of broad-spectrum antibiotics tends to result in “false-negative” results for conventional culture methodology; in contrast, mNGS is less influenced by prior antibiotic treatment. A study reported that no significant difference in sensitivity was observed between mNGS and culture in non-antibiotic-exposed patients (43.3% vs. 36.7%; P = 0.10) [8]. However, mNGS showed significantly higher sensitivity than that of culture (52.7% vs. 34.4%; P < 0.01) in patients with prior antibiotic usage.

**Airway microbiome analyses**

The airway microbiome is the sum of microbes that coexist in the airways of healthy subjects and patients with respiratory diseases. The advent of NGS has greatly promoted the boom of microbiome analyses, as it has allowed sequencing to become more accessible and time-efficient. Additionally, the development of mNGS is the cornerstone of advances in the area of microbiome analysis. After mapping the available sequencing information into microbiology resource databases, mNGS can overcome the limitations of targeted detection methods to characterize all microorganisms within human body systems using a single test.

It has been reported that 20–50% of healthy individuals’ airways are colonized by opportunistic pathogens, such as *Streptococcus*...
The analytical performance of mNGS in the diagnosis of respiratory tract infections.

| Study                        | Platform          | Samples                                                                 | Sequencing                                      |
|------------------------------|-------------------|------------------------------------------------------------------------|-------------------------------------------------|
| Langelier C et al. [9]       | Illumina HiSeq    | Clinical microbiologic testing; 92 YA samples                          | RNA and DNA based mNGS                          |
| Charalampous T et al. [3]    | MinION qPCR       | Respiratory samples (sputum, DNA based mNGS                            | DNA based mNGS                                  |
| Wang J et al. [7]            | NA                | Conventional tests (smear, culture, pathology, GM test, Xpert MTB)     | DNA based mNGS                                  |
| van Rijn AL et al. [71]      | Illumina NextSeq  | 96                                                                     | RNA and DNA based mNGS                          |
| Li H et al. [5]              | BGISEQ-500        | Lung biopsy tissues                                                     | DNA based mNGS                                  |
| Huang J et al. [54]          | BGISEQ-100        | Samples (lung tissue, BALF, endotracheal secretions and ETAs)           | RNA based mNGS                                  |
| Shi CL et al. [13]           | Illumina HiSeq    | Culture, microscopic examination, and AFS                              | DNA based mNGS                                  |
| van Boheemen S et al. [19]   | Illumina and NextSeq | 19 Nasopharyngeal washings, 2 sputa with BALF, 1 bronchoalveolar washing, and 1 | DNA based mNGS                                  |
|                              |                    |                                                                        | RNA based mNGS                                  |
|                              |                    |                                                                        | DNA based mNGS                                  |

Abbreviations: AFS, acid-fast stain; BALF, bronchoalveolar lavage fluid; NA, no accessible; NPV, negative predict value; PPV, positive predict value; PSB, protected specimen brushes; TA, tracheal aspirate.

Human host response analyses

Normally, the RNA based mNGS approach is more complex than DNA based mNGS approach, but the former could address several issues that cannot be tackled by the latter [18]. For example, RNA based mNGS approach is particularly useful for identification of viral infections which will be missed in DNA based mNGS approach [19]. DNA-based mNGS puzzles over the differentiation between viable and unviable bacterial cells, but RNA based mNGS has been demonstrated useful for the identification of viable pathogens [20].

DNA-based mNGS is expected to measure the number of genomes for each species, while RNA based mNGS could provide a measurement of gene expression. Thus, RNA based mNGS not only allows taxonomic analysis but also provides host response assessment in a single experiment, which serves as an auxiliary tool to differentiate non-infectious or infectious illnesses, bacterial or viral infections in clinical settings [21]. In individuals hospitalized with acute respiratory illnesses, significantly increased expression of gene sets correlated with immune responses has been observed in patients with confirmed LRI pathogens compared to those in whom definite pathogens have not been identified (median, 94.9 vs. 33.1, Wilcoxon rank sum P = 0.017). Another study performing mNGS for 92 acute respiratory failure patients demonstrated that LRIs were related to reduced intra-patient diversity and interpatient diversity of the airway microbiome [9]. Furthermore, they utilized diversity, which was assessed using RNA based mNGS to predict LRIs and obtained a receiver-operating curve (ROC) of 0.80 (95% CI, 0.89–1.00) in the validation cohort. Although no microbiome-based tests have performed clinically validation for the diagnosis of disease, mNGS provides an efficient tool for airway microbiome analyses, which is helpful for serving as a biomarker for distinguishing infectious and non-infectious diseases.

**Table 1**

The analytical performance of mNGS in the diagnosis of respiratory tract infections.
in viral infections correlated with antiviral immune processes, for example, interferon genes. Making the best use of auxiliary genomic information from mNGS for human host response analyses provides a new perspective and approach to comprehensively characterize infection status.

**Prediction of drug resistance**

Providing only the pathogen identification results is far from satisfactory, it is necessary to detect clinically relevant antibiotic resistance genes (ARGs) and further predict pathogen resistance to guide patient management. Currently, most clinical microbiology laboratories depend on conventional culture-dependent phenotypic methods (e.g., gradient diffusion, disk diffusion, etc.) and culture-independent molecular methods to evaluate the resistance status [22]. However, all culture-dependent methods have inherent shortcomings, such as time-consuming, labor-intensive, the bias from predominant microbial populations, and the risk of contamination overgrowth. Molecular methods are rapid, but only a small set of prominent ARGs derived from limited common pathogens can be targeted by traditional molecular methods.

With the development of sequencing technologies, culture-dependent techniques relying on whole-genome sequencing (WGS) and culture-independent techniques relying on mNGS are available for predicting resistance. Compared with WGS, mNGS forgoes culture bottlenecks, which is more convenient for assessing the resistance status of slowly growing or uncultivable pathogens and dead pathogens due to antibiotic exposure. Wang et al. used both the MinION and BGISEQ-500 platforms to make a bacteriologic diagnosis from a culture-negative lung tissue sample [23]. Not only *Klebsiella pneumoniae* was identified by both platforms as the most top dominant pathogen, additional information of ARGs was also provided. The MinION platform provided for an extremely fast TAT that the resistance genes *blaSHV-12*, *aac(3)-Ila* and *blaKPC-2* were identified at 29th, 38th, and 56th min of sequencing, respectively.

Although culture-independent mNGS is attractive, there remain many hurdles to overcome. A recent study using nanopore metagenomics for bacterial LRI diagnosis has identified 183 ARGs across 41 respiratory samples from patients [3]. However, only 24 (13.11%) were matched to observed resistances by antimicrobial susceptibility testing. Among the other detected genes, 16 genes did not match the phenotype of cultured isolates, nearly 1/3 of the detected genes (56/183) likely originated from the normal or colonizing respiratory flora, some genes were even unlikely to derived from the cultured species in the laboratory. Currently, most mNGS diagnosis platforms are based on short reads sequencing, it is challenging to determine the detected ARGs originated from the genome of the causative agent rather than normal flora, or contaminations in environment. Nanopore sequencing with rapid TAT could produce long reads that are sufficient in length to span repeat regions, which holds promise for ARGs analysis, but the accuracy needs to be further improved [18]. What’s more, there is no isolate to confirm genomic resistance prediction with true phenotypic susceptibility testing [3]. Thus, it is uncertain whether the identification of ARGs is relevant to the resistance phenotypes. Besides, it requires isolation and sequencing of all bacteria present in the sample to determine the specificity and sensitivity of ARGs detection, which poses a great challenge for the assessment of its analytical validation, clinical validation, and clinical utility [22].

**Challenges of mNGS for lower respiratory infection**

**Sequencing platforms**

Sequencer is the material basis of metagenomic sequencing and the choice of the instrument is mainly according to the performance index and clinical needs. Currently, Illumina’s sequencing cannot be matched in terms technology maturity and wide range of platforms, which still dominates the short-read sequencing industry for metagenomic studies [24]. Illumina-based platforms use a strategy of bridge amplification that a single molecule of DNA template first hybridizes with a slide-bound adapter on the flow cell and then amplifies locally into a clonal cluster [25]. Then sequencing by synthesis reaction occurs, in which a single 3’-blocked deoxynucleotides (dNTPs) is added to build the complementary DNA per cycle, and the optical readout of the fluorescently labelled nucleotides determines the dNTP (A, G, T, or C). Illumina offers a popular series of platforms ranges from small, low-throughput benchtop units to large ultra-high throughput instruments. In order to achieve good sample coverage, the higher output instruments such as the HiSeq and NextSeq are widely used in mNGS. The HiSeq series platforms have the advantages of high-throughput and relatively long read length (125 bp/150 bp), however, long run time (3.5 days) limits its use in rapid pathogen detection in clinical [26,27]. In 2018, HiSeq platform has been obsolesced, NextSeq series sequencing systems occupy the mainstream sequencing platform in clinical pathogen detection owing to their moderate throughput and short sequencing time (12–30 h per run) [14,26]. Table 2 describes the parameters of the available sequencing platforms in pathogens detection currently.

The BGI platform is also popular in pathogen detection owing to its low cost and short sequencing time [28]. This technology clones single-stranded circular DNA using rolling circle amplification to produce DNA nanoballs (DNBs) [29]. Then DNBs are adsorbed onto silicon substrates using DNB loading technology, whereby DNA molecular anchors and fluorescent probes are polymerized. Similarly, the resulting optical signals are captured by a high-resolution imaging system and converted into the sequence information. This technology presents a very high accuracy (~99.99%), because each base is probed multiple times [25].

The Ion Torrent platform offered by Thermo Fisher Scientific is the first NGS platform without optical sensing, which detects the released H+ ions as each dNTP is incorporated [30]. This platform has superior read lengths (up to 600 bp) compared to other short-read sequencers [31]. The Ion Torrent platform provides several types of chips to meet the different needs of the researcher, the output of the chips ranges from ~50 Mb to 15 Gb and the runtime is between 2 and 7 h [25].

The application of third-generation sequencing (also known as single-molecule sequencing technology) is another major turning point in the field of mNGS. Currently, there are two representative types of third-generation sequencing technologies: single-molecule real-time sequencing (SMRT) and nanopore sequencing. SMRT sequencing relies on the principle of sequencing by synthesis and utilizes nanoscale zero-mode waveguide to achieve real-time sequencing of single DNA molecules. The SMRT sequencing has the advantage of long read length and rapid sequencing speeds that the SMRT Sequel series by PacBio have an average read length of 10–20 kb, and can achieve 160 GB data output within 6 h [26]. However, due to the bulky equipment and expensive hardware, SMRT seems less popular than nanopore sequencing in the area of pathogens detection.

In 2014, Oxford Nanopore Technologies unveiled the first consumer prototype of the MinION sequencer, which was characterized by inexpensive (starting pack available for 1000$) and portable (4 in. long) [32]. Nanopore sequencing is based on the principle that when the ssDNA/RNA fragment passes through the nanopore, the changes of electrical current are translated into a specific sequence of nucleotides [18]. Nanopore sequencing can generate very long reads (>200 kb) [33], as no DNA amplification occurs during library preparation. This an important improvement in metagenomic analysis since long reads makes de novo genome
assembly more easily and accurately [34]. Besides, nanopore sequencing enables real-time analysis of sequencing data that the pathogens and ARGs can be identified in 6 h [3,35]. However, the features that high error rate, lower throughput, and higher per-read costs limit the widespread adoption of these technologies [36].

Depletion of human nucleic acid

Respiratory tract samples usually contain large amounts of human nucleic acid, which decrease the sensitivity of assays for low-abundance pathogens. 95% of raw NGS reads are derived from the human DNA in nasopharyngeal aspirate samples [37]. Sequencing of unwanted human DNA reads and performing computational human host subtraction from large NGS datasets are wasteful and time-consuming processes. Depleting irrelevant human DNA or RNA increases the relative proportion of microorganism-derived sequences.

Current approaches for depleting human-derived nucleic acids can be implemented before nucleic acid extraction (pre-extraction) or after nucleic acid extraction (post-extraction). Pre-extraction approaches utilize chemical reagents (e.g., saponin) or osmotic lysis to selectively lyse human cells, followed by using deoxyribonuclease (DNase) or propidium monoazide (PMA) to degrade the released human genomic content, remaining only the intact microorganism for downstream analysis [3,38]. Hasan et al. compared the efficiency of saponin, Tween-20, Triton X-100, and Chaps Cell Extract 158 Buffer (New England Biolabs) for selective lysis of human cells and used Turbo DNase for post-lysis treatment [39]. Saponin at a concentration of 0.025% was the most effective in both cerebrospinal fluid (CSF) and NPA specimens, which increased approximately 30- to 100-fold of pathogen DNA to human DNA ratios. Furthermore, there was a significant enrichment of ~ 40- and ~ 170-fold compared to the unprocessed specimens in the CSF and NPA specimens, respectively. A recent nanopore sequencing study presented an optimized saponin-based host DNA depletion method for bacterial LRI diagnosis, which removed up to 99.99% of the human DNA from respiratory samples [3]. Marotz et al. developed a novel method that using osmotic lysis followed by 10 μM PMA treatment to enrich the microbial DNA from human oral samples and compared its performance with four commercially available kits used for host depletion: 5-μM filtration (Fil), QIAamp DNA Microbiome Kit (QIA), MolYsis™ Basic (Mol), and NEBNext Microbiome DNA Enrichment Kit (NEB) [38]. Compared to the average proportion of human reads in the raw samples (89.29 ± 0.61%), treatment with lyPMA (8.53 ± 2.08%), QIA (29.17 ± 5.04%), and Mol (62.88 ± 3.46%) significantly depleted host reads, but no difference was observed in NEB kit (90.83 ± 0.77%) treated samples. Collectively, pre-extraction approaches assume that the human cell membrane is more fragile than most viral capsids or microbial cell walls. These methods sacrifice the sensitivity for detecting some special pathogens without cell walls (such as parasites or Mycoplasma spp.) and free nucleic acids from dead organisms. Additionally, there is a risk of indiscriminately increasing exogenous background contamination from the use of additional reagents. Physically separating (such as physical filtration and centrifugation) the cells and cell-free compartments of clinical samples during the preanalytical phase is a convenient approach to decrease human-derived nucleic acids [36]. A caveat to this approach is that it decreases microbial reads after discarding intact or intracellular microorganisms [36].

Post-extraction approaches can overcome these issues of pre-extraction approaches and provide an alternative for DNA libraries, one approach takes advantage of the methyl-CpG binding domain to selectively separate the methylated host DNA from microbial DNA [40]. This method decreases host genomes sequence reads by 50-folds and increases bacterial and Plasmodium reads by 8–11.5-folds [40]. For RNA libraries, the methods for the subtraction of abundant human rRNA or mitochondrial RNA sequences are mature and have been described for transcriptome analysis [41,42]. The depletion of rRNA or mitochondrial RNA sequences would indirectly increase the ratio of microbial reads and improve the analytical sensitivity for pathogen detection. The depletion methods for RNA libraries include using capture probes followed by binding to magnetic beads for subtraction [43] or by RNase H treatment [41], using antibodies against human and mitochondrial rRNA [42], using CRISPR-Cas9 to selectively target unwanted species for cleavage [44]. In recent years, a series of simpler, cost-effective, and optimized methods have been developed for RNA depletion. Culviner et al. recently developed a method (“do-it-yourself” kit), which was based on the specific hybridization of biotinylated oligonucleotides to the 23S, 16S, and 55 rRNAs, followed by precipitation of these complexes by magnetic streptavidin-coated beads [45]. After processing, 75–80% of reads in RNA based mNGS were derived from mRNA. In 2016, Gu et al. introduced CRISPR-associated nuclelease Cas9 technology to selectively deplete unwanted high-abundance sequences from eukaryotic cDNA libraries [44]. This method, also called DASH

| Platform                  | Maximum Read Length [bp] | Maximum Reads per Run | Maximum Output | Run Time (hours) | Advantages                                      | Disadvantages                                      | Reference |
|---------------------------|--------------------------|-----------------------|----------------|------------------|------------------------------------------------|---------------------------------------------------|-----------|
| Illumina NextSeq 550      | PE150                    | 400 million           | 120 Gb         | 12–30            | Moderate throughput and short running time     | Short read length                                 | [14,67]   |
| Illumina NextSeq 1000 & 2000 | PE150                 | 1.1 billion           | 330 Gb         | 11–48            | High throughput                                 | Short read length                                 | [73]      |
| Illumina NovaSeq         | PE250                    | 20 billion            | 6 Tb           | 13–44            | High throughput and long read length            | Long running time                                 | [74,75]   |
| MGISEQ-200               | PE150                    | 100 million           | 150 Gb         | 9–40             | Low cost and high accuracy                     | Short read length                                 | [76]      |
| MGISEQ-2000              | SE400/PE200              | 1500–1800 million      | 1440 Gb        | FSC: 17–37; PE: 3–21.5 | Low cost, high throughput and short running time | Short read length and long running time | [77]      |
| Ion GeneStudio 55        | 600                      | 130 million           | 15 Gb          | FCE: 17–109      | High compatibility and short running time       | Long read length                                 | [78]      |
| Oxford Nanopore Minion   | > 4 Mb                   | /                     | 50 Gb          | Up to 20         | Long read length and short running time         | Low accuracy and high cost                        | [79]      |
| PacBio Sequel system     | 1–1.8 Kb                 | /                     | 3.5–7 Gb       | Up to 20         | Long read length and short running time         | Low accuracy and high cost                        | [80]      |
(Depletion of Abundant Sequences by Hybridization), has a >99% reduction of the mitochondrial rRNA on eukaryotic samples. Recently, Prezza et al. evaluated the performance of DASH for bacterial short-read RNA-seq, which removed 56–86% of rRNA reads in RNA libraries from *Salmonella enterica* and *Bacteroides thetaiotaomicron* [46].

### Nucleic acid contamination

Nucleic acid contamination from exogenous microorganisms is ubiquitous and can be introduced at every step of the wet process. Failure to cope with contamination may lead to false-positive results, even swamping the signals from the low-biomass samples (for example, skin swabs) [47]. The types of contaminants in mNGS assays include external contamination and internal or cross-contamination (Fig. 2). External contamination arises from microorganisms outside the samples, such as the operators' bodies, laboratory environment, consumables, and reagents [47]. The contaminant taxonomic profile is unique between different laboratories and changes over time along with different researcher and time [47]. The latter is caused by the cross-contamination from other samples in the same run during sample processing or sequencing. For example, substandard operation processes or microbial aerosols may introduce microorganisms from high pathogen loads into other samples treated at the same time. In addition, index-hopping during Illumina sequencing [48], bar-coded primers, or adapter contamination during synthesis [49] could result in cross-talk between different samples.

Although it is impossible to eliminate contamination completely and it is difficult to distinguish contaminating microbial sequences from true microbial sequences, it remains necessary to attempt to minimize the impact of contamination (Fig. 2). Lessons can be learned from other molecular testing methods to reduce contamination, such as maintaining a unidirectional workflow and strict physical separation, using 10% sodium hypochlorite to extensively clean the materials and surfaces more frequently, using ultra-purification reagents, documenting lot numbers of kits, and performing periodic swipe tests [33]. A template-free control that will undergo all steps of the workflow parallel to the real samples should be checked in each run [33]. Upright negative controls are also recommended to trace the sources of potential contamination; for example, incorporation of phage lambda in different reagents to monitor contamination in kits [33].

Many *in silico* contaminant removal methods have been proposed, which serve as a supplement to the above laboratory approaches. The most common *in silico* decontamination method involves discarding sequences below the relative abundance threshold [47]. However, there is a risk of expunging low-frequency true sequences in the samples and the remaining abundant contaminants that interfere with downstream analysis. Another simple method subtracts sequences that appear in negative samples in the same run during sample processing or sequencing. For example, substandard operation processes or microbial aerosols may introduce microorganisms from high pathogen loads into other samples treated at the same time. In addition, index-hopping during Illumina sequencing [48], bar-coded primers, or adapter contamination during synthesis [49] could result in cross-talk between different samples.

### Bioinformation challenges

Nowadays, high-throughput sequencing technologies are available, which can produce large amounts of data from clinical samples in a single test. This data explosion has created challenges in terms of data storage and security. The quantity, location, and period of data storage must be carefully considered and adequate measures must be taken to protect patient sequence data and information [56, 57]. Deciphering clinically relevant data from large datasets quickly and accurately is the chief difficulty of mNGS for infectious disease diagnostics. The typical mNGS bioinformatics pipeline is a complicated process, including pre-processing for depletion low-complexity and low-quality reads and the trimming of adapters, human host subtraction, alignment to a reference database, and taxonomic classification of aligned reads [56, 58]. Several user-friendly and automated platforms have been developed to facilitate these processes, such as SURPI+ [58], Taxonomer [59], CosmosID [60], and OneCodex [61]. The development of these easy-to-use software packages will further promote the incorporation of mNGS into clinical microbiology laboratories. Another challenge is that the choice of databases may dominate the accuracy and reliability of the metagenomics analysis. The large and com-
The comprehensive National Center for Biotechnology Information (NCBI) nucleotide database contains the genomic information of all known organisms, which increases the possibility of detecting the rare infections. However, the NCBI contains erroneous information, such as low-complexity sequences, incorrect species annotation, contaminants from human DNA, sequencing vectors, and adaptors, all of which may lead to false-positive results [61]. Some limited but highly curated databases are available now, such as FDA-ARGOS or the FDA Reference Viral Database, which ensure the accuracy and reliability of the initial microbial call [62,63]. These incomplete databases include limited microorganisms, which may result in false-negative results. Incorporating annotated sequences from multiple databases is helpful for improving the accuracy of microorganism identification [56]. Additionally, clinical reference databases must be updated regularly to track the latest version and modify the mis-annotations and other database errors [56].

Other challenges

The complex workflow of mNGS used in clinical practice involves multiple processes, which poses challenges for its extensive use. Apart from technical challenges, the other factors that limit the broad implementation of mNGS such as imperfect quality assurance, high costs, the sequencing depth consideration, and so on.

The complete analytical validation of mNGS is the first step from bench to bedside. There are currently no US Food and Drug Administration (FDA)-cleared methods, instruments, and/or databases for mNGS. Currently, all mNGS tests developed in-house are laboratory-developed tests (LDTs). Before LDTs are implemented clinically, proof-of-concept validation with established performance metrics should be performed, which is time-consuming and extremely expensive [24]. In November 2020, the European Society for Clinical Virology Network on Next-Generation Sequencing (ENNGS) established and published recommendations for the wet lab procedure of viral mNGS; recommendations for the bioinformatics procedure are available recently [33]. Internal quality control (IQC) procedures must be adopted to monitor the performance of the entire testing process of every run, including the pre-analytical, analytical, and post-analytical phases of testing. A series of proficiency testing (PT) programs for mNGS will be launched in succession, which will contribute to the implementation and standardization of mNGS [33].

Currently, the high cost is one of the bottlenecks that restrict the widespread of mNGS in clinical practice. Although the cost of sequencing has dropped sharply since 2014, the average cost of mNGS ranges from US$1,000–2,500 per sample [5]. A study mentioned that the average cost of mNGS is 3,000 renminbi (RMB) (approximately $400) per specimen in China, which is higher than that for any single traditional pathogenic test (600–700 RMB for culture test, 320 RMB for Cryptococcus antigen test, 600 RMB for Aspergillus serological test, and 600 RMB for T.SPOT) [8]. However, mNGS is able to identify all potential pathogens in a single test, which may be more cost-effective than a series of traditional pathogen screening tests. More prospective clinical studies and economic data focusing on the cost-effectiveness of mNGS in improving patient outcomes are urgently needed to justify its clinical utility [56]. Whenever, encouraging the innovation of the sequencing technology and the widespread of laboratory automation will contribute to the cost reduction.

How many reads are needed for mNGS is another question waits for answer. Currently, none hard and fast rules are accessible for how much sequence are required. The choice of read depth is highly dependent on desired outcome and budget [64]. For instance, if the mNGS tests aim to ARGs analysis, higher sequencing depth (10–100-fold) is required than only identification the unknown pathogen [65]. ENNGS recommended > 10 million reads per sample for virus diagnostics due to the generally low proportion of viral reads in clinical samples [33]. For the diagnosis of LRIs,

![Fig. 2. The possible contaminations sources of mNGS and the tips for eliminating the contaminations. The sources of contaminations mainly include four components: (1) Laboratory environment and operators' bodies (2) Consumables and reagents (3) Cross contamination and (4) Reference database contamination. It is necessary to adopt some strategies to minimize the impact of contamination.](image-url)
When sequenced to less than 5 million per sample, they found that the sequencing depth of 5 to 25 million reads has been increasingly employed for unbiased detection of pathogen in clinical samples from infectious patients. LRIs, as a leading cause of infection, is an area where mNGS can make a difference. With the development of automation technology, current barriers, including complex manual operations, complex data analysis, and high costs, will be removed. To ensure accuracy, all ongoing mNGS LDTs should be validated to establish their performance metrics. IQC and PT programs should be launched regularly to promote the standardization of mNGS tests. Once analytical validity has been demonstrated, well-designed prospective clinical trials are required to demonstrate the clinical utility of mNGS in the diagnosis of LRIs or other infectious diseases. In particular, given the background of the SARS-CoV-2 pandemic, mNGS has demonstrated its pivotal role in monitoring and tracking outbreaks. We envisage that mNGS will become an important tool in the field of infectious disease diagnosis in the next decade.

**Funding**

This work was supported by the “AIDS and Hepatitis, and Other Major Infectious Disease Control and Prevention” Programme of China under Grant [No. 2018ZX10102001] and by the National Natural Science Foundation of China under Grant [No. 81703276].

---

### Table 3
Thresholds for identification pathogen by mNGS in the diagnosis of lower respiratory tract infections.

| Study                | Disease                                  | Sample types                          | Platform   | Thresholds for identification pathogen by mNGS |
|---------------------|------------------------------------------|---------------------------------------|------------|-----------------------------------------------|
| Zhang Y. et al. [81] | Pneumocystis pneumonia                   | Sputum, blood, lung tissue and BALF   | NA         | Pneumocystis jirovecii ‘s specific reads ranking among top15 or its relative reads proportion in fungi higher than 85% |
| Wang J. et al. [7]  | Mixed pulmonary infection                | BALF and lung tissue                   | NA         | The infectious pathogen was determined if it met any of the following thresholds: culture and/or histopathological examination positive of bacteria, virus or fungi; IPA was defined using galactomannan antigen and PCR; at least 50 unique reads from a single species of bacteria, virus or fungi; for pathogens with unique reads less than 50, it can still be diagnosed as infectious pathogens with the consistent clinical situation; at least one unique read from MTBC. |
| Li H. et al. [5]    | Pulmonary infection                      | Lung tissues                          | BGISEQ-500 | The infectious pathogens were determined if it met any of the following thresholds: 30% relative abundance at the genus level in bacteria or fungi; culture and/or histopathological examination positive and at least 50 unique reads from a single species of bacteria or fungi; at least one unique read from MTBC. |
| Huang J. et al. [54]| Peripheral pulmonary infection           | Lung tissue, BALF, and protected-specimen brush. | BGISEQ-100 | The criteria for a positive mNGS test result included: the relative abundance of bacteria (Mycobacterium tuberculosis excluded) and fungi was >30% at the genus level; Mycobacterium tuberculosis was considered to be positively detected if at least one read was aligned to the reference genome at species or genus level; positive virus detection was considered when the SMRN was no less than 3; when the pathogen was detected by traditional pathogen detection methods and the mNGS reads number was >50, this pathogen can also be considered as positively detected by mNGS. |
| Langelier C. et al. | LRTIs                                     | BALF                                  | illumina   | Microbes identified as confirmed pathogens if both clinical testing and mNGS identified the microbe; there existed literature evidence of pathogenicity in the lungs; a developed Z-score was at least twofold greater than that of any other microbe of the same type (virus, bacteria, or fungus) identified in the patient. |
| Charalampous, T. et al. [3] | Bacterial lower respiratory infection | Sputum, BALF and ETas                | MinION     | Microbes were considered new potential pathogens if mNGS alone identified the microbe and criteria 2 and 3 described here were met; all other microbes were considered unlikely or uncertain pathogens. |
| Wang, H. et al. [55] | Severe nonresponding pneumonia          | BALF                                  | BGISEQ-100 | Bacterial/mycoplasma/chlamydia: SDSMRN ≥ 3, if SDSMRN ≥ 3, species was reported; otherwise, the genus was reported. DNA Virus/fungus: SDSMRN ≥ 3. RNA Virus: SDSMRN ≥ 1. Parasite: SDSMRN ≥ 100. MTC: SDSMRN ≥ 1. |

**Abbreviations:** BALF, broncho-alveolar lavage fluid; ETA, endotracheal tube aspirate; IPA, invasive pulmonary aspergillosis; MTBC, Mycobacterium tuberculosis complex; MTC, Mycobacterium tuberculosis complex; NA, no accessible; SDSMRN, the number of reads stringently mapped to pathogen species; SMRN, stringent map read number; WIMP, “What’s In My Pot” pipeline.
Compliance with ethics requirements
Not applicable.

CRediT authorship contribution statement
Zhenli Diao: Writing – original draft. Dongsheng Han: Writing – review & editing. Rui Zhang: Writing – review & editing. Funding acquisition. Jiming Li: Writing – review & editing. Funding acquisition.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References
[1] Estimates of the global, regional, and national morbidity, mortality, and [[21]]
[2] Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol 2019;37(7):783–92. doi:https://doi.org/10.1038/s41587-019-0156-5
[3] Zinter MS, Dvorak CC, Mayday MY, et al. Pulmonary Metagenomic Sequencing When Applied to Clinical Practice. Front Cell Infect Microbiol. 2020;10:182. doi:
[4] Tsalik EL, Henao R, Nichols M, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. Sci Trans Med. 2016 Jan 20;8(32):322a11. doi: 10.1126/scitranslmed.aal6683. PubMed PMID: 26791949; PubMed Central PMCID: PMCPMC6716032. eng.
[5] Babiker A, Bradley HL, Stittleburg VD, et al. Metagenomic sequencing to detect Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. Am J Respir Crit Care Med. 2018 Feb 15;197(4):524-528. doi: 10.1164/rccm.201706-1097LE. PubMed PMID: 28668513; PubMed Central PMCID: PMCPMC5821905. eng.
[6] van Boheemen S, van Rijn AL, Pappas N, Carbo EC, Vorderman RHP, Sidovar I, et al. A Prospective, Multicenter, Multisite Study of Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric Patients. J Mol Diagn 2020;22(2):196–207. doi: 10.1016/j.jmoldx.2019.10.007
[7] Yang M, Courisuea A, Liu X, et al. Direct Metatranscriptome RNA-seq and Multiplex RT-PCR Amplicon Sequencing on Nanopore MinION - Promising Strategies for Multiplex Identification of Viable Pathogens in Food. Frontiers in microbiology. 2020;11:514. doi: 10.3389/fmicb.2020.00514. PubMed PMID: 32328039; PubMed Central PMCID: PMCPMC716032. eng.
[8] Tsalk EL, Henao R, Nichols M, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. Sci Trans Med. 2016 Jan 20;8(32):322a11. doi: 10.1126/scitranslmed.aal6683. PubMed PMID: 26791949; PubMed Central PMCID: PMCPMC6716032. eng.
[9] Zhenli Diao: Writing – original draft. Dongsheng Han: Writing – review & editing. Rui Zhang: Writing – review & editing. Funding acquisition. Jiming Li: Writing – review & editing. Funding acquisition.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References
[1] Estimates of the global, regional, and national morbidity, mortality, and [[21]]
[2] Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol 2019;37(7):783–92. doi:https://doi.org/10.1038/s41587-019-0156-5
[3] Zinter MS, Dvorak CC, Mayday MY, et al. Pulmonary Metagenomic Sequencing When Applied to Clinical Practice. Front Cell Infect Microbiol. 2020;10:182. doi:
[4] Tsalik EL, Henao R, Nichols M, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. Sci Trans Med. 2016 Jan 20;8(32):322a11. doi: 10.1126/scitranslmed.aal6683. PubMed PMID: 26791949; PubMed Central PMCID: PMCPMC6716032. eng.
[5] Babiker A, Bradley HL, Stittleburg VD, et al. Metagenomic sequencing to detect Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. Am J Respir Crit Care Med. 2018 Feb 15;197(4):524-528. doi: 10.1164/rccm.201706-1097LE. PubMed PMID: 28668513; PubMed Central PMCID: PMCPMC5821905. eng.
[6] van Boheemen S, van Rijn AL, Pappas N, Carbo EC, Vorderman RHP, Sidovar I, et al. A Prospective, Multicenter, Multisite Study of Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric Patients. J Mol Diagn 2020;22(2):196–207. doi: 10.1016/j.jmoldx.2019.10.007
[7] Yang M, Courisuea A, Liu X, et al. Direct Metatranscriptome RNA-seq and Multiplex RT-PCR Amplicon Sequencing on Nanopore MinION - Promising Strategies for Multiplex Identification of Viable Pathogens in Food. Frontiers in microbiology. 2020;11:514. doi: 10.3389/fmicb.2020.00514. PubMed PMID: 32328039; PubMed Central PMCID: PMCPMC716032. eng.
[8] Tsalk EL, Henao R, Nichols M, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. Sci Trans Med. 2016 Jan 20;8(32):322a11. doi: 10.1126/scitranslmed.aal6683. PubMed PMID: 26791949; PubMed Central PMCID: PMCPMC6716032. eng.
[9] Zhenli Diao: Writing – original draft. Dongsheng Han: Writing – review & editing. Rui Zhang: Writing – review & editing. Funding acquisition. Jiming Li: Writing – review & editing. Funding acquisition.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References
[1] Estimates of the global, regional, and national morbidity, mortality, and [[21]]
[2] Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol 2019;37(7):783–92. doi:https://doi.org/10.1038/s41587-019-0156-5
[3] Zinter MS, Dvorak CC, Mayday MY, et al. Pulmonary Metagenomic Sequencing When Applied to Clinical Practice. Front Cell Infect Microbiol. 2020;10:182. doi:
[4] Tsalik EL, Henao R, Nichols M, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. Sci Trans Med. 2016 Jan 20;8(32):322a11. doi: 10.1126/scitranslmed.aal6683. PubMed PMID: 26791949; PubMed Central PMCID: PMCPMC6716032. eng.
[5] Babiker A, Bradley HL, Stittleburg VD, et al. Metagenomic sequencing to detect Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. Am J Respir Crit Care Med. 2018 Feb 15;197(4):524-528. doi: 10.1164/rccm.201706-1097LE. PubMed PMID: 28668513; PubMed Central PMCID: PMCPMC5821905. eng.
[6] van Boheemen S, van Rijn AL, Pappas N, Carbo EC, Vorderman RHP, Sidovar I, et al. A Prospective, Multicenter, Multisite Study of Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric Patients. J Mol Diagn 2020;22(2):196–207. doi: 10.1016/j.jmoldx.2019.10.007
[7] Yang M, Courisuea A, Liu X, et al. Direct Metatranscriptome RNA-seq and Multiplex RT-PCR Amplicon Sequencing on Nanopore MinION - Promising Strategies for Multiplex Identification of Viable Pathogens in Food. Frontiers in microbiology. 2020;11:514. doi: 10.3389/fmicb.2020.00514. PubMed PMID: 32328039; PubMed Central PMCID: PMCPMC716032. eng.
[8] Tsalk EL, Henao R, Nichols M, et al. Host gene expression classifiers diagnose acute respiratory illness etiology. Sci Trans Med. 2016 Jan 20;8(32):322a11. doi: 10.1126/scitranslmed.aal6683. PubMed PMID: 26791949; PubMed Central PMCID: PMCPMC6716032. eng.
[9] Zhenli Diao: Writing – original draft. Dongsheng Han: Writing – review & editing. Rui Zhang: Writing – review & editing. Funding acquisition. Jiming Li: Writing – review & editing. Funding acquisition.
European Society of Clinical Microbiology. 2021;40(1):95–102. doi: https://doi.org/10.1007/s10096-020-03996-4

[80] Ardui S, Ameur A, Vermeesch JR, et al. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic acids research. 2018 Mar 16;46(5):2159-2168. doi: 10.1093/nar/gky066. PubMed PMID: 29401301; PubMed Central PMCID: PMCPMC5861413. eng.

[81] Zhang Y, Ai JW, Cui P, et al. A cluster of cases of pneumocystis pneumonia identified by shotgun metagenomics approach. J Infect. 2019 Feb;78(2):158-169. doi: 10.1016/j.jinf.2018.08.013. PubMed PMID: 30149030; eng.

Dr. Jinming Li is the deputy director of National Center for Clinical Laboratories and director of clinical molecular and immune department. He is a doctoral supervisor of Peking Union Medical College and Peking University Medical Department. Since the mid-1990s, he has systematically put forward the concept and method of quality management and standardization of clinical molecular diagnosis in China. His research direction is clinical molecular diagnostic methods and standardization. He has successively undertaken 7 National Natural Science Foundation and 2 national major special projects as project leader. As the first and corresponding author, he has published more than 200 papers.

Dr. Rui Zhang received M.D. degrees from Peking Union Medical College. She is a master tutor of Peking Union Medical College, Peking University Medical College and Chinese Academy of Sciences. She is working at National Center for Clinical Laboratories, China. Her research interest revolves clinical molecular detection methods and standardization. As the project leader, she presides over 10 projects including general projects of National Natural Science Foundation of China and top-notch young talents of national ten thousand talents program. She has published more than 40 papers in Clinical Chemistry and other famous academic journals.

Zhenli Diao received B.S. degrees from Qingdao University, China in 2019. She is currently a M.M. student under the supervision of Prof. Jinming Li in Peking University Health Science Center, China. Her major is diagnostics of clinical laboratory. Her research area is metagenomics next-generation sequencing and molecular biology.