

# High-Throughput Metagenomics for Identification of Pathogens in the Clinical Settings

Na Li, Qingqing Cai, Qing Miao, Zeshi Song, Yuan Fang, and Bijie Hu\*

The application of sequencing technology is shifting from research to clinical laboratories owing to rapid technological developments and substantially reduced costs. However, although thousands of microorganisms are known to infect humans, identification of the etiological agents for many diseases remains challenging as only a small proportion of pathogens are identifiable by the current diagnostic methods. These challenges are compounded by the emergence of new pathogens. Hence, metagenomic next-generation sequencing (mNGS), an agnostic, unbiased, and comprehensive method for detection, and taxonomic characterization of microorganisms, has become an attractive strategy. Although many studies, and cases reports, have confirmed the success of mNGS in improving the diagnosis, treatment, and tracking of infectious diseases, several hurdles must still be overcome. It is, therefore, imperative that practitioners and clinicians understand both the benefits and limitations of mNGS when applying it to clinical practice. Interestingly, the emerging third-generation sequencing technologies may partially offset the disadvantages of mNGS. In this review, mainly: a) the history of sequencing technology; b) various NGS technologies, common platforms, and workflows for clinical applications; c) the application of NGS in pathogen identification; d) the global expert consensus on NGS-related methods in clinical applications; and e) challenges associated with diagnostic metagenomics are described.

global sequencing market.<sup>[3]</sup> Nevertheless, the development of each generation of sequencing technology has served to significantly improve a myriad of fields in genome and medical research, drug development, infectious diseases, etc. (Table 1).

## 1.1. First-Generation Sequencing Technologies

The development of the first-generation sequencing technologies began in the 1970s. In 1975, Sanger and Coulson established a simple and rapid method for determining DNA sequences using primed synthesis with DNA polymerase.<sup>[4]</sup> In 1977, Maxam and Gilbert further developed a DNA sequencing technique via chemical degradation.<sup>[5]</sup> Both of these initial platforms relied on manual sequencing. Whereas the development of the fluorescence detection technology and capillary electrophoresis technologies catapulted DNA sequencing into the era of automatic sequencing. At the end of the 20th century, automatic plate electrophoresis sequencing and capillary sequencing were introduced, with the latter

## 1. History of Sequencing Technology

Since the development of first-generation Sanger sequencing in 1977, DNA sequencing technology has made considerable strides.<sup>[1]</sup> From the first-generation to the fourth-generation, the length of sequenced reads has changed from long to short, while from the second to the fourth-generation, the read length has changed from short to long. Although the third- and fourth-generation sequencing technologies have developed rapidly,<sup>[2]</sup> the current next-generation short-read and long-read sequencing technologies maintain a dominant position in the

playing a particularly important role in the sequencing work performed for the Human Genome Project (HGP).<sup>[6]</sup>

The primary feature of first-generation sequencing technology, known as the gold standard method, is the ability to read DNA with a length of  $\approx 1000$  bp with 99.999% accuracy. However, its high cost (half a dollar per 1000 bases), low throughput (96–384 samples per round of sequencing), among other disadvantages, significantly impeded its large-scale application. Nevertheless, Sanger sequencing continues to be used for whole genome sequencing of pathogens with small genomes, such as human influenza A viruses, and Zika virus, as well as for bacteria 16S rRNA gene sequencing and fungi 18S rRNA gene sequencing.<sup>[7–10]</sup> It is also used to verify the authenticity of microorganisms detected by next-generation sequencing (NGS). For example, in 2017, Chinese scientists detected Pseudorabies Virus in vitreous humor for the first time using NGS technology, while Sanger sequencing was used to further verify their findings.<sup>[11]</sup>

## 1.2. Second-Generation Sequencing Technologies

Second-generation sequencing technologies, also commonly referred to as NGS or high-throughput sequencing

N. Li, Q. Miao, Prof. B. Hu  
Department of Infectious Diseases  
Zhongshan Hospital  
Fudan University  
Shanghai 200032, China  
E-mail: hu.bijie@zs-hospital.sh.cn  
Q. Cai, Z. Song, Dr. Y. Fang  
Genoxor Medical Science and Technology Inc.  
Zhejiang 317317, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/smt.202000792>.

DOI: 10.1002/smt.202000792

**Table 1.** Overview of high throughput sequencing platform parameters in pathogens detection.

Platform	Method	Read length [bp]	Read type	Advantages	Disadvantages	Refs.
<b>Second-generation sequencing</b>						
Illumina MiSeq	Sequencing by synthesis and reversible termination	150, 250	PE	The read length is relatively long	The sequencing results were obtained in 24–36 h, low throughput	[196,215]
Illumina NextSeq500/550		75, 150	SE, PE	The sequencing results were obtained in 11–29 h	Short read length	[25,26,236]
Illumina HiSeq 4000		125, 150	PE	High throughput and long read length	The sequencing results were obtained in 84 h	[21,25]
Illumina NovaSeq 6000		150	PE	High throughput and long read length	The sequencing results were obtained in 40 h	[237,238]
BGISeq-50	Combinatorial probe anchor ligation and DNA nanoball	50	SE	Low cost, received medical device certification clearance	Short read length	[24,132,239]
BGISeq-100		50	SE	Low cost		[29,30]
MGISeq-2000		100	PE	Low cost, received medical device certification clearance		[240,241]
<b>Third-generation sequencing</b>						
PacBio Sequel	Single molecule real time (SMRT) sequencing	1–1.8 kb	SE	Long read length	Low accuracy and high cost	[242]
Oxford Nanopore MinION	Nanopore sequencing	/	1D	Long read length and rapid sequence time	Low accuracy and high cost	[174]

technologies, improved upon the low throughput issue associated with first-generation sequencing methods. Since the year 2000, massively parallel sequencing technologies have made high-throughput sequencing possible. Several developments such as pyrosequencing, reversible terminator chemistry sequencing, as well as supported oligonucleotide ligation and sequencing have greatly increased throughput. Simultaneously, the cost of sequencing has been greatly reduced with time.<sup>[12]</sup> The most common current applications of NGS in diagnostic microbiology laboratories include targeted NGS with different methods for enrichment including amplification or probe hybridization, and metagenomics next-generation sequencing (mNGS). Currently, the most widely employed second-generation sequencer in clinical settings is manufactured by Illumina, followed by Beijing Genomics Institute (BGI).

### 1.2.1. Illumina Sequencing by Synthesis

Illumina-based platforms perform sequencing using a synthesis and reversible terminator chemistry strategy,<sup>[13,14]</sup> in addition to the bridge amplification method, whereby single molecules of DNA are first attached to a flow cell and subsequently amplified locally into a clonal cluster.<sup>[13]</sup> Next, a sequencing by synthesis reaction occurs, in which synthesis of complementary DNA occurs via addition of one nucleotide per cycle, and the optical readout of the fluorescently labeled nucleotides then determines its identity (A, G, T, or C).<sup>[13]</sup> Illumina offers a popular series of platforms, including HiSeq, MiniSeq, MiSeq, NextSeq, and NovaSeq, which are all suitable for ultralarge-scale sequencing and to meet the needs of different

sequencing scales and applications.<sup>[15–18]</sup> The Illumina MiSeq instrument was the first platform to be applied for clinical pathogen detection after it was successfully used to diagnosis *Leptospira* infection within 48 h in a boy with severe comprehensive immunodeficiency. Since this time, a new era has begun for the clinical application of mNGS.<sup>[19]</sup> Specifically, the HiSeq series platform is generally used to carry out clinical pathogenic detection on large sample sizes due to its high-throughput and relatively long sequencing read length, however, each run can take up to 3.5 days, which is not suitable for the rapid detection of single samples.<sup>[20–22]</sup> Alternatively, NextSeq500 and NextSeq550 series sequencing systems offer the advantages of moderate throughput and short running time (12–30 h per run), making them particularly suitable for the detection of clinical pathogens, for which they are widely used.<sup>[23–26]</sup>

### 1.2.2. BGI Sequencing

Sequencing using the combinatorial probe anchor ligation (cPAL) and DNA nanoballs (DNB) methods is the core of BGI.<sup>[27]</sup> Single-stranded circular DNA is amplified using rolling circle amplification (RCA) by 2–3 orders of magnitude; the amplification products are designated, DNB. The nanoballs are then fixed on an arrayed silicon chip using DNB loading technology, which relies on joint probe anchoring polymerization. DNA molecular anchors and fluorescent probes are polymerized on the DNB, and the resulting optical signals are collected by a high-resolution imaging system and digitally processed to obtain the sequence of interest.<sup>[28]</sup> The BGI sequencing platform is being increasingly used in clinical pathogen mNGS detection due to its low cost, short sequencing time and other

advantages. BGISEQ sequencers are most commonly employed to detect various pathogens, including those causing focal infections, *Mycobacterium tuberculosis* (MTB), etc.<sup>[24,29,30]</sup>

The development of second-generation sequencing technologies has not only greatly promoted the development of biological and medical research, but has also ushered in a new era for the clinical detection of pathogens. However, regardless of the platform or technology used, the resulting read lengths using mainstream methods are short at  $\approx 50\text{--}300$  bp, and subsequent analysis depends on fragment splicing, which may introduce errors. Moreover, PCR amplification is performed during library preparation to amplify low amounts of starting DNA and to aid sequence enrichment. However, sequences with low content may not be amplified, resulting in the loss of certain information. Further, the run time required for second-generation sequencing is long with most sequencers requiring more than 12 h, hence, the sequencing data cannot be obtained in real-time. Hence, there remains room for improvement in these sequencing technologies.

### 1.3. Third-Generation Sequencing Technology

The third-generation sequencing technology is also known as single-molecule sequencing technology.<sup>[31]</sup> Compared with NGS, third-generation sequencing technologies can detect ten nucleotides per second, which greatly reduces the time required for sequencing. Moreover, the full-length mRNA sequence can be derived from the ultralong read lengths obtained. Third-generation sequencers can also directly sequence the original DNA/RNA samples without requiring PCR amplification, with no preference for CG nucleotides, and can directly detect and obtain methylation information. However, despite the many advantages of third-generation sequencing, it has not been widely adopted in clinical applications due to its high error rate and high associated cost.<sup>[2]</sup>

#### 1.3.1. Single-Molecule Real-Time (SMRT) Sequencing

SMRT sequencing is a representative of third-generation sequencing technology that applies the principle of sequencing by synthesis. SMRT cell contains nanoscale zero-mode waveguide (ZMW) wells. Each ZMW contains a DNA polymerase molecule and DNA sample chain, which are template fragments that are processed and ligated to hairpin adapters at each end, resulting in a circular DNA molecule with constant single-stranded DNA (ssDNA) regions at each end, and the double-stranded DNA (dsDNA) template in the middle. This platform allows for single-molecule sequencing and real-time detection of the fluorescence signal from the nucleotides. Incorporation of nucleotides into the growing chain results in the dissociation of the labeling groups, which reduces steric hindrance and helps maintain the continuous synthesis of DNA chains, prolonging the sequence read length.<sup>[32,33]</sup>

Additionally, the newly launched SMRT Sequel series by PacBio provides an average read length of 10–20 kb, and can achieve sequencing throughputs of 160 GB. However, in spite of these features, the bulky equipment and expensive hardware

make this system unpopular for clinical diagnosis of pathogens. Nevertheless, during the novel coronavirus pandemic, the PacBio system has been employed numerous times for whole genome sequencing and reassembly of SARS-CoV-2.<sup>[34,35]</sup>

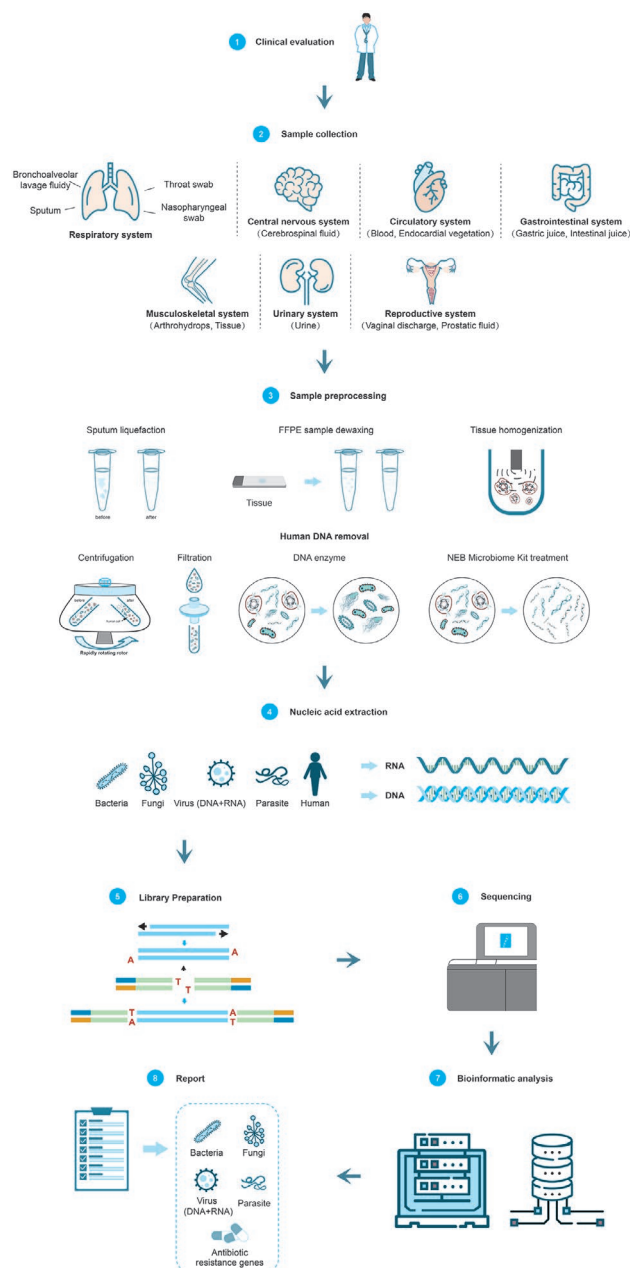
#### 1.3.2. Nanopore Sequencing

Nanopore sequencing technology is another representative third-generation sequencing technology, however, it is also considered to be a fourth-generation platform as it can perform real-time data acquisition and analysis.<sup>[36]</sup> A core component of nanopore sequencing is the use of nanoscale pores composed of transmembrane proteins, which is a significant departure from the previous sequencing technologies.<sup>[37,38]</sup> In 2012, Oxford Nanopore Technologies released the GridION and MinION sequencing platforms.<sup>[39]</sup> MinION is a mobile single-molecule nanopore sequencing device that is 4 in. long and can be connected through a USB3.0 port to a laptop computer. It is, therefore, conveniently portable and meets DNA sequencing needs under several conditions. Notably, nanopore DNA sequencing is orders of magnitude faster than other strategies for the diagnosis of infectious diseases. The data output of a single flow cell in the platform is 10–25 Gb, with the sequencing time ranges from 0 to 48 h, and read lengths of 800 kb can be achieved, through the error rate can be as high as 5–40%.<sup>[40–42]</sup> Due to the advantages of nanopore sequencing technology, it has been widely adopted in the field of epidemic outbreak investigation to detect infectious pathogens, antimicrobial resistance, as well as other infectious areas of concern.<sup>[43–46]</sup> However, the high number of sequencing errors, and higher per-read costs compared to other NGS platforms may limit its utility for certain applications.

## 2. Sequencing Methods and Bioinformatic Analysis

The process of high-throughput sequencing of pathogens primarily includes two components: experimental manipulations (wet lab) and bioinformatic analysis (dry lab). The wet lab manipulations include sample pretreatment, nucleic acid extraction, library construction, and sequencing. The dry lab bioinformatic analysis includes quality control of data, removal of human sequences, sequence alignment of sequences from microbial species, and analysis of genes for drug resistance or virulence.

However, the diagnosis of infectious diseases requires that prior to pretreatment of samples, specific samples must first be collected from the primary site of infection. For example, bronchoalveolar lavage fluid (BALF) and sputum are generally recommended for pulmonary infection, while cerebrospinal fluid (CSF) is recommended for central nervous system (CNS) infection. Although the library construction, sequencing and bioinformatics analysis of different samples are the same, pretreatment and nucleic acid extraction differs depending on the sample source. The following section will describe the mNGS detection process for samples suitable for different infection systems (Figure 1).



**Figure 1.** mNGS workflow in clinical application. This workflow consists of eight components. 1) Clinical evaluation: the test is suitable for patients with infectious diseases. 2) Sample collection: collecting samples from the primary site of infection can greatly increase the detection rate. 3) Sample preprocessing: the pretreatment methods for different types of samples are different, sputum needs liquefaction treatment, FFPE samples are dewaxed, and tissue needs homogenization. The percentage of human DNA in samples can be reduced using methods such as filtration, differential centrifugation, DNA enzymatic hydrolysis, and methylation reagent treatment. 4) Nucleic acid extraction: there are differences between DNA and RNA extraction. 5) Library preparation: library construction method is selected according to the sequencing platform and purpose. 6) Sequencing: at present, the mainstream second-generation sequencing platforms are produced by Illumina and BGI. 7) Bioinformatic analysis: based on the analysis of the raw data, the information of species and antibiotic resistance genes in the samples were obtained. 8) Report: the possible pathogens were screened out according to the analysis results.

## 2.1. Sample Preprocessing

### 2.1.1. Sputum

For respiratory tract infections, the most common samples are BALF and sputum. However, the fiberoptic bronchoscopy procedure, required for collection of BALF, is invasive and often deemed intolerable by patients, thus sputum becomes an alternative sample.<sup>[21]</sup> Sputum samples are generally highly viscous and heterogeneous; additionally, the distribution of microorganisms in sputum is uneven. Therefore, sputum liquefaction treatment is performed prior to extracting nucleic acids from the samples. The existing sputum liquefaction schemes include the following methods: i) standard *N*-acetyl-L-cysteine treatment and digestion with 2% NaOH and decontamination; ii) NaOCl liquefaction and sedimentation; iii) chitin solution liquefaction; iv) the use of sputum and an equal volume of phosphate buffer containing  $1 \text{ g L}^{-1}$  protease K to liquefy sputum; and v) classical dithiothreitol (DTT) liquefaction, in which a chitin solution homogenizes the mucus sputum more quickly than the *N*-acetyl-L-cysteine and NaOCl methods.<sup>[47–49]</sup> With the advancement of automated processing, a portable, low-power preprocessing device can perform all of the required steps for sputum preprocessing, including liquefaction, homogenization, dissolution, and inactivation.<sup>[50]</sup>

### 2.1.2. Tissue Sample

Tissue samples account for one of the most common types of clinical samples as they allow for the detection of various infections, including lung, brain, skin, and soft tissue. For patients who require surgery, the relevant tissues can be obtained during the operation without requiring additional sampling steps.<sup>[51–53]</sup> Meanwhile, formalin fixed paraffin embedded (FFPE) samples can introduce contaminating microorganisms and cause damage to DNA and RNA due to the complex processing steps required. Therefore, although these samples can be preserved for an extended period of time and are important for the detection of infectious diseases, they can be difficult to process and analyze effectively. FFPE samples are first dewaxed by treating  $25 \text{ }\mu\text{m}$  thick sections in  $1.5 \text{ mL}$  tubes with  $120 \text{ }\mu\text{L}$  of xylene for  $10 \text{ min}$  twice, followed by  $1200 \text{ }\mu\text{L}$  of 100% ethanol for  $10 \text{ min}$  three times with constant gentle agitation for deparaffination.<sup>[54]</sup> Alternatively, fresh tissue is commonly used for analysis, which requires tissue homogenization. The FastPrep-24 Instrument can be used for tissue fragmentation, after which the centrifuged supernatant is then used for further experiments.<sup>[55]</sup> Tissue digestion is performed using a lysis buffer composed of 0.5% Tween 20, 2 mg proteinase K,  $3.5 \times 10^{-3} \text{ M}$   $\text{MgCl}_2$ ,  $15 \times 10^{-3} \text{ M}$  ammonium sulfate, and  $60 \times 10^{-3} \text{ M}$  Tris-HCl at  $56 \text{ }^\circ\text{C}$  for  $1 \text{ h}$  to obtain transparent cleavage products.<sup>[54]</sup> Although tissue samples often have high human cell background content, their sequences can be eliminated from analysis by using an appropriate dehosting process to improve the sensitivity of overall detection.

## 2.2. Nucleic Acid Extraction

Nucleic acid extraction is an important step in the mNGS detection process. Sufficient DNA or RNA of appropriate quality is



extracted from cells for downstream processing. To achieve this, various sample extraction strategies are employed. The raw input varies based on the source and type of sample, as well as the method of preprocessing used.<sup>[56]</sup> The extraction strategy for cell-free DNA (cfDNA) or cell-free RNA (cfRNA) is generally used in samples from peripheral blood.<sup>[57]</sup> Accordingly, each commercial vendors often offers several different kits for manual or automated extraction using liquid-handling robots. After dewaxing, DNA from FFPE samples can be extracted using nucleic acid extraction kits suitable for different target microorganisms.<sup>[54]</sup> With the advancement of nucleic acid extraction kits, mature commercial kits (such as the Molzym Ultra-Deep Microbiome Prep kit) are now suitable for DNA extraction from various types of samples, including biopsies, which are capable of removing host DNA.<sup>[54,58]</sup>

### 2.3. Library Preparation and Sequencing

Currently, several platforms and instruments are available for mNGS, each with different schemes for library construction, the complexity of which is greater for second-generation sequencing libraries compared to third-generation. In this section, we introduce the commonly used database construction schemes for Illumina, BGI, Oxford Nanopore technologies, and the PacBio sequencing platforms.

The Illumina sequencing platform is currently the most widely used in the field of mNGS. Its library construction process involves the random breakage of genomic DNA or double-stranded cDNA (obtained by RNA reverse transcription) into small fragments by enzymatic cleavage or using ultrasound, enzyme-based flattening of both ends of the DNA molecule, addition of an A base to the 3'-end using the Klenow enzyme, and adapter ligation.<sup>[59]</sup> However, the reliance of all existing RNA sequencing on RNA reverse transcription to produce cDNA, followed by second-strand synthesis via additional enzymes and purification steps, can introduce sequence-dependent bias. Recently, Chinese scientists combined the Tn5 transposase with transposon sequencing to develop a new rapid library construction method, called the SHERRY method. Compared to existing methods, it significantly simplifies the process of database construction and reduces the initial template input, which is relevant for samples with low starting concentrations, including that of coronaviruses.<sup>[60,61]</sup>

Certain similarities and differences exist between the BGISEQ technologies and the Illumina sequencing platforms. In both cases, five sample preparation steps are required: sample fragmentation, end repair, addition of an A nucleotide, adaptors addition, and PCR; however, the specific experimental steps performed and the enzymes used are quite different. In the BGI prescribed method, the DNA overhangs are filled in to form blunt-ended molecules, and a sequence of connectors is added to the two DNA fragments ends to connect them, forming a ring structure. A special molecule of cyclic DNA is then thermally denatured together, which is added in reverse to a special chain of the PCR product, and the single-stranded molecule is connected by DNA ligase. The remaining single-stranded linear molecules are digested by exonuclease to obtain a single-stranded circular DNA library. PMseq was sequenced

by single-stranded ring DNA amplification to form DNA nanospheres (DNB), which are fixed on arrayed silicon chips by rolling ring amplification (RCA).<sup>[62,63]</sup>

Oxford Nanopore DNA library construction methods can be divided into three categories: 1D, 1D2, and 2D methods.<sup>[64]</sup> The 1D2 library construction method is an improvement over the 1D method, and though a higher throughput can be obtained using the 1D library, a higher sequence accuracy is achieved with the 1D2 library.<sup>[37,65]</sup> Further, libraries constructed using the Nanopore kits do not require the reverse transcription of RNA into cDNA, as it is possible to directly sequence the single RNA strand. Therefore, construction of the RNA libraries can be divided into two categories based on end-use: 1) direct sequencing; 2) sequencing of the cDNA library by reverse transcription. The time required to construct the two libraries is similar (115 min vs 125 min, respectively); however, direct sequencing requires ~500 ng of input, RNA template, whereas the cDNA library only needs ~50 ng.<sup>[66,67]</sup>

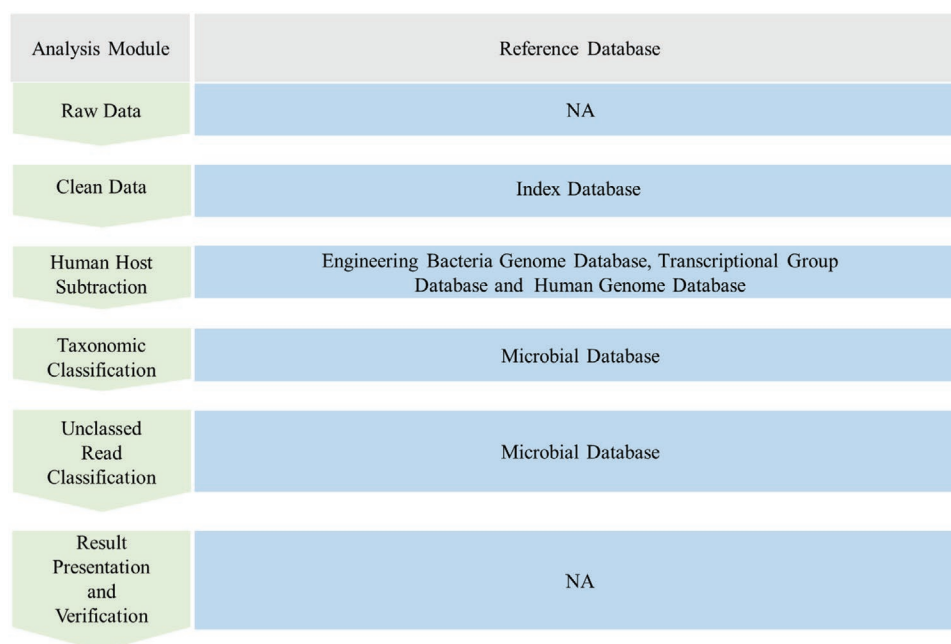
The SMRT bell libraries used in PacBio sequencers are dumbbell-shaped and offer the advantage of circularizing whole molecules. Hence, the molecules can be sequenced repeatedly, allowing for the generation of long-read lengths. The dumbbell-shaped molecules are composed of two parts: the hairpin adapter and the dsDNA template. Construction of the sequencing library involves DNA fragmentation to obtain dsDNA fragments with sticky ends, and subsequently, AMPure PB magnetic beads used for size selection and purification of the target fragments. Next, damage repair and end repair of the DNA template is performed to enable its use in direct sequencing and to avoid negative effects on read length. Subsequently, the flat end hairpin joints are connected to the DNA template, which is then purified and quantified.<sup>[68]</sup> Additionally, Coupland et al.<sup>[69]</sup> reported that small DNA molecules can be sequenced directly from 1 ng of DNA without requiring standard library preparation.

### 2.4. Bioinformatic Analysis

Bioinformatic analysis is the final step required for mNGS-based detection of pathogenic microorganisms, which enables their rapid and accurate identification. Pathogenic infections require varying treatments, making it imperative that their detection is accurate, direct, and verifiable. Errors and contamination can be introduced throughout the process during sampling, assaying, or sequencing of samples, as well as by the instrument consumables or production environment. Errors may also be introduced during bioinformatic analysis, resulting in false-negative or false-positive results. From the sequencing read-based identification of a species, it is necessary to optimize and verify every step of the process, including quality control, human host subtraction, microbial identification, verification, etc. (Figure 2).

#### 2.4.1. Raw Data

Raw sequencing data can be generated using different platforms, for which the first in silico analysis step involves the separation of data for each sample into independent files per index of the individual sample library. The sequenced reads are



**Figure 2.** The bioinformatics analysis process starts with the fastq data, including the removal of low-quality and low-complex sequences, host and engineering bacteria sequences, and the identification of pathogens. Samples qualified for sequencing (Q30 qualified), the proportion of removing low-quality and low-complex sequences is about 5%, which accounts for about 20% of the whole process time. Removal of hosts, engineering bacteria, plasmids, different types of samples and different treatment methods, and the host proportion varies greatly, between 40% and 99%; this step accounts for about half of the whole process time. For taxonomic classification, different alignment software is quite different, such as k-mer algorithm, there are 30% to 80% of the sequences, which can be assigned to microbes. The proportion of unclassified sequences is high, so can choose to compare with NR database to identify distant sequences. This step accounts for about 30%.

usually retrieved as FASTQ files. Particular attention must be paid to the amount of sequencing data for each sample (at least 10 Mb is currently required) and the quality of the sequenced reads (Q20/Q30). The sequencing quantity and quality should be of an acceptable standard before further processing, otherwise additional data may be required.

#### 2.4.2. Clean Data

The next step in the analysis of sequencing data requires removal of low-quality reads, which may interfere with the downstream analysis or cause false positives. The quality control process includes trimming of reads to remove adaptors/adaptor trimming, quality filtering of reads, removal of low-quality reads, removal of short reads, discarding reads shorter than 36 nucleotides, low-complexity read filtering, and removal of duplicate reads. **The Trimmomatic or Fastp tools are commonly used for the removal of low-quality reads, adaptors, and short reads.**<sup>[70,71]</sup> Additionally, tools such as DustMasker are used to mask low-complexity regions, **while duplicate reads can be removed using PRINSEQ**, and FastQC can be used to graphically analyze the data for each sample.<sup>[72]</sup>

#### 2.4.3. Human Host Subtraction

The presence of contaminating human DNA is common in the sequencing results for various microbes, making it necessary

to remove interference from the host sequence to shorten the analysis time. To this end, sequence comparison with the human reference genome can be made, and the host sequence removed, though software and reference genome selection does affect the degree to which host DNA removal is effective.

Next, sequence alignments are performed relative to the reference genome using mapping software, such as Bowtie 2, BWA, HISAT2, etc.<sup>[73–75]</sup> HISAT2 is based on the HISAT and Bowtie2 implementations, small indexes, and alignment speed. These tools are widely used in genome resequencing and RNASeq data analysis, providing alignment accuracy, optimization of called mutations, and discarding unmapped reads from the final output. Software such as BMTagger and CS-SCORE are specifically used to remove host DNA,<sup>[76,77]</sup> compared to Bowtie2 and BWA, these programs offer better host DNA removal rates.<sup>[78]</sup> Further, BMTagger is the standard operating procedure tool used for the Human Microbiome Project.

The human reference genome (current version: GRCh38.p13, release date 2020.04) consists of chromosomal scaffolds, unplaced scaffolds, and “alternate” scaffolds, with the latter representing sequences that are divergent from the primary chromosome sequence. The earliest release was in October 2014; however, the previously released version GRCh37 from April 2011 (updated in December 2013; UCSC version hg19) is also widely used as annotation libraries are available. When removing host DNA, priority should be given to the latest reference genome version. It is advisable to not use versions optimized for mapping, which include “traps,” Hepatitis B virus sequences, and chromosome optimizations. For more efficient

removal of the host sequences, the human mRNA reference sequence of human can be added to the reference genome simultaneously. Further, given the differences between populations, different versions of the human genome can be added to the host database for detection of DNA sequences from different populations, as well as sequences from bacteriophages, plasmids, engineered bacteria, and potentially contaminating bacteria, to remove the corresponding interfering sequences.

#### 2.4.4. Reference Databases

For taxonomic classification, alignments are performed relative to reference databases regardless of the algorithm used. Various reference genomes can be used for library comparisons, including the National Center for Biotechnology Information (NCBI) nonredundant nucleotide sequence database (NT/NR), NCBI Reference Sequence (RefSeq) and genome database, NCBI GenBank database, and microbial reference sequences or genome collections.<sup>[79]</sup> Based on the sequence similarity and algorithm, the NT/NR library and reference genome library (with higher assembly level) can be selected based on marker specific region recognition. Further, a k-mer counting algorithm is recommended, as it is fault-tolerant and enables the analysis of microbial genome sequences, which is helpful to improve the accuracy and positive rate of identification.<sup>[80]</sup>

Microbial species are extremely diverse and include viruses, bacteria, fungi, and parasites. To enable their identification, it is necessary to analyze their genomes and construct a microbial reference genome database. However, the construction of reference databases should consider potential discrepancies in the study, as in some cases, the similarity between species is greater than that between strains of the same species when classification is based on morphology, as shown by a comparison of genomic differences.<sup>[81]</sup> Further, differences may exist in the reference genome sequence due to variations in genome quality, including at the contig, scaffold, and whole genome levels. Hence, the various methods available for database construction must be carefully analyzed for choice of sequencing platform, sequencing depth, splicing software, and establishment of a standard genome grouping mechanism. These methods must be combined with comparative genome analysis, phylogeny, **genome quality assessment using tools such as CheckM**, establishment of a filtering mechanism, quality assessment of contig and scaffold sequences in the genome for assessment of contaminating sequences, and removal of low-quality sequences.<sup>[82]</sup> **Additionally, publicly available curated microbial genome databases, such as FDA-Argos and the FDA Reference Viral Database (RVDB), have been made available.**<sup>[83]</sup> A combined approach that incorporates annotated sequences from multiple databases may encourage greater confidence in the sensitivity and specificity of microorganism identification.

#### 2.4.5. Taxonomic Classification

Metagenomics classification tools, including the Kraken series represented by the k-mer algorithm include Kraken, Bracken, KrakenUniq, Kraken 2, Centrifuge, and CLARK.<sup>[80,84–86]</sup> For a

certain length (k-mer length) reference genome, the sliding window technique using 1 bp window slides is used. If the k-mer appears in two species, it is assigned to the smallest taxon lowest common ancestor (LCA), and the genome sequences of all microorganisms are analyzed in turn, establishing the k-mer database. During the identification of classifier sequences, the trimming of reads is similar to the defining of k-mers for database query, and the classification of reads is determined based on the information of each k-mer. Results containing false positives are a significant challenge in metagenomics classification, which has been addressed by the KrakenUniq algorithm, which optimizes the results in the Kraken classification results. Bracken calculates the abundance of detected species based on Kraken. The Kraken 2 algorithm optimizes the storage index of k-mers, which reduces the memory required and improves the speed by reducing memory usage by 85%.<sup>[85]</sup> Alternatively, the Centrifuge algorithm uses an indexing scheme based on the Burrows–Wheeler transform (BWT) and the Ferragina–Manzini (FM) index, and is optimized specifically to resolve metagenomic classification related issues. Methods, such as MetaPhlAn2, that analyze the composition of microbial communities, use unique marker genes for classification, which can be optimized for specific groups, such as mycobacteria. Although this platform offers superior identification accuracy and more professional library construction, it is possible that certain identified reads are lost.<sup>[87]</sup>

In studies investigating macrogenomic communities, most species are unknown, and their identification is conducted by comparing with **protein sequences, using tools such as DIAMOND and Kaiju**. Samples with a large number of unclassified reads indicate that there may be unknown microbe species present in the community. This can be confirmed using the nr protein library.<sup>[88,89]</sup> The alignment-based methods, such as BLAST and Mega BLAST,<sup>[90]</sup> are widely used for the identification of taxonomic differences using sequencing data and can be regarded as the gold standard. However, due to the amount of data generated by NGS, the operation speed of such aligners for these datasets is slow, making them not ideal, and not commonly used, for this purpose. Nevertheless, the results obtained from identification analyses can be verified using such aligners.

#### 2.4.6. Report

The mNGS technique is highly sensitive, capable of readily detecting pathogens in samples, and identifying microorganisms in reagents, consumables, and the environment. However, to achieve accurate results, data must be filtered according to a certain threshold. Commonly used filtering indicators include sequencing reads and relative abundance.

The type of pathogen influences the amount of sequencing data obtained. Considering the differences in microbial genome size, the detected sequences are first standardized based on reads per million (RPM). **The threshold for viruses is generally at, or above, 3 RPM and exceeds 1000 RPM for retroviruses. For the identification of fungi, the threshold is above 5 RPM, and the credibility of the applied threshold is greater when there is an increase in the number of sequenced reads that can reach 100 RPM.** For intracellular bacteria, such as MTB

and *Legionella*, the threshold is relatively low, and credible data can be obtained with a threshold of 1 RPM. With an increase in the number of sequences, credibility gradually increases, however, upon reaching 30 RPM, further increments no longer significantly increase the credibility of the threshold. For parasites, the applied threshold should be above 10 RPM, and the specificity of the sequence should be strictly confirmed.<sup>[3,91–93]</sup>

Different thresholds may need to be set based on data from the negative control and sequence detection data. For example, in a virus, if the number of detected sequences covers more than three nonoverlapping regions of the virus reference genome, the detected sequences are first standardized as RPM. Next, the ratio (RPM-r) of the clinical sample RPM to that of the negative quality no template control (NTC) is calculated. An RPM value  $\text{RPM-r} = \text{RPM sample} / \text{NTC} \geq 10$  indicates that the species should be included in the clinical report.<sup>[22,94]</sup>

The pipeline for the evaluation and analysis of microbe testing can include DisCVR for virus identification and real-time fluorescence quantitative PCR (RT-PCR) analysis as a reference. The optimal threshold of 850 k-mers for DisCVR and 150 reads for CLARK and Kraken is appropriate.<sup>[95]</sup> Based on relative abundance, viruses in the blood samples can be filtered for detection, and in clinical samples, thresholds of 1% for relative abundance and 0.01% for total reads classified have been applied.<sup>[96]</sup>

For some bacteria, it is necessary to focus on distinguishing between colonization, contamination of samples, and the presence of real pathogens and filter them according to the background bacterial database and baseline detection. A more accurate report for pathogens in samples can be derived after excluding human microflora, contamination from the laboratory environment, and reagent consumables using the background microbial databases.<sup>[97]</sup>

To evaluate the credibility of the identification results, the reads obtained can be compared with the reference genome of the species to calculate the coverage and sequencing depth by mapping. The higher the coverage and depth, the more reliable the results. Hence, if the coverage is low, but the depth is high, the credibility for identification is relatively low. Meanwhile if both coverage and depth are very low, minimal usable data is available for identification of the species, and other detection methods, such as RT-PCR, can be used. The data can be visualized using the genome coverage curve, hot spot charts, pie charts, the Sanger chart, etc.<sup>[98]</sup>

### 3. Clinical Application of NGS

Infectious diseases remain a leading cause of human morbidity and mortality worldwide. Fast and accurate diagnosis of the etiologic pathogen can be challenging since a wide variety of microorganisms may cause clinically indistinguishable diseases, whereas the spectrum of detectable pathogens is relative narrow by current methods. Generally, a battery of combined tests, including culture, type-specific serologic assays, and nucleic acid amplification tests, are required for establishing a diagnosis. Further, these methods may be time-consuming, for example, delivering results for commonly encountered pathogens by culture can take at least 2–5 days, or even longer

(several weeks to months), for more fastidious or insidious organisms, such as mycobacteria, *Nocardia*, and fungi.<sup>[24,99]</sup> Yet many common pathogens are impossible to culture in vitro, such as viruses,<sup>[99,100]</sup> and samples may be difficult to obtain, requiring invasive biopsy.<sup>[101,102]</sup> Additionally, the administration of antibiotics may affect the sensitivity of pathogens identification by traditional methods, such as culture.<sup>[103,104]</sup> Meanwhile, the introduction of syndromic multiplex PCR panels, 16S ribosomal DNA detection, and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) have served to dramatically reduce the turnaround time for testing, however, the etiology remains unknown in up to 60% of infectious diseases.<sup>[105,106]</sup> Moreover, missed or delayed diagnosis, caused by the disadvantages of traditional microbe identification methodologies, drives the abuse of empirical broad spectrum antibiotics or antifungal drugs, thus preventing the use of targeted and curative treatments.<sup>[3]</sup> Alternatively, mNGS enables quick detection and comprehensive identification of bacteria, fungi, viruses, and parasites without the need for prior presumption of the causative culprits. Pathogens can be identified directly from various clinical specimens with a higher sensitivity and accuracy, outperforming culture-based methods, especially for mycobacteria, anaerobes, atypical pathogens, and viruses.<sup>[29,99]</sup> Furthermore, mNGS is less affected by prior antibiotic exposure.<sup>[99]</sup> Therefore, mNGS shows great potential for infectious diseases diagnosis and differential diagnoses, which has been supported by a series of studies (Table 2).

#### 3.1. Infections in Different Organ Systems

##### 3.1.1. Bloodstream Infections

The constitution of causative pathogens varies with septic patients in different clinical backgrounds, with cases of Gram-negative bacteria, anaerobes, and fungal sepsis increasing over the past decade, although Gram-positive bacteria remain the most frequently identified.<sup>[107]</sup> The causative organisms remain unidentified in approximately half of sepsis patients, namely, culture-negative sepsis. Alternatively, acute onset and severe sepsis can limit the available time available for diagnosis of the causative pathogen, contributing to high mortality rates,<sup>[108]</sup> which is further amplified in patients that receive inappropriate or mismatched antimicrobial therapy.<sup>[109]</sup> Hence, there is an urgent need for more rapid and sensitive pathogen identification technologies.

Multiple preliminary studies and case reports have shown that the genomic DNA or RNA fragments from circulating or noncirculating pathogens associated with the infections can be detected as cfDNA or cfRNA in purified plasma,<sup>[57,110–112]</sup> which allows for the use of mNGS for rapid and precise identity of pathogens causing sepsis, along with information regarding the abundance and genetic relatedness.<sup>[113,114]</sup> The diagnosis agreement of mNGS with tests based on blood culture in patients with a sepsis can be as high of 93.7% with the results are equivalent to, or superior to, those obtained by other direct molecular diagnostic methods, indicating the potential for mNGS to identify a wide range of infections, including bloodstream infection caused by various etiologies.<sup>[115]</sup> Further, multiple studies



**Table 2.** Studies of NGS for infectious diseases diagnosis.

Year	Samples and populations	Methods	Results	Conclusions	Refs.
<b>Big data studies</b>					
2018	511 specimens (blood, respiratory tract samples, body fluids and pus, CSF <sup>a</sup> ), urine, and swabs) from 561 cases suspected infections from April 2017 to December 2017.	A retrospective study	<ul style="list-style-type: none"> <li>The sensitivity and specificity of mNGS for diagnosing infectious disease were 50.7% and 85.7%, respectively, outperformed those of culture, especially for MTB<sup>a</sup>, viruses, anaerobes, and fungi.</li> <li>The sensitivity of mNGS was superior to that of culture in cases with antibiotic exposure.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS has higher sensitivity for pathogen identification.</li> <li>mNGS is less affected by prior antibiotic exposure.</li> </ul>	[99]
2019	132 clinical samples (sputum, CSF, pus, etc.) from 105 patients presenting with suspected active MTB infection between June 1, 2017 and May 21, 2018.	A prospective study	<ul style="list-style-type: none"> <li>The specificity of mNGS for diagnosing active MTB was 98%, while the sensitivity was 44%, similar to Xpert (42%) and much higher than conventional methods (29%).</li> <li>mNGS has a significantly much higher sensitivity in pretreatment samples (76%) than the post-treatment ones (31%).</li> <li>Combining Xpert and mNGS together, the identify rate increased to 60%.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS had a higher sensitivity for active MTB diagnosis.</li> <li>Combined mNGS and Xpert showed an overall superior advantage.</li> </ul>	[24]
2019	163 specimens (CSF, blood, and throat swabs) from 105 patients suspected with viral encephalitis/meningitis or respiratory infection from May 2017 to June 2019.	A prospective study	<ul style="list-style-type: none"> <li>The positive and negative percent agreement of mNGS is 65% and 95%, respectively.</li> <li>The overall percent agreement of mNGS is 81% in CSF, 68% in blood samples, and 100% in the tested throat swabs.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS has advantages for virus diagnostics.</li> <li>mNGS can be used as a supplement of current routine tests.</li> </ul>	[127]
<b>Focal infection</b>					
2019	Tissues from 98 suspected focal infection cases.	A single-center retrospective study	<ul style="list-style-type: none"> <li>mNGS showed a positive percent agreement of 86.30% in a variety of tissues, higher than culture (45.21%) and other conventional methods (57.53%).</li> <li>mNGS detected an extra 34 pathogenic microorganisms.</li> <li>Time requirement was shorter than culture.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS is a promising diagnostic tool in focal infections.</li> <li>Help to make more timely and targeted therapeutic decisions.</li> </ul>	[29]
<b>Blood stream infection and febrile illness</b>					
2016	78 plasma samples from ICU <sup>a</sup> patients, and 10 plasma samples from healthy volunteers between July 2014 and August 2014.	A prospective study	<ul style="list-style-type: none"> <li>The overall diagnostic sensitivity was significantly increased from 12.82% (10/78) by bacterial culture alone to 30.77% (24/78) by NGS alone for ICU patients.</li> </ul>	<ul style="list-style-type: none"> <li>NGS can be applied as an emerging diagnostic tool for septic patients.</li> </ul>	[57]
2019	Plasma samples from 40 returning travelers presenting with a fever of $\geq 38^{\circ}\text{C}$	A single center, proof-of-principle study	<ul style="list-style-type: none"> <li>8/40 patients were diagnosed with a viral infection by mNGS, higher than that of standard-of-care diagnostics.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS improves the diagnosis sensitivity of viral infections.</li> <li>mNGS has potential to be an all-in-one diagnostic test.</li> </ul>	[196]
2019	60 serum, 90 nasopharyngeal, and 10 stool specimens were collected from 94 children with febrile illness.	A retrospective exploratory study	<ul style="list-style-type: none"> <li>The most common microbes identified by mNGS were <i>Plasmodium falciparum</i> (51.1% of samples) and parvovirus B19 (4.4%) from serum; human rhinoviruses A and C (40%), respiratory syncytial virus (10%), and human herpesvirus 5 (10%) from nasopharyngeal swabs; and rotavirus A (50% of those with diarrhea) from stool.</li> <li>3 new viral species were also detected and identified.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS helps to identify multiple potential pathogens associated with fever.</li> </ul>	[171]

**Table 2.** Continued.

Year	Samples and populations	Methods	Results	Conclusions	Refs.
<b>Central nervous system (CNS) infections</b>					
2018	CSF samples from 99 pediatric bacterial meningitis patients.	A retrospective observational study	<ul style="list-style-type: none"> <li>mNGS showed higher sensitivity (68.7%) than culture (55.6%).</li> <li>The main pathogens identified in this study were <i>Streptococcus pneumoniae</i> (<math>n = 29</math>), group B <i>Streptococcus</i> (<math>n = 15</math>), <i>Staphylococcus aureus</i> (<math>n = 7</math>), and <i>Escherichia coli</i> (<math>n = 7</math>).</li> </ul>	<ul style="list-style-type: none"> <li>NGS could be a promising alternative diagnostic approach for critically ill pediatric patient with meningitis.</li> </ul>	[122]
2019	CSF samples from 135 pediatric bacterial meningitis cases.	A retrospective study	<ul style="list-style-type: none"> <li>Taking culture as the gold standard, the sensitivity, specificity, PPV<sup>a</sup>, and NPV<sup>a</sup> of mNGS for diagnosing <i>S. pneumoniae</i> meningitis were 73.1%, 88.1%, 59.4%, and 93.2%, respectively, and 70.3%, 93.9%, 81.4%, and 89.3%, respectively when compared with combined tests (culture and <i>S. pneumoniae</i> antigen test).</li> <li>The number of unique reads of <i>S. pneumoniae</i> in CSF sample (&lt;14 days onset) was significantly higher than CSF sample (&gt;14 days from onset).</li> </ul>	<ul style="list-style-type: none"> <li>mNGS has high sensitivity and specificity for <i>S. pneumoniae</i> identification.</li> <li>mNGS was less affected than culture by prior antibiotics exposure.</li> <li>CSF collection time affect unique reads of <i>S. pneumoniae</i> mNGS.</li> </ul>	[128]
2019	CSF samples from 204 pediatric and adult patients at eight hospitals.	A 1 year, multicenter, prospective study	<ul style="list-style-type: none"> <li>mNGS identified 32/58 diagnosed CNS infections, 13 of them were not identified by clinical testing. 11 were diagnosed by serologic testing only, and 7 were diagnosed from tissue samples.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS of CSF improved diagnosis of neurologic infections.</li> </ul>	[22]
2020	CSF samples from 248 adult patients suspected with CNS infections.	A single-center prospective cohort study	<ul style="list-style-type: none"> <li>mNGS had a sensitivity of 90.00% and 66.67% in culture-positive patients without or with empirical treatment, respectively.</li> <li>mNGS detected an extra of 48 bacteria and fungi in culture-negative patients.</li> <li>mNGS provided a higher detection rate compared to culture in patients with (34.45% vs 7.56%) or without empirical therapy (50.00% vs 25.00%).</li> <li>mNGS detection rate was significantly higher in patients with CSF WBC<sup>a</sup> &gt;300 × 10<sup>6</sup> L<sup>-1</sup>, CSF protein &gt;500 mg L<sup>-1</sup>, or glucose ratio ≤0.3.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS had an overall superior diagnosis efficiency in CNS infections to culture, especially in empirically treated patients.</li> <li>CSF laboratory results were related with the results of mNGS.</li> <li>mNGS could dynamically monitor disease progression.</li> </ul>	[117]
2020	CSF samples from 51 patients with suspected tuberculous meningitis from January 2017 to December 2018	A retrospective analysis	<ul style="list-style-type: none"> <li>The sensitivity, specificity, PPV, and NPV of mNGS in CSF for tuberculous meningitis were 84.44%, 100%, 100%, and 46.15%, respectively.</li> <li>The diagnostic sensitivity of mNGS (84.4%) was significantly higher than that of AFB<sup>a</sup> (0%), MGIT960 culture (22.2%), MTB PCR (24.4%), and Xpert MTB/RIF (40%).</li> <li>CSF protein quantification and cell count might be valuable in the prediction of mNGS positive MTB.</li> </ul>	<ul style="list-style-type: none"> <li>CSF mNGS had high sensitivity, specificity, and PPV in the diagnosis of TBM<sup>a</sup>.</li> <li>Significant increase in CSF cell number and protein quantification increased the likelihood of positive MTB detection of NGS.</li> </ul>	[132]
2020	CSF samples from 213 patients with infectious and noninfectious CNS diseases from November 2016 to May 2019.	A prospective multicenter study	<ul style="list-style-type: none"> <li>The mNGS-positive detection rate of CNS infections was 57.0%.</li> <li>The mNGS performance was optimal in the diagnosis of definite viral encephalitis and/or meningitis at an SSRN<sup>a</sup> ≥2, TBM at a GSRN<sup>a</sup> ≥1, bacterial meningitis at an SSRN ≥5 or 10, and cryptococcal meningitis and cerebral aspergillosis at SSRN ≥2, with a positivity rate of 42.6%, 27.3%, 73.3%, and 80%, respectively.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS of CSF effectively identified pathogens causing CNS infections.</li> <li>mNGS should be used in combination with conventional microbiological testing.</li> </ul>	[130]

**Table 2.** Continued.

Year	Samples and populations	Methods	Results	Conclusions	Refs.
<b>Respiratory infections</b>					
2018	Samples of sputum, blood, or BALF from 178 severe pneumonia patients in the ICU.	A retrospective study	<ul style="list-style-type: none"> <li>The bacterial detection rate of NGS group was significantly higher than the control group.</li> <li>The 28 day and 90 day mortality of NGS group was significantly lower than the control group.</li> </ul>	<ul style="list-style-type: none"> <li>NGS might be superior to conventional detection methods in severe pneumonia in ICU.</li> </ul>	[150]
2019	Specimens (pulmonary biopsy and BALF) from 55 cases (36 with mixed and 19 with nonmixed pulmonary infection) collected between July 2018 and March 2019.	A prospective study	<ul style="list-style-type: none"> <li>The sensitivity of mNGS in mixed pulmonary infection diagnosis was much higher than that of conventional test (97.2% vs 13.9%), but the specificity was the opposite (63.2% vs 94.7%).</li> <li>The PPV and NPV of mNGS was 83.3% and 92.3%, respectively.</li> <li>mNGS improved the diagnosis of pulmonary fungal infections.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS is a promising technique to detect copathogens in mixed pulmonary infection.</li> </ul>	[151]
2019	88 nasopharyngeal swabs from 63 patients with chronic obstructive pulmonary disease exacerbations.	A prospective study	<ul style="list-style-type: none"> <li>The sensitivity and specificity of mNGS is 96% and 98% as compared with PCR.</li> <li>mNGS detected additional viral pathogens.</li> <li>A positive correlation was found between Cq value and mNGS viral normalized species reads.</li> </ul>	<ul style="list-style-type: none"> <li>The mNGS was highly sensitive and specific for semiquantitative detection of respiratory viruses.</li> </ul>	[149]
2020	Lung biopsy tissues from 121 patients diagnosed with peripheral pulmonary lesions (PPLs) and lung infection.	A prospective randomized study	<ul style="list-style-type: none"> <li>The positivity rate of mNGS in R-EBUS<sup>3</sup>-guided TBLB<sup>3</sup> (78.7%) was significantly higher than the TBLB group (60.0%).</li> <li>Tissue collected by R-EBUS within the lesion produced higher positivity rate than samples adjacent to the lesions.</li> </ul>	<ul style="list-style-type: none"> <li>R-EBUS improves positivity pathogens detection rate of mNGS.</li> </ul>	[157]
2020	Samples (including lung tissue, BALF, and brush) collected from patients suspected with pulmonary infection from June 2018 to August 2019	A retrospective study	<ul style="list-style-type: none"> <li>mNGS identified at least one microbial species in 89% pulmonary infection cases.</li> <li>mNGS detected microbes related to human diseases in 94.49% of samples from pulmonary infection patients who had received negative results from traditional pathogen detection.</li> <li>The accuracy and sensitivity of mNGS are higher than those of traditional pathogen detection.</li> <li>mNGS could simultaneously detect and identify a large variety of pathogens.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS provided fast and precise pathogen detection and identification.</li> <li>mNGS helps in prompt and accurate treatment of peripheral pulmonary infection.</li> </ul>	[148]
<b>Bone and joint infections</b>					
2020	37 patients suspected with periprosthetic joint infection (PJI) who underwent prosthetic joint revision surgery from July 2016 to December 2018	A retrospective cohort study	<ul style="list-style-type: none"> <li>The sensitivity, specificity, and accuracy of preoperative synovial fluid mNGS were (92%, 91.7%, and 83.7%, respectively) higher than that of culture (52%, 91.7%, and 43.7%, respectively).</li> </ul>	<ul style="list-style-type: none"> <li>mNGS can be used as an effective supplemental method to improve diagnostic efficiency during the preoperative period.</li> </ul>	[160]
2020	44 periprosthetic tissues collected intraoperatively from patients who were suspected of PJI and underwent surgery.	A prospective study	<ul style="list-style-type: none"> <li>The sensitivity, specificity, PPV, NPV, and accuracy of periprosthetic tissue mNGS were (95.45%, 90.91%, 91.3%, 95.24%, and 93.18%, respectively) in PJI diagnosis were higher than that of culture (72.72%, 77.27%, 76.19%, 73.91%, and 75%, respectively).</li> </ul>	<ul style="list-style-type: none"> <li>mNGS can improve the diagnosis of PJI.</li> <li>Periprosthetic tissue can be used as an alternative to synovial fluid and sonication fluid.</li> </ul>	[51]

**Table 2.** Continued.

Year	Samples and populations	Methods	Results	Conclusions	Refs.
2017	131 sonication fluid samples from patients undergoing revision arthroplasty or removal of other orthopedic devices.	A laboratory method development study	<ul style="list-style-type: none"> <li>Compared to results from sonication fluid culture, the species-level and genus-level sensitivity of mNGS were 88% and 93%, respectively.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS can provide accurate diagnostic information in PJI.</li> </ul>	[162]
<b>Digestive system infection and urinary tract infection (UTI)</b>					
2017	Serum samples from 204 adult acute liver failure (ALF) patients collected from 1998 to 2010.	A retrospective cohort study	<ul style="list-style-type: none"> <li>8 cases of previously unrecognized viral infection were diagnosed by mNGS, as well as 7 missed dual or triple infections.</li> </ul>	<ul style="list-style-type: none"> <li>mNGS is helpful in screening for uncommon viruses and co-infections in ALF patients.</li> </ul>	[173]
2018	Urine/semen/rectal swab samples from 112 patients in different areas of urology for prevention and treatment purpose.	Clinical application of NGS in different clinical phase I–II trials	<ul style="list-style-type: none"> <li>Application of NGS in different clinical phase I–II trials demonstrated that this novel approach extends our knowledge about the microbiome of the urogenital tract.</li> <li>DNA sequence has a high sensitivity to detect a bacterial and fungal association with resistant genes to antibiotics.</li> </ul>	<ul style="list-style-type: none"> <li>The NGS DNA sequence technology helps patients with UTI gain an individualized approach for a more accurate diagnosis, prevention, prophylaxis, and treatment.</li> </ul>	[167]
<b>Infections in immunocompromised patients</b>					
2017	Blood samples, and in some cases, nasopharyngeal swabs and/or biological fluids from 101 immunocompromised adults.	A multicenter, blinded, prospective, proof-of-concept study	<ul style="list-style-type: none"> <li>Detection rate of clinically relevant viruses and bacteria identified by untargeted NGS were significantly higher (36%) than conventional methods (11%), and even when the latter were continued over 30 days (19%).</li> <li>Untargeted NGS had a high NPV compared with conventional methods.</li> </ul>	<ul style="list-style-type: none"> <li>Untargeted NGS is a promising method for microbiological diagnosis in immunocompromised adults.</li> </ul>	[188]

<sup>a)</sup>Abbreviations: mNGS, metagenomic next-generation sequencing; CSF, cerebrospinal fluid; MTB, *Mycobacterium tuberculosis*; WBC, white blood cell; PPV, positive predictive value; NPV, negative predictive value; AFB, acid-fast bacilli; TBM, tuberculous meningitis; SSRN, species-specific read number; GSRN, genus-specific read number; R-EBUS, radial endobronchial ultrasound; TBLB, transbronchial lung biopsy.

have reported that the overall diagnostic sensitivity of mNGS is significantly greater than bacterial culture used alone, while offering superior results for culture-negative sepsis cases.<sup>[57,116]</sup>

Note, bacterial DNA may periodically exist in the blood of healthy people, which was previously considered a sterile microenvironment, however, the bacterial diversity differs significantly from that of septic patients. A significant predominance of anaerobic bacteria, primarily belonging to the order of *Bifidobacteriales*, was observed in healthy volunteers, whereas aerobic or microaerophilic microorganisms dominated in septic patients. For example, the abundance of *Actinobacteria* phyla decreased in sepsis, while *Proteobacteria* phyla decreased in healthy volunteers.<sup>[112]</sup> These results may help interpret mNGS reports for patients with suspected sepsis, while reminding us that bacteria continuously translocate into the blood, but do not always cause sepsis.

To summarize, mNGS has undeniable advantages over culture-based methods in the diagnosis and treatment guidance of bloodstream infection, as it offers significantly reduced time required for pathogens identification to less than 24 h, regardless of the type of microorganism, and is less affected by antibiotics administration.<sup>[113]</sup> Moreover, this approach may prove invaluable in patients infected with fastidious pathogens, or

less commonly fungi, *Mycobacterium*, and parasites, allowing physicians to accurately diagnose and regulate treatment.<sup>[99]</sup> In addition, mNGS detects the presence of viral infections<sup>[117]</sup> or mixed-infections,<sup>[118–120]</sup> avoiding the abuse of antibiotics in sepsis-like patients, while informing proper and targeted antibiotics usage in septic patients.

### 3.1.2. Central Nervous System Infections

Various pathogenic microbes can infect the CNS, presenting as meningitis, encephalitis, and abscess, which are often life-threatening; however, routine microbiological tests are often insufficient to detect all neuroinvasive pathogens, especially for rare pathogens. Moreover, obtaining the relevant samples to detect the etiologic pathogens require invasive procedures, such as lumbar puncture or brain biopsy, which are limited by the availability and volume of CSF or brain tissues. Therefore, the etiology of CNS infections often remains undiagnosed, which occurs in up to 50% of acute meningoencephalitis cases.<sup>[121]</sup> Thus, methods for timely and accurate diagnosis are urgently required, where mNGS shows its superiority in recent years.<sup>[19,122]</sup> In fact, multiple case studies have reported



the detection of viruses,<sup>[120,123–127]</sup> bacteria,<sup>[19,128]</sup> fungi, and parasites<sup>[120,129]</sup> using mNGS in CSF and brain tissues.<sup>[130]</sup> Additionally, mNGS also proven valuable in diagnosing subacute or chronic meningitis.<sup>[120]</sup> Specially, elevated levels of white blood cells (WBCs) and protein, as well as decreased glucose ratio in the CSF may correlate with increased detection of CNS infections via mNGS.<sup>[117]</sup> A system review had recommend that NGS be considered as a front-line diagnostic test for chronic and recurring infections, and as a second-line technology for acute cases of encephalitis.<sup>[131]</sup>

In addition to its reasonably high diagnostic performances, mNGS may also help guide more targeted therapy for CNS infections and accelerate the identification and subsequent treatments for noninfectious causes.<sup>[22]</sup> However, although mNGS has an overall superior detection rate compared to conventional method, it cannot replace the necessity for traditional methodologies, such as culture.<sup>[117,130]</sup> Rather, mNGS is recommended to be applied in combination with conventional microbiological testing to improve the pathogenic diagnosis of CNS infections. Moreover, it may prove useful for ruling out an active infection in patients with suspected autoimmune encephalitis, thereby providing the necessary information for clinicians to confidently initiate immunosuppressive therapies for autoimmune disease with reduced concerns regarding missed detection of microbial infections.<sup>[22,130]</sup> Furthermore, mNGS results are less affected by prior use of antibiotics before collection of CNS samples,<sup>[128]</sup> and therefore the technique has an advantage over other methods in cases whereby disease-causing microbial detection in patients is greatly affected by antibiotic use. However, it should be noted that with prolonged treatment times, the detection rate of pathogenic microorganisms via mNGS is decreasing.<sup>[117]</sup>

Studies have shown that mNGS offers high sensitivity, specificity, and positive predictive value (PPV) in the diagnosis of tuberculous meningitis in the CSF.<sup>[30]</sup> Indeed, the sensitivity of mNGS was significantly higher than culture alone, and a combination of mNGS and conventional methods significantly increased the detection rate significantly. Moreover, patients with a significant increase in cell number and protein in their CSF have a higher chance of being MTB positive when detected by NGS.<sup>[132]</sup> These results indicate that mNGS has the potential to be used as a first-line CSF test to detect the presence of mycobacterial DNA. Since MTB complex members exhibit >99.99% genomic sequence similarity, genus stringent mapped reads number (GSMRN) was considered reliable when interpreting reports for the MTB complex.<sup>[130]</sup>

mNGS of CSF can also effectively identify fungi causing CNS infections. A large, prospective, multicenter series case study showed that the sensitivity of mNGS using species-specific read number (SSRN, equivalent of stringent mapped reads number, SMRN)  $\geq 2$  in the diagnosis of cryptococcal meningitis and cerebral aspergillosis was 76.92% and 80%, respectively,<sup>[130]</sup> which can be significantly improved with the combination of traditional methods, thereby informing the choice of appropriate antifungal agents and therapeutic courses.<sup>[118]</sup> However, wide adoption of mNGS for clinical purposes to diagnose cryptococcal meningitis is unlikely as the current diagnostic tests available (CSF CrAg) are highly sensitive and specific, with rapid turnaround times.<sup>[133]</sup>

Definitive diagnosis of viral infection in the CNS is dependent upon virus isolation from CSF or brain tissues, which is challenging and generally only performed in the laboratory. Therefore, treatment primarily depends on empirically provided therapy. The identification of certain CNS viral pathogens remains difficult, as many viral families that infect the CNS failed to be identified by the gold standard pathogen-specific PCR assays. Although DNA viruses (mostly herpes virus) are often identified by mNGS, its prediction of viral encephalitis and meningitis has not significantly improved,<sup>[134–137]</sup> which may be partially due to the absence of RNA detection methods, since RNA-based mNGS has not been launched extensively.<sup>[130]</sup> RNA viruses, such as enteroviruses and the Japanese encephalitis virus, are important causative agents of viral encephalitis and meningitis.<sup>[138]</sup> Hence, simultaneous extraction of DNA and RNA followed by cosequencing may improve the detection rates of viruses in CSF samples.<sup>[139]</sup>

Further, mNGS also has values for the characterization of complex and rare pathogens present in culture-negative and undiagnosed cases. mNGS has been shown to help detect the presence of microbes, such as *Listeria monocytogenes*,<sup>[140]</sup> Brucellosis-causing species,<sup>[93,141]</sup> *Naegleria fowleri*,<sup>[142]</sup> neurocysticercosis-causing parasites,<sup>[143,144]</sup> and *Vibrio vulnificus*.<sup>[145]</sup> In CNS toxoplasmosis, mNGS may be useful in cases when the toxoplasmosis IgG is negative, CSF PCR is negative, and imaging is not classic, or when there is a lack of response to antitoxoplasmosis therapy.<sup>[119,133]</sup>

Additionally, mNGS can be used to dynamically monitor disease progression using semiquantitative value analysis.<sup>[117]</sup> However, negative tests must be interpreted with caution owing to the higher risk of false-negative results.<sup>[22]</sup>

### 3.1.3. Respiratory Infections

Upper and lower respiratory tract infections are among the most common illnesses leading to medical consultation, and are associated with significant mortality.<sup>[146]</sup> Hundreds of pathogens, including bacteria, fungus, viruses, and parasites, can cause pulmonary infection, and the situations in immunocompromised patients may be even more complex. Undoubtedly the identification and characterization of pathogens is crucial for precision treatment and improved prognosis of patients. However, frequent and inappropriate use of antibiotics in respiratory tract infections limits the sensitivity and reliability of culture-based testing, which often occurs in clinical settings.

mNGS provides relatively fast and precise detection and identification of a large variety of pathogens, contributing to the prompt and accurate treatment of pulmonary infection,<sup>[147–149]</sup> particularly for critically ill patients and those with mixed-infection.<sup>[150,151]</sup> The most common pathogens causing lower respiratory pulmonary infectious have been identified as bacteria, such as *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, and *Acinetobacter baumannii*.<sup>[99,152]</sup> In addition, a greater number of fastidious bacteria, including MTB, nontuberculous mycobacteria (NTM), *Nocardia*, and various Actinomycetes, were identified by mNGS compared to conventional culture methods.<sup>[99]</sup> Moreover, the use of mNGS has also been shown to improve the diagnosis of pulmonary invasive fungal

infections.<sup>[147,151]</sup> The sensitivity of mNGS is further highlighted in mixed pulmonary infection diagnosis and severe nonresponding pneumonia.<sup>[153–155]</sup> In fact, the detection accuracy rate of mNGS may even reach 100% in the analysis of immunocompromised patients.<sup>[156]</sup>

Additionally, mNGS provides more strain-specific information and helps to identify new pathogens.<sup>[153]</sup> However, the interpretation of the specificity differs,<sup>[151]</sup> depending on whether the pathogens are the causative agents of infection due to the presence of normal flora, commensal oral flora, or colonizers, therefore a tradeoff is generally needed. NGS is currently the preferred methodology for virus discovery, particularly for emerging pneumonia-causing viruses (i.e., SARS-CoV, MERS-CoV, and H7N9), and could potentially help to trace and control outbreaks.<sup>[154]</sup>

Interestingly, the specificity of mNGS is reportedly higher for transbronchial lung biopsy (TBLB) tissue than for BALF; however, detection in BALF has been shown to have a higher sensitivity in the diagnosis of peripheral pulmonary infectious lesions.<sup>[152]</sup> However, in our experience, the overall sensitivity of mNGS does not differ among respiratory tract sample types and is not superior to that of culture for recognizing common bacteria. Nevertheless, mNGS has been reported to have a significantly higher sensitivity than culture in sputum and lung tissue samples.<sup>[99]</sup> Moreover, NTM, rather than MTB, *Aspergillus*, or *Cryptococcus*, was more often detected in BALF than sputum by mNGS.<sup>[99]</sup> Meanwhile, R-EBUS-guided-TBLB significantly facilitates the accurate insertion of the bronchoscope into the lesions in the pulmonary, thereby improving the positivity rate of mNGS analysis in pathogen detection.<sup>[157]</sup>

### 3.1.4. Bone and Joint Infection (BJI)

BJIs are serious and potentially life-threatening, particularly for periprosthetic joint infection (PJI), which is a devastating complication that can occur after arthroplasty and multiple revision surgeries.<sup>[158]</sup> Conventional diagnostic challenges in BJI due to the fastidious nature of the organisms, biofilm formation on implant surfaces, prior antibiotic administration, or limited sample availability, all contribute to the high rate of culture-negative PJIs, the microbial etiology for which is challenging to diagnose.<sup>[159,160]</sup>

In most cases, mNGS, with high sensitivity, specificity, and positive and negative predictive values, can be used as an effective supplemental method to improve the diagnostic efficiency of BJI. Synovial fluid, sonication fluid from explants (i.e., prosthetic joint and other orthopedic devices), and periprosthetic tissue, all can be used to identify the causative pathogens via mNGS.<sup>[51,161]</sup> Remarkably, sonication fluid is potentially more advantageous than synovial fluid for distinguishing the primary causative pathogens from contaminants, possibly because sonication may enrich the microorganism abundance in samples by dislodging biofilms from the surface of implants.<sup>[162–164]</sup> Fastidious bacteria NTM can be detected by mNGS from all sample types, enabling surgeons to collect specimen from various samples. Moreover, assigned reads in the sonicated fluid is significantly higher than in synovial

fluid (more than 200-fold times) and tissues,<sup>[165]</sup> which need further research to verify.<sup>[162–164]</sup> mNGS should be highly recommended for the cases of culture-negative BJI, limited sample volume obtained by joint cavity puncture, or when the patient has poor responses to empirical antibiotics therapy or debridement.<sup>[160,165]</sup>

### 3.1.5. Urinary Tract Infection (UTI)

UTI is a common community-acquired infection commonly caused by bacteria, as well as certain fungi or virus (i.e., polyomavirus and the human herpesvirus).<sup>[166–168]</sup> *Escherichia coli* ( $\approx 75\%$ ) is the most common causative agent for both uncomplicated and complicated UTI, with the latter comprising a complex constitution of causative microorganisms, making it rather challenging to diagnose the pathogens.<sup>[166,169]</sup> However, mNGS is capable of identifying pathogens more quickly and with more sensitivity than culture methods, providing clinicians with the information necessary to make accurate diagnoses, as well as to determine the preventive, prophylactic, and appropriate treatment options.<sup>[166,167]</sup>

However, interpretation of urine mNGS reports may prove difficult for various reasons. First, urine samples are readily contaminated by improper procedures before and during urine collection. Second, colonizers inhabit the distal urethra, the skin around the urethral meatus, as well as the vagina, which must be taken into consideration, as specific organisms may become pathogenic in immunocompromised patients. Third, asymptotically persistence or latent infections often occur. Therefore, even with precise quantitative distribution of all microbial associations, it may be difficult to define which bug or superbug is causative of UTI.<sup>[166,168]</sup>

### 3.1.6. Digestive System Infection

The major enteropathogenic bacterial species include invasive enteropathogens originating from outside environments, and pathobionts originating from commensal gut species. Although the study of the intestinal microbiome is a very relevant topic popular among researchers, only a few have attempted to diagnose the associated infectious diseases, such as diarrhea, using mNGS techniques.<sup>[170]</sup> As the intestines are normally colonized by commensal bacteria, mNGS of stool samples primarily focus on nonbacterial species, such as *Rotavirus A*, *Cryptosporidium*, and the human parechovirus, which are often detected in the stool samples of febrile children.<sup>[171]</sup> mNGS has also been used to detect potential causative agents in patients with acute cholecystitis,<sup>[172]</sup> screen for the presence of uncommon viruses, or identify co-infections in patients with acute liver failure.<sup>[173]</sup> In addition to pathogen identification, mNGS is also a rapid and agnostic diagnostic approach for investigating resistome, and can be used for identifying antibiotic resistance genes (ARGs) in the gut, helping detect multidrug resistant organisms (MDROs), thus leading to early implementation of infection prevention practices, antimicrobial optimization, and prevention of invasive infections.<sup>[174,175]</sup>

### 3.1.7. Complex and Atypical Pathogen Infections

Many cases have demonstrated the superiority of mNGS in the identification of i) complex and atypical infections, such as *Chlamydia psittaci*,<sup>[176,177]</sup> *Legionella spp.*,<sup>[178]</sup> anaerobes that are difficult to culture,<sup>[179]</sup> and mycobacteria;<sup>[165,180]</sup> ii) zoonotic pathogens, such as *Streptococcus suis*,<sup>[181]</sup> *Orientia tsutsugamushi*,<sup>[182]</sup> *Leptospira santarosai*,<sup>[19]</sup> and *Toxoplasma gondii*;<sup>[119]</sup> iii) emerging RNA and DNA viruses, such as *Bocaparvovirus*,<sup>[183]</sup> Ebola virus,<sup>[184]</sup> Zika and chikungunya viruses,<sup>[185]</sup> bornavirus,<sup>[186]</sup> and SARS-CoV-2.<sup>[187]</sup>

## 3.2. Infections in Special Populations

### 3.2.1. Infections in Immunocompromised Patients

Infections in immunocompromised hosts have drawn a great deal of attention by clinicians, particularly in recent years, since they are generally more complex and serious with the potential to result in disastrous consequences, such as failure of a transplanted solid organ. mNGS has demonstrated distinct advantages in the detection of pathogens in this group of people,<sup>[188]</sup> and was recommended as a first-line diagnostic tool for patients undergoing solid organ transplantation and hematopoietic stem cell transplantation,<sup>[23,110,188–190]</sup> or as an addition to routine test panels.<sup>[191]</sup>

Allograft health monitoring, including infections and transplant rejection, which occur frequently, and sometimes simultaneously, is an important component of post-transplant therapy. mNGS detection of cfDNA can be applied in monitoring organ transplant rejection via genome transplant dynamics, and may have the potential to replace invasive techniques, such as lung and endomyocardial biopsies as the symptoms of infection and rejection are often difficult to discriminate.<sup>[110,192]</sup> Additionally, data from mNGS may provide important information regarding the relationship between the human virome, the state of the immune system, and the effects of pharmacological treatment. For example, the total viral load is reportedly more significantly affected by immunosuppression than the bacterial microbiome.<sup>[193]</sup> The infecting viruses include those that are common, such as *Cytomegalovirus*, and Epstein–Barr virus, as well as undiagnosed DNA viruses, such as the human herpesvirus, polyomavirus, adenovirus, and the torque teno virus.<sup>[23,110,190]</sup>

### 3.2.2. Febrile Illness

The global burden of febrile illness has been difficult to quantify. Identification of etiologies in patients with fever of unknown origin (FUO) is the chief issue, among which infectious factors remain the primary cause.<sup>[194]</sup> This poses a challenge for diagnostic and therapeutic determination, and may result in inappropriate antibiotic use, making it difficult to predict, detect, or evaluate potential outbreaks or emerging infections. Unbiased mNGS represents a powerful tool to fill the gaps in our understanding toward the etiology of febrile illness, thus informing the improvement of diagnostic algorithms, therapeutic guidelines, and public health strategies.<sup>[171,195]</sup>

Viruses play an important role in FUO, and distinctions have been observed between the viromes of febrile and afebrile groups using short-read Illumina sequencing. The assessment and identification of known, or potentially novel, viruses have improved the medical management of children with FUO, helping to avoid the administration of unnecessary antibiotic therapy in those with viral infections.<sup>[195]</sup> Another worldwide challenge is travel-associated infections, with a broad spectrum of potential etiologies. mNGS has greatly improved the pathogens diagnostic efficiency of returning travelers with acute febrile illness, particularly for viruses, such as dengue virus, Ebola virus, hepatitis E and hepatitis A, and Chikungunya virus.<sup>[196]</sup> mNGS is considered to have the potential of being used as an all-in-one diagnostic test for FUO patients, helping differentiate infections and noninfections, identify causative agents, and discover novel or emerging pathogens.<sup>[196]</sup> However, a cost–benefit study reported that current conditions do not warrant a widespread rush to deploy metagenomic testing to resolve any and all uncertainty, but rather as a front-line technology that should be used in specific contexts, such as acute and seriously ill cases, or as a supplement to, rather than a replacement for, careful clinical judgment.<sup>[197]</sup>

### 3.2.3. Unexplained Pneumonia

Metagenomic analysis is of great value for the monitoring and response to rapidly emerging infectious diseases, such as unexplained pneumonia, among which COVID-19 is currently the most popular.<sup>[198]</sup> Metagenomic sequencing, with its unbiased nature, plays a crucial role in the etiological diagnosis of unexplained pneumonia, particularly when encountering an atypical or novel causative agent.<sup>[187,199,200]</sup> During the initial outbreak of SARS-CoV-2, a commercial sequencing company provided an incorrect report stating the identification of SARS virus, causing widespread panic. Later, upon careful analysis using additional sequencing data, the causative agent was determined to be a novel coronavirus.<sup>[200]</sup> Further, combined with multiplex PCR amplification and nanopore sequencing, the phylogenetic structure of the novel coronavirus were analyzing during the early days of the pandemic.<sup>[201]</sup>

However, important lessons should be learned from this outbreak; i.e., although metagenomics sequencing lacks the capability of identifying the novel pathogen, it proved valuable for ruling out common pathogens as first-line diagnosis of unexplained pneumonia.<sup>[200]</sup>

## 3.3. Infection Control

The hospital infection prevention control management is an important aspect of healthcare, nosocomial infections occur worldwide and affect hundreds of millions of patients, leading to tens of billions of dollars lost each year, and remain a particularly aggressive problem in the pediatric healthcare context.<sup>[202]</sup> The mNGS of clinical or environment samples is a promising approach for rapidly identifying pathogen sequences for the detection and epidemiological determination of transmission.<sup>[203–205]</sup>

## 4. The Global Expert Consensus for NGS in the Diagnosis of Pathogens

mNGS, with recent advances, reduced costs, user-friendly data analysis tool, and accurate and comprehensive databases, has been a revolutionary technology that has disrupted traditional clinical diagnostic microbiology.<sup>[106]</sup> mNGS can overcome the limitations of traditional diagnostic tests, enabling a hypothesis-free, culture-independent, and universal pathogen detection platform, with which clinical specimens can be directly used for microbe identification of bacteria, viruses, fungi, or parasites, while also assisting in the identification of novel organisms.<sup>[3]</sup> Yet, few associated expert consensus for the use of NGS in infectious diseases diagnosis. However, the general consensus is that its relatively high cost is a barrier to mNGS becoming the first-line of detection method for clinical in the short term. However, it is recommended as the first-line detection method in situations involving patients with difficult or complicated infectious diseases, including those who are acutely and critically ill, suffering from immunodeficiency,<sup>[188]</sup> or in special populations, as well as in situations of infectious diseases outbreaks of unknown origin.<sup>[187,206]</sup> In most cases, mNGS is recommended as a supplement to conventional detection methods, particularly in cases where traditional diagnostic assays are negative<sup>[133]</sup> (Table 3).

## 5. The Challenges for NGS in Pathogens Diagnosis

The multiple achievements made in high-throughput sequencing technology for the diagnosis of infectious diseases are encouraging; yet, some hurdles remain to be addressed, including the limitations of the technology, clinical application, high costs, the lack of current standardization of methods and data analysis, interpretation guidelines for clinicians, generation of appropriate work pipelines, the regulation, and supervision. Moreover, it remains to be determined how to: i) minimize the influence of human host background and extraneous sources of nucleic acids; ii) differentiate colonization from infection; iii) improve the detection efficiency of intracellular

microorganisms; iv) generate more reasonable standardization of the methods, data storage, protection, analysis, and interpretation. Here, we comprehensively discuss the challenges of the clinical application of high-throughput sequencing technologies and propose possible solutions.

### 5.1. The Level of Technology

#### 5.1.1. The Influence of Human Host Background

Samples collected for testing samples can be of different types depending on the site of infection, and may be obtained from the CSF, peripheral blood, sputum, BALF, or other sites. As the human genome is much larger than the microbial genome, the host DNA background is also much higher than that of microbes; therefore, even a small number of human cells can greatly affect the accuracy of microbial detection, decreasing the useful coverage of the sequencing data, and increasing the sequencing cost. Currently available methods for the removal of human DNA and enriching microbial DNA include: i) the removal of host cells based on differences in cell size relative to microbial cells using methods such as differential centrifugation and filtration, and ii) the selective lysis of host cells and subsequent treatment with enzymatic or chemical reagents to remove exposed DNA.<sup>[207]</sup>

Centrifugation is a simple, convenient, and low-cost method for the removal of human DNA and is widely used in mNGS-based detection processes. For example, CSF samples from patients with meningitis are centrifuged at  $5000 \times g$  for 15 min to remove cell precipitates, nucleic acid is then extracted from the supernatant, and the sequencing library is constructed. Processing the samples before library construction can reduce the host genome proportion, significantly increase the sequence coverage of pathogens in the samples, and improve the detection sensitivity.<sup>[20]</sup> Further, centrifugation and filtration can be combined to enrich pathogenic microorganisms. For second-generation sequencing to detect viruses, samples can be centrifuged at a low speed of  $6000 \times g$  for 10 min, and the supernatant filtered using a  $0.80 \mu\text{m}$  filter to remove eukaryotic and bacterial cells without affecting

**Table 3.** Global expert consensus and guidelines of NGS in pathogens diagnosis.

Countries/regions	Year	Viewpoint	Refs.
Winnipeg	2018	NGS has obvious advantages in the detection and typing of HIV drug resistance genes.	[243]
International	2018	NGS provides new means for the cognition of joint microorganisms and is expected to better understand joint colonization.	[244]
UK	2018	NGS should be considered as a front-line diagnostic test in chronic and recurring presentations and, given current sample-to-result turn-around times, as second-line in acute cases of encephalitis.	[131]
Singapore	2018	Current conditions do not warrant a widespread rush to deploy metagenomic testing to resolve any and all uncertainty (i.e., pyrexia of unknown origin PUO), but rather as a front-line technology that should be used in specific contexts, as a supplement to rather than a replacement for careful clinical judgment.	[197]
China	2020	As a comprehensive direct detection method, the current high cost makes it impossible for mNGS to become a clinical first-line detection method in the short term, but in difficult and complicated diseases, critically ill, immunodeficiency, and other special population, it still has the potential to become a quasi-first-line detection method for pathogen diagnosis.	[92]



the extraction of RNA viruses.<sup>[208]</sup> For example, recovery of the human genome from saliva samples by sifting through a 5 µm filter was  $89.69 \pm 0.84\%$  compared to unfiltered original samples  $89.29 \pm 0.61\%$  recovery. These results indicate that a considerable amount of extracellular host DNA in the samples may not be removed via filtration method based on cell size alone. To overcome this deficiency, human cells can be selectively lysed, and enzymes or chemical reagents can be used to remove exposed DNA. The percentage of human DNA in samples can be reduced using methods such as propidium monoazide treatment after selective mammalian cell lysis by sonication (human DNA recovery is 25.6%), and via osmotic lysis with H<sub>2</sub>O (1.7%) followed by mammalian cell removal by differential centrifugation (1.4%). The treatment of raw saliva samples with propidium monoazide has been shown to reduce the percentage of human DNA (16.8%), suggesting that the majority of human DNA in saliva is already exposed.<sup>[209]</sup> In addition to ultrasonic fragmentation, nonionic surfactant saponins can also be used to lyse human cells and subsequently, the digested DNA by enzymes has been shown to increase the number of sequences retrieved for *Cryptococcus neoformans*.<sup>[20]</sup>

Human genomic DNA undergoes methylation, and methylated host DNA can be excluded from samples using the methyl-CpG binding domain to enrich for microbial DNA, resulting in a 50-fold reduction in the number of reads mapping to the human genome.<sup>[210]</sup> A study on the removal of host DNA from CSF showed that upon processing the sample using the NEB Next Microbiome DNA Enrichment Kit, the relative proportion of pathogen DNA relative to human reads was enriched by  $\approx 2.5$ -fold.<sup>[211]</sup> However, results from another study showed that the proportion of the human genome in saliva treated with the same kit remained high at  $90.83 \pm 0.77\%$ , which was not significantly different than the untreated original samples ( $89.29 \pm 0.61\%$ ).<sup>[207]</sup> At present, the evaluation of the effect of dehosting nucleic acids based on the characteristics of host DNA methylation is different, and is related to the sample type, host DNA content, and other factors. The true dehosting efficiency remains to be confirmed by additional experimental data. Currently, all dehosting DNA schemes have certain inherent limitations, and a combination of several schemes may be considered in the future to achieve maximum host DNA removal.

### 5.1.2. mNGS Efficiency in Samples Containing Intracellular Bacteria and Fungi

mNGS can potentially detect the presence of all microorganisms in a sample, and the content of the particular microbial DNA in the sample partially affects the detection rate of the species. The detection efficiency of fungi and intracellular bacteria, such as tuberculosis, is lower than that of other microorganisms. Therefore, a suitable cell lysis method that can efficiently lyse the cell walls of fungi and intracellular bacteria may improve the detection sensitivity of these microorganisms.

At present, three kinds of methods for microbial cell lysis exist: heating and enzymatic digestion of cell walls, mechanical (sonication and using bead-beaters), and chemical.<sup>[212]</sup> In a comparative study of several cell wall disrupting methods, lysozyme was found to be effective in the enzymatic hydrolysis

of cells; however, its use is limited to Gram-positive bacteria that are sensitive to lysozyme. Although bead-beating is more generally applicable, the cell wall cracking efficiency can be highly variable depending on equipment, buffer, and the cell type. For spores and other cells, such as mycobacteria, that are difficult to destroy through bead-beating.<sup>[213]</sup> In another study using six different methods for cell wall disruption and extraction of MTB DNA, the results showed that Infection Diagnostic, Inc. (IDI) lysing tubes had a higher cleavage and extraction efficiency than the other five treatment methods, including the QIAGEN QIAamp DNA mini kit.<sup>[214]</sup> Many schemes can be used to improve the cell wall disruption efficiency in fungi and intracellular bacteria, though it is important to consider the corresponding effect on RNA extraction. In a study by Huang et al., the sample treatment program used underestimated the possibility of RNA virus infection. As the cell walls of fungi and intracellular bacteria are thick, cell wall disrupting methods are used to extract DNA, which may affect the detection of RNA viruses.<sup>[148]</sup> Therefore, it is important to consider the use of a variety of cell wall disrupting techniques to improve the efficiency of cell breakage and the detection rate for RNA viruses.

### 5.1.3. Stability of RNA in the Detection Process

Infections resulting from RNA viruses are common. For example, detection of the novel SARS-CoV-2 coronavirus, mNGS technology was used to extract total RNA from BALF samples, resulting in the successful identification of the virus.<sup>[215]</sup> RNA forms unstable molecules that are easily degraded by RNases in the environment. Therefore, RNA extraction immediately after sample collection can help to obtain high-quality RNA. However, samples often need to be sent to other locations for testing, and thus, timely extraction cannot be guaranteed. To ensure the accuracy of the subsequent test results, it is necessary to ensure that the RNA in the samples is not degraded during collection and transportation. At present, RNA samples are transported using dry ice or after the addition of an RNA stabilizer. Although a low-temperature environment can slow the rate of nucleic acid degradation, multiple freeze-thaw steps leads to the release of cellular nucleases, resulting in RNA degradation.<sup>[216]</sup> Commonly used reagents for sample preservation include RNAlater (Thermo Fisher Scientific and Qiagen) and RNastable (Sigma-Aldrich).<sup>[217]</sup> As transcriptome technology has matured over time, sample processing is carried out in strict accordance with standard operating procedures to ensure the accuracy of the test results.

### 5.1.4. Detection of Drug Resistance Genes and Virulence Factors

The development of sequencing technology allows for a more in-depth understanding of the changes and evolution of microorganisms, as well as a comprehensive and accurate description of strain genotypes, virulence, antimicrobial resistance spectra, and systematic genetic background.<sup>[218]</sup> Mason et al. analyzed the data for resistance causing genes using whole genome sequencing (WGS) of 1379 isolates of *Staphylococcus*

*aureus*. Their results showed that the correlation between drug resistance genotypes and phenotypes can be as high as 98.3%, indicating that before obtaining results related to drug sensitivity, drug resistance in infected pathogens can be detected using WGS technology to avoid the ineffective use of drugs.<sup>[219]</sup> Greater attention is being paid to virulence genes, and an in-depth study on virulence factors is likely to provide a strong basis for the distinction between pathogenic and colonized bacteria. A genome-wide study on the virulence genes of *K. pneumoniae* clinical isolates identified 21 new genes in addition to the *iroBCD/iucABCD* and *rmpA/A2* genes already known to be associated with virulence. These 21 genes may be factors contributing to the high virulence phenotype of *K. pneumoniae*.<sup>[220]</sup> Although sequencing technology has made various strides in the detection of virulence factors and drug resistance genes, there are certain differences between mNGS- and WGS-based detection, with challenges persisting in the application of mNGS for practical detection. First, samples generally expend a large amount of sequencing output consuming human DNA sequences, leading to insufficient coverage for the detection of pathogen DNA and limiting the comprehensive analysis of drug resistance genes. Second, second-generation sequencing generates short fragments and even after the detection of drug resistance genes, it may be difficult to determine the pathogen to which the gene belongs due to sequence splicing. Third, the drug resistance mechanism is complex and diverse; the current databases for drug resistance genes and virulence genes require improvement, and there are inconsistencies in the reports of genotypes and phenotypes of pathogens. These considerations have caused resistance to the application of mNGS for the detection of drug resistance genes and virulence factors. Meanwhile, third-generation sequencing technology offers the advantages of delivering long reads, high-throughput, and synchronous data analysis, and is a promising technique for the detection of drug resistance genes and virulence factors.

## 5.2. Biological Information Analysis

### 5.2.1. Optimization of Microbial Databases

Public databases containing microbial reference genomes are constantly being updated, requiring laboratories to track the version being used at any given time, and to address potential erroneous annotations and other database errors. Larger and more complete databases that contain publicly stored sequences, such as the NCBI nucleotide database, are more comprehensive and more accurate than the limited databases, such as FDA-Argos or RVDB, though the larger databases may contain more errors. A combinatorial method using multiple database annotation sequences can improve the sensitivity and specificity of microbial identification.

In the laboratory, the analysis software and reference databases used are best determined before verification and clinical use. Many laboratories maintain the production version and the latest development version of the clinical reference database (for example, the NCBI nucleotide database is updated fortnightly), as databases are updated regularly at prespecified

intervals. Standardized datasets can be used to validate the database after updates to ensure the accuracy and repeatability of results, as newly saved sequences and clinical metadata may introduce errors.

In addition to the aspects mentioned above, it is also important to take note of the following: i) incomplete sequences of rare pathogens or emerging pathogenic strains, ii) biased references toward species, iii) genetically similar pathogens (e.g., contamination of mycobacterial strains), and iv) the presence of normal flora in reagents, which is a common phenomenon that limits specificity. To provide accurate results, the reported algorithm may need to account for the test process and sample quality, pathogens that are rare in current and previous tests, the relative and absolute abundance of microorganisms, whether there are genetically similar microbes, and microbial genome coverage. The workflow for reporting results requires that the standards for quality control and the interpretation of results should be pre-established, which may include expert reviews of all cases or a subset of cases that meet the defined criteria or in diseases with unusual outcomes.

### 5.2.2. Virus Classification

Viruses are the simplest biological organisms with varying genome types, ranging from single-stranded RNA or single-stranded DNA, which encode only a few proteins, to dsDNA containing viruses that are similar in size to bacterial genomes. The average genome size of viruses is smaller than 50 kb. Viruses evolve and mutate faster than other pathogens, hence, the classification of viruses has greater challenges compared to those for bacteria or fungi, and there are often different classification rules for different groups of viruses.<sup>[221–223]</sup> The International Committee on Taxonomy of Viruses (ICTV) is responsible for updating information relating to virus classification that depends on genomic information.<sup>[223,224]</sup> The standardization of genome assemblies has resulted in the creation of standards for the quality control of sequencing data, splicing information, quality evaluation after splicing, and additional problems that require manual identification.<sup>[225]</sup> Standardization of the genome annotation process includes the prediction of viral genes, while classification and identification criteria are based on the COG protein homologous family, with the accuracy of classification ensured by using the characteristics of genomes with different classification units.<sup>[226,227]</sup>

RVDB is a standard genome database used for virus classification that contains nucleic acid sequences related to all identified viral, virus-related, virus-like, and endogenous non-retroviral elements, as well as endogenous retroviruses and retrotransposons, however, excludes bacteriophages and other nonviral sequences. The virus genomic sequences are checked and verified manually, similarity comparisons are made, redundancies are removed, and the classification of virus species is confirmed by a comparison with clustered representative sequences.<sup>[228]</sup> Virus identification software, such as DisCVR, also uses the k-mer algorithm and compares the length of the k-mer (18, 22, 26, and 30). Ultimately, 22 is selected and, compared with CLARK and Kraken, it has higher specificity and sensitivity.<sup>[95]</sup> Therefore, along with the use of virus

genome databases such as NT or NCBI, there is a need for the optimization of the alignment algorithms of existing databases based on the inclusion standard for accurate virus classification. Viral genomes from public databases in the future can be included in the local identification database through the confirmation of classification, filtering of contaminating sequences, providing genome quality ratings, etc., and simultaneously, it is necessary to carry out parallel or confirmatory identification and confirmation.<sup>[229]</sup>

### 5.3. The Level of Interpretation

A well interpreted mNGS report can guide clinicians in diagnosis, differential diagnosis, treatment, and prognosis evaluation. The human microbiome is complex as portions of the human body are naturally colonized by numerous bacteria, fungi, and viruses, primarily within the respiratory and gastrointestinal tract, skin, and vagina, whereas most areas inside the body of healthy individuals are physiologically sterile (i.e., blood) and microorganisms appear only under certain conditions (i.e., sepsis). Therefore, the interpretation of mNGS reports is rather difficult and requires a battery of clinical, laboratory, and microbiological research experience.

Due to the associated technical and clinical complexity, there is no absolutely standard method for interpreting mNGS results. Hence, clinicians working with the sequencing results for an infectious disease face several constraints. To address these challenges, first resumption of working groups should be implemented. Some institutions have assembled teams working on precision medicine, consisting of representatives from clinician groups (particularly infectious diseases), microbiology laboratories, and professional technical teams, who can assess the results of sequencing and provide an appropriate interpretation before the reports of the mNGS are delivered to clinicians. Another strongly advised approach is to have an experienced laboratory or clinician director manually revise the results prior to release.<sup>[230]</sup>

Second, to provide accurate results, several constraints should be considered when formulating the algorithms for interpreting the reports. i) the quality of the samples and test process; ii) types and collection methods of samples, whether from a site with colonized flora or not; iii) detection efficiency for different microorganisms, such as common bacteria, mycobacteria, fungi, virus, or parasites; iv) certain pathogens may be similar genetically (i.e., species of mycobacteria); v) the reference databases bias or incompleteness for rare pathogens or emerging strains of pathogens; vi) contamination from the reagents and environment. Laboratories should establish an appropriate workflows depending on its own unique situation, including metrics for quality control and interpreting the findings, an expert review should then be included when defining criteria or reviewing unusual findings.<sup>[106,231,232]</sup>

A large proportion of the samples are collected from the respiratory tract, where the microbiome is particularly difficult to characterize, with diverse pathogens causing infectious diseases, as well as colonizers. For example, the mNGS of lung biopsies or BALF yields reads of bacteria from the oropharyngea flora, as well as viruses or *Candida*, which may

not be considered causes of pulmonary infections in immunocompetent patients, but may be considered pathogenic in immunocompromised hosts.<sup>[147]</sup> Thresholds such as the relative abundance of a potential pathogen at 30% of all genera present and the sampling of at least 100 unique reads could be used to indicate the presence of potential pathogens relative to the background lung microbiome. In general, higher coverage rate, relative abundance, and SMRN are typically reliable with true infections, whereas lower indices are somewhat discredited for infections or are considered to be more likely associated with the presence of commensal/colonizer/contaminant microorganisms of unknown clinical significance. Thus, the entire clinical picture must be taken into consideration in any cases when determining the clinical significance of the detected microbes.<sup>[115]</sup>

### 5.4. Limitations Compared with Traditional Pathogen Identification Methods

Massive case reports, case series reports, and studies on clinical application have sprung up and showed the superiority of mNGS over traditional microbial detection methods, such as microscopy and staining, culture, nucleic acid amplification tests, immunological assays, 16S rRNA gene sequencing, and MALDI-TOF MS, as presented in detail in Table 2. mNGS works better in comprehensively cataloging the individual members comprising a specific microbiome, optimizing the precise treatment strategy and improving the prognosis of patients, and understanding how microbial communities function and influence host–pathogen interactions.<sup>[233]</sup> However, at present, it is generally recognized that mNGS will not likely universally replace traditional microbiology techniques in the short term for the following reasons: i) some traditional detection methods are highly sensitive and specific with rapid turnaround times, such as CSF CrAg, rapid and specific PCR methods; ii) indirect approaches such as serological tests will continue to play a key part in the diagnostic workup for infections, especially for the emerging infectious diseases;<sup>[234]</sup> iii) functional assays such as culture, phenotypic susceptibility testing, and drug sensitive test will likely always be useful for research studies.<sup>[235]</sup>

In summary, there are considerable barriers, including sensitivity, interpretation, actionability, popularity, turnaround time, antimicrobial susceptibility, clinical utility, laboratory workflow, cost, etc., for the adoption of diagnostic metagenomics. Current limitations suggest that mNGS can be a complementary, and perhaps an essential test in certain clinical settings, rather than being an “one-in-all” test.<sup>[148]</sup>

### 5.5. Recommendations for the Use of mNGS in Clinical Applications

The application of mNGS in the detection of clinical pathogens remains in the early stage, with no mNGS workflows or methods approved by the Food and Drug Administration or National Medical Products Administration for the diagnosis of infectious diseases. Therefore, the current mNGS processes are highly personalized, indicating the importance and difficulties

associated with supervising the clinical applications of this technology. Clinical regulation exists in all aspects, including for testing population selection, method standardization, quality control, and workflow validation. The expert consensus on the application of mNGS technology in the detection of infections of critical importance was released in China in 2019,<sup>[92]</sup> which recommended the following six indicators for mNGS detection: i) the pathogen should be identified as soon as possible in critical cases; ii) the pathogen should be identified as soon as possible for special patients such as immunosuppressed hosts, patients with underlying diseases, and patients with severe infection who were hospitalized repeatedly; iii) patients in which the traditional microbiological detection techniques were repeatedly negative and the treatment effect was not good; iv) cases in which the presence of new pathogens is suspected, indicating that there may be a certain degree of clinical infectivity; v) cases where special pathogen infections are suspected, and vi) long-term fever and/or presence of infection with additional clinical symptoms of unknown causes. Screening the test subjects and selecting the appropriate population for the test is the first step in clinical supervision, as these measures can help avoid wasted of resources and benefit patients who urgently require testing.

The mNGS experimental process is more complicated than other detection techniques as it involves sample preprocessing, nucleic acid extraction, library construction, sequencing, data analysis, and data interpretation. The more complex the steps, the more difficult it is to validate the whole workflow and establish a quality control system. At present, a few research teams have tried to establish quality control standards, however, no unified index has yet been generated.<sup>[106]</sup> Quality control is required for various steps in the workflow, including during the extraction, as well as for the concentration and purity of nucleic acids used, nucleic acid fragment distribution range, library concentration, the establishment of reference database standards, sequencing data volume, and pathogen screening principles. Additionally, different quality control standards may need to be established for different types of samples.

Currently, research validating the performance of mNGS assays in clinical applications remains scarce. In 2019, the Miller team performed a validation for mNGS using CSF samples, and comprehensively analyzed and evaluated the lowest detection limit, precision, accuracy, stability, and other indicators.<sup>[211]</sup> Similarly, the performance of mNGS in other types of samples such as BALF and sputum requires evaluation. We recommend that laboratories and third-party testing companies verify the performance of the workflow using different sample types before testing clinical samples to confirm the accuracy and stability of the workflow. This is particularly important as the formation of standardized detection methods that utilize mNGS technology will enable better clinical supervision. For example, this work depends on the joint efforts of relevant government agencies, scientific research institutions, and commercial companies. In May 2019, more than 20 units participated in a joint study launched by the National Institutes of Food and Drug Control to evaluate the quality of the metagenome second-generation sequencing reagents used for pathogen detection. The testing of different mNGS processes for unified samples promotes the research process for quality control and standardization. At present, there are great differences in

mNGS workflows among different units, and there is still a long way to go to achieve standardized detection procedures.

In summary, although there are many reported successful applications of mNGS in the clinical setting, most of them are case reports or retrospective studies. Thus, large-scale prospective studies with large sample sizes are needed to further demonstrate the reliability of this technology. These studies should focus on when and how to select mNGS, their diagnostic value compared, or combined with, traditional methods, how can mNGS help direct patients' management, including diagnosis and differential diagnosis of infectious diseases, guide treatment, assess efficacy and prognosis, in addition with cost-benefit analysis. In technology, the main goal should always be the further optimization of the sequencing process, database construction, and bioinformatic analysis, making mNGS more accurate, cost-effective, and improving its universality. Currently, there are few expert consensuses on the clinical application of mNGS, however, it is generally accepted that mNGS cannot yet replace the place of traditional methods in the short term, but is recommended as an auxiliary or necessary detection method for detecting microorganisms. We hope that high-quality clinical studies on mNGS will be carried out in the coming years, giving mNGS a greater role in clinical application. In addition, with the emergence of new pathogens in recent decades, mNGS, along with the continuous improvement of the regulatory network and popularization of third-generation sequencing technology, will also play an irreplaceable role in monitoring and tracking the outbreak of novel pathogen infections.

## Acknowledgements

N.L. and Q.C. contributed equally to this work. The authors acknowledge the special fund for clinical research of Zhongshan Hospital, Fudan University (No. 2018ZSLC06).

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

clinical application, infectious disease, metagenomics, next-generation sequencing

Received: August 31, 2020

Revised: October 24, 2020

Published online: December 13, 2020

- [1] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463.
- [2] T. Charalampous, G. L. Kay, H. Richardson, A. Aydin, R. Baldan, C. Jeanes, D. Rae, S. Grundy, D. J. Turner, J. Wain, R. M. Leggett, D. M. Livermore, J. O'Grady, *Nat. Biotechnol.* **2019**, *37*, 783.
- [3] P. J. Simner, S. Miller, K. C. Carroll, *Clin. Infect. Dis.* **2018**, *66*, 778.
- [4] F. Sanger, A. R. Coulson, *J. Mol. Biol.* **1975**, *94*, 441.
- [5] A. M. Maxam, W. Gilbert, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560.



- [6] A. T. Woolley, R. A. Mathies, *Anal. Chem.* **1995**, *67*, 3676.
- [7] Y. M. Deng, N. Spirason, P. Iannello, L. Jelley, H. Lau, I. G. Barr, *J. Clin. Virol.* **2015**, *68*, 43.
- [8] G. B. Cabral, J. L. P. Ferreira, R. P. Souza, M. S. Cunha, A. Luchs, C. A. Figueiredo, L. F. M. Brigido, *Mem. Inst. Oswaldo Cruz* **2018**, *113*, 38.
- [9] A. J. Sabat, E. van Zanten, V. Akkerboom, G. Wisselink, K. van Slochteren, R. F. de Boer, R. Hendrix, A. W. Friedrich, J. W. A. Rossen, A. Kooistra-Smid, *Sci. Rep.* **2017**, *7*, 3434.
- [10] S. Banos, G. Lentendu, A. Kopf, T. Wubet, F. O. Glockner, M. Reich, *BMC Microbiol.* **2018**, *18*, 190.
- [11] J. W. Ai, S. S. Weng, Q. Cheng, P. Cui, Y. J. Li, H. L. Wu, Y. M. Zhu, B. Xu, W. H. Zhang, *Emerging Infect. Dis.* **2018**, *24*, 1087.
- [12] D. Han, Z. Li, R. Li, P. Tan, R. Zhang, J. Li, *Crit. Rev. Microbiol.* **2019**, *45*, 668.
- [13] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. R. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, E. C. M. Chiara, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoshler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G. D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Racz, V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovskiy, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, A. J. Smith, *Nature* **2008**, *456*, 53.
- [14] Z. Hu, L. Cheng, H. Wang, *Methods Mol. Biol.* **2015**, *1231*, 91.
- [15] M. N. Wright, D. Gola, A. Ziegler, *Methods Mol. Biol.* **2017**, *1666*, 629.
- [16] G. Prakash, A. Kumar, N. Sheoran, R. Aggarwal, C. T. Satyavathi, S. K. Chikara, A. Ghosh, R. K. Jain, *Microbiol. Resour. Announce.* **2019**, *8*, e01499.
- [17] D. Korostin, N. Kulemin, V. Naumov, V. Belova, D. Kwon, A. Gorbachev, *PLoS One* **2020**, *15*, e0230301.
- [18] R. L. Marine, L. C. Magana, C. J. Castro, K. Zhao, A. M. Montmayeur, A. Schmidt, M. Diez-Valcarce, T. F. F. Ng, J. Vinje, C. C. Burns, W. A. Nix, P. A. Rota, M. S. Oberste, *J. Virol. Methods* **2020**, *280*, 113865.
- [19] M. R. Wilson, S. N. Naccache, E. Samayoa, M. Biagtan, H. Bashir, G. Yu, S. M. Salamat, S. Somasekar, S. Federman, S. Miller, R. Sokolic, E. Garabedian, F. Candotti, R. H. Buckley, K. D. Reed, T. L. Meyer, C. M. Seroogy, R. Galloway, S. L. Henderson, J. E. Gern, J. L. DeRisi, C. Y. Chiu, *N. Engl. J. Med.* **2014**, *370*, 2408.
- [20] X. C. Ji, L. F. Zhou, C. Y. Li, Y. J. Shi, M. L. Wu, Y. Zhang, X. F. Fei, G. Zhao, *J. Mol. Neurosci.* **2020**, *70*, 659.
- [21] C. Langelier, K. L. Kalantar, F. Moazed, M. R. Wilson, E. D. Crawford, T. Deiss, A. Belzer, S. Bolourchi, S. Caldera, M. Fung, A. Jauregui, K. Malcolm, A. Lyden, L. Khan, K. Vessel, J. Quan, M. Zinter, C. Y. Chiu, E. D. Chow, J. Wilson, S. Miller, M. A. Matthay, K. S. Pollard, S. Christenson, C. S. Calfee, J. L. DeRisi, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E12353.
- [22] M. R. Wilson, H. A. Sample, K. C. Zorn, S. Arevalo, G. Yu, J. Neuhaus, S. Federman, D. Stryke, B. Briggs, C. Langelier, A. Berger, V. Douglas, S. A. Josephson, F. C. Chow, B. D. Fulton, J. L. DeRisi, J. M. Gelfand, S. N. Naccache, J. Bender, J. Dien Bard, J. Murkey, M. Carlson, P. M. Vespa, T. Vijayan, P. R. Allyn, S. Campeau, R. M. Humphries, J. D. Klausner, C. D. Ganzon, F. Memar, N. A. Ocampo, L. L. Zimmermann, S. H. Cohen, C. R. Polage, R. L. DeBiasi, B. Haller, R. Dallas, G. Maron, R. Hayden, K. Messacar, S. R. Dominguez, S. Miller, C. Y. Chiu, *N. Engl. J. Med.* **2019**, *380*, 2327.
- [23] M. L. Carpenter, S. K. Tan, T. Watson, R. Bacher, V. Nagesh, A. Watts, G. Bentley, J. Weber, C. Huang, M. K. Sahoo, A. Hinterwirth, T. Doan, T. Carter, Q. Dong, S. Gourguechon, E. Harness, S. Kermes, S. Radhakrishnan, G. Wang, A. Quiroz-Zarate, J. Ching, B. A. Pinsky, *J. Clin. Microbiol.* **2019**, *57*, e01113.
- [24] X. Zhou, H. Wu, Q. Ruan, N. Jiang, X. Chen, Y. Shen, Y. M. Zhu, Y. Ying, Y. Y. Qian, X. Wang, J. W. Ai, W. H. Zhang, *Front. Cell. Infect. Microbiol.* **2019**, *9*, 351.
- [25] S. van Boheemen, A. L. van Rijn, N. Pappas, E. C. Carbo, R. H. P. Vorderman, I. Sidorov, P. J. van 't Hof, H. Mei, E. C. J. Claas, A. C. M. Kroes, J. J. C. de Vries, *J. Mol. Diagn.* **2020**, *22*, 196.
- [26] L. Chen, Y. Xu, C. Liu, H. Huang, X. Zhong, C. Ma, H. Zhao, Y. Chen, *BMC Infect. Dis.* **2020**, *20*, 435.
- [27] R. Drmanac, A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Keremian, P. Carnevali, I. Nazarenko, G. B. Nilsen, G. Yeung, F. Dahl, A. Fernandez, B. Staker, K. P. Pant, J. Baccash, A. P. Borchering, A. Brownley, R. Cedeno, L. Chen, D. Chernikoff, A. Cheung, R. Chirita, B. Curson, J. C. Ebert, C. R. Hacker, R. Hartlage, B. Hauser, S. Huang, Y. Jiang, V. Karpinchyk, M. Koenig, C. Kong, T. Landers, C. Le, J. Liu, C. E. McBride, M. Morensoni, R. E. Morey, K. Mutch, H. Perazich, K. Perry, B. A. Peters, J. Peterson, C. L. Pethiyagoda, K. Pothuraju, C. Richter, A. M. Rosenbaum, S. Roy, J. Shafto, U. Sharanovich, K. W. Shannon, C. G. Sheppy, M. Sun, J. V. Thakuria, A. Tran, D. Vu, A. W. Zaranek, X. Wu, S. Drmanac, A. R. Oliphant, W. C. Banyai, B. Martin, D. G. Ballinger, G. M. Church, C. A. Reid, *Science* **2010**, *327*, 78.
- [28] F. Y. Zhu, M. X. Chen, N. H. Ye, W. M. Qiao, B. Gao, W. K. Law, Y. Tian, D. Zhang, D. Zhang, T. Y. Liu, Q. J. Hu, Y. Y. Cao, Z. Z. Su, J. Zhang, Y. G. Liu, *Plant Methods* **2018**, *14*, 69.
- [29] H. C. Zhang, J. W. Ai, P. Cui, Y. M. Zhu, W. Hong-Long, Y. J. Li, W. H. Zhang, *J. Infect.* **2019**, *79*, 419.
- [30] S. Wang, Y. Chen, D. Wang, Y. Wu, D. Zhao, J. Zhang, H. Xie, Y. Gong, R. Sun, X. Nie, H. Jiang, J. Zhang, W. Li, G. Liu, X. Li, K. Huang, Y. Huang, Y. Li, H. Guan, S. Pan, Y. Hu, *Front. Microbiol.* **2019**, *10*, 1993.
- [31] L. M. Petersen, I. W. Martin, W. E. Moschetti, C. M. Kershaw, G. J. Tsongalis, *J. Clin. Microbiol.* **2019**, *58*, e01315.
- [32] S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat. Rev. Genet.* **2016**, *17*, 333.

- [33] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, *Science* **2009**, 323, 133.
- [34] X. Yu, W. Jiang, Y. Shi, H. Ye, J. Lin, *J. Cell. Mol. Med.* **2019**, 23, 7143.
- [35] A. S. Gonzalez-Reiche, M. M. Hernandez, M. J. Sullivan, B. Ciferri, H. Alshammary, A. Obla, S. Fabre, G. Kleiner, J. Polanco, Z. Khan, B. Albuquerque, A. van de Guchte, J. Dutta, N. Francoeur, B. S. Melo, I. Oussenko, G. Deikus, J. Soto, S. H. Sridhar, Y. C. Wang, K. Twyman, A. Kasarskis, D. R. Altman, M. Smith, R. Sebra, J. Aberg, F. Krammer, A. Garcia-Sastre, M. Luksa, G. Patel, A. Paniz-Mondolfi, M. Gitman, E. M. Sordillo, V. Simon, H. van Bakel, *Science* **2020**, 369, 297.
- [36] Y. Feng, Y. Zhang, C. Ying, D. Wang, C. Du, *Genomics, Proteomics Bioinf.* **2015**, 13, 4.
- [37] R. M. Leggett, M. D. Clark, *J. Exp. Bot.* **2017**, 68, 5419.
- [38] R. Bowden, R. W. Davies, A. Heger, A. T. Pagnamenta, M. de Cesare, L. E. Oikkonen, D. Parkes, C. Freeman, F. Dhalla, S. Y. Patel, N. Popitsch, C. L. C. Ip, H. E. Roberts, S. Salatino, H. Lockstone, G. Lunter, J. C. Taylor, D. Buck, M. A. Simpson, P. Donnelly, *Nat. Commun.* **2019**, 10, 1869.
- [39] M. Eisenstein, *Nat. Biotechnol.* **2012**, 30, 295.
- [40] S. Goodwin, J. Gurtowski, S. Ethe-Sayers, P. Deshpande, M. C. Schatz, W. R. McCombie, *Genome Res.* **2015**, 25, 1750.
- [41] C. L. C. Ip, M. Loose, J. R. Tyson, M. de Cesare, B. L. Brown, M. Jain, R. M. Leggett, D. A. Eccles, V. Zalunin, J. M. Urban, P. Piazza, R. J. Bowden, B. Paten, S. Mwaigwisya, E. M. Batty, J. T. Simpson, T. P. Snutch, E. Birney, D. Buck, S. Goodwin, H. J. Jansen, J. O'Grady, H. E. Olsen, MinION Analysis and Reference Consortium, *F1000Research* **2015**, 4, 1075.
- [42] F. J. Rang, W. P. Kloosterman, J. de Ridder, *Genome Biol.* **2018**, 19, 90.
- [43] J. Quick, N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouedraogo, B. Afrough, A. Bah, J. H. Baum, B. Becker-Zajia, J. P. Boettcher, M. Cabeza-Cabrero, A. Camino-Sanchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallash, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, R. Trina, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, B. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O'Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Woelfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keita, A. Rambaut, P. Formenty, S. Gunther, M. W. Carroll, *Nature* **2016**, 530, 228.
- [44] I. Prazsak, N. Moldovan, Z. Balazs, D. Tombacz, K. Megyeri, A. Szucs, Z. Csabai, Z. Boldogkoi, *BMC Genomics* **2018**, 19, 873.
- [45] A. Bainomugisa, T. Duarte, E. Lavu, S. Pandey, C. Coulter, B. J. Marais, L. M. Coin, *Microb. Genomics* **2018**, 4, e000188.
- [46] L. R. Runtuwene, J. S. B. Tuda, A. E. Mongan, W. Makalowski, M. C. Frith, M. Imwong, S. Srisutham, L. A. Nguyen Thi, N. N. Tuan, Y. Eshita, R. Maeda, J. Yamagishi, Y. Suzuki, *Sci. Rep.* **2018**, 8, 8286.
- [47] P. Farnia, F. Mohammadi, Z. Zarifi, D. J. Tabatabaei, J. Ganavi, K. Ghazisaeedi, P. K. Farnia, M. Gheydi, M. Bahadori, M. R. Masjedi, A. A. Velayati, *J. Clin. Microbiol.* **2002**, 40, 508.
- [48] L. Terranova, M. Oriano, A. Teri, L. Ruggiero, C. Tafuro, P. Marchisio, A. Gramegna, M. Contarini, E. Franceschi, S. Sottotetti, L. Cariani, A. Bevivino, J. D. Chalmers, S. Aliberti, F. Blasi, *Int. J. Mol. Sci.* **2018**, 19, 3256.
- [49] J. Guiot, S. Demarche, M. Henket, V. Paulus, S. Graff, F. Schleich, J. L. Corhay, R. Louis, C. Moermans, *J. Visualized Exp.* **2017**, 130, e56612.
- [50] H. J. Park, A. Woo, J. M. Cha, K. S. Lee, M. Y. Lee, *Sci. Rep.* **2018**, 8, 17994.
- [51] Y. Cai, X. Fang, Y. Chen, Z. Huang, C. Zhang, W. Li, B. Yang, W. Zhang, *Int. J. Infect. Dis.* **2020**, 96, 573.
- [52] Q. Wang, Q. Miao, J. Pan, W. Jin, Y. Ma, Y. Zhang, Y. Yao, Y. Su, Y. Huang, B. Li, M. Wang, N. Li, S. Cai, Y. Luo, C. Zhou, H. Wu, B. Hu, *Int. J. Infect. Dis.* **2020**, 100, 414.
- [53] P. S. Ramachandran, M. R. Wilson, *Nat. Rev. Neurol.* **2020**, 16, 547.
- [54] H. Frickmann, C. Kunne, R. M. Hagen, A. Podbielski, J. Normann, S. Poppert, M. Looso, B. Kreikemeyer, *BMC Microbiol.* **2019**, 19, 75.
- [55] C. Kohl, A. Brinkmann, P. W. Dabrowski, A. Radonic, A. Nitsche, A. Kurth, *Emerging Infect. Dis.* **2015**, 21, 48.
- [56] P. Dedhia, S. Tarale, G. Dhongde, R. Khadapkar, B. Das, *Asian Pac. J. Cancer Prev.* **2007**, 8, 55.
- [57] Y. Long, Y. Zhang, Y. Gong, R. Sun, L. Su, X. Lin, A. Shen, J. Zhou, Z. Caiji, X. Wang, D. Li, H. Wu, H. Tan, *Arch. Med. Res.* **2016**, 47, 365.
- [58] K. Helmersen, H. V. Aamot, *Sci. Rep.* **2020**, 10, 2985.
- [59] E. A. Bennett, D. Massilani, G. Lizzo, J. Daligault, E. M. Geigl, T. Grange, *BioTechniques* **2014**, 56, 289.
- [60] L. Di, Y. Fu, Y. Sun, J. Li, L. Liu, J. Yao, G. Wang, Y. Wu, K. Lao, R. W. Lee, G. Zheng, J. Xu, J. Oh, D. Wang, X. S. Xie, Y. Huang, J. Wang, *Proc. Natl. Acad. Sci. USA* **2020**, 117, 2886.
- [61] B. Lu, L. Dong, D. Yi, M. Zhang, C. Zhu, X. Li, C. Yi, *eLife* **2020**, 9, e54919.
- [62] J. Huang, X. Liang, Y. Xuan, C. Geng, Y. Li, H. Lu, S. Qu, X. Mei, H. Chen, T. Yu, N. Sun, J. Rao, J. Wang, W. Zhang, Y. Chen, S. Liao, H. Jiang, X. Liu, Z. Yang, F. Mu, S. Gao, *GigaScience* **2017**, 6, 1.
- [63] J. Wu, W. Zhang, S. Huang, Z. He, Y. Cheng, J. Wang, T. W. Lam, Z. Peng, S. M. Yiu, *Bioinformatics* **2013**, 29, 2971.
- [64] C. de Lannoy, D. de Ridder, J. Risse, *F1000Research* **2017**, 6, 1083.
- [65] H. Lu, F. Giordano, Z. Ning, *Genomics, Proteomics Bioinf.* **2016**, 14, 265.
- [66] M. Seki, E. Katsumata, A. Suzuki, S. Sereewattanawoot, Y. Sakamoto, J. Mizushima-Sugano, S. Sugano, T. Kohno, M. C. Frith, K. Tsuchihara, Y. Suzuki, *DNA Res.* **2019**, 26, 55.
- [67] C. Soneson, Y. Yao, A. Bratus-Neuenschwander, A. Patrignani, M. D. Robinson, S. Hussain, *Nat. Commun.* **2019**, 10, 3359.
- [68] A. Rhoads, K. F. Au, *Genomics, Proteomics Bioinf.* **2015**, 13, 278.
- [69] P. Coupland, T. Chandra, M. Quail, W. Reik, H. Swerdlow, *BioTechniques* **2012**, 53, 365.
- [70] A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **2014**, 30, 2114.
- [71] S. Chen, Y. Zhou, Y. Chen, J. Gu, *Bioinformatics* **2018**, 34, i884.
- [72] R. Schmieder, R. Edwards, *Bioinformatics* **2011**, 27, 863.
- [73] J. Piatigorsky, A. Tyler, *J. Cell Sci.* **1968**, 3, 515.
- [74] H. Li, R. Durbin, *Bioinformatics* **2009**, 25, 1754.
- [75] B. Langmead, S. L. Salzberg, *Nat. Methods* **2012**, 9, 357.
- [76] M. Eisenstein, *Nat. Biotechnol.* **2012**, 30, 295.
- [77] M. M. Haque, T. Bose, A. Dutta, C. V. Reddy, S. S. Mande, *Genomics* **2015**, 106, 116.
- [78] M. D. Czajkowski, D. P. Vance, S. A. Frese, G. Casaburi, *Bioinformatics* **2019**, 35, 2318.
- [79] D. H. Haft, M. DiCuccio, A. Badretdin, V. Brover, V. Chetvernin, K. O'Neill, W. Li, F. Chitsaz, M. K. Derbyshire, N. R. Gonzales, M. Gwadz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, C. Zheng, F. Thibaud-Nissen, L. Y. Geer, A. Marchler-Bauer, K. D. Pruitt, *Nucleic Acids Res.* **2018**, 46, D851.

- [80] R. Ounit, S. Wanamaker, T. J. Close, S. Lonardi, *BMC Genomics* **2015**, 16, 236.
- [81] S. Wittouck, S. Wuyts, C. J. Meehan, V. van Noort, S. Lebeer, *mSystems* **2019**, 4, e00264.
- [82] D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, *Genome Res.* **2015**, 25, 1043.
- [83] H. Sichtig, T. Minogue, Y. Yan, C. Stefan, A. Hall, L. Tallon, L. Sadzewicz, S. Nadendla, W. Klimke, E. Hatcher, M. Shumway, D. L. Aldea, J. Allen, J. Koehler, T. Slezak, S. Lovell, R. Schoepp, U. Scherf, *Nat. Commun.* **2019**, 10, 3313.
- [84] F. P. Breitwieser, D. N. Baker, S. L. Salzberg, *Genome Biol.* **2018**, 19, 198.
- [85] D. E. Wood, J. Lu, B. Langmead, *Genome Biol.* **2019**, 20, 257.
- [86] D. Kim, L. Song, F. P. Breitwieser, S. L. Salzberg, *Genome Res.* **2016**, 26, 1721.
- [87] D. T. Truong, E. A. Franzosa, T. L. Tickle, M. Scholz, G. Weingart, E. Pasolli, A. Tett, C. Huttenhower, N. Segata, *Nat. Methods* **2015**, 12, 902.
- [88] P. Menzel, K. L. Ng, A. Krogh, *Nat. Commun.* **2016**, 7, 11257.
- [89] B. Buchfink, C. Xie, D. H. Huson, *Nat. Methods* **2015**, 12, 59.
- [90] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, 215, 403.
- [91] O. M. Allicock, C. Guo, A. C. Uhlemann, S. Whittier, L. V. Chauhan, J. Garcia, A. Price, S. S. Morse, N. Mishra, T. Briesse, W. I. Lipkin, *mBio* **2018**, 9, e02007.
- [92] Consensus Group of Experts on Application of Metagenomic Next Generation Sequencing in the Pathogen Diagnosis in Clinical Moderate and Severe Infections, Professional Committee of Sepsis and Shock, Chinese Research Hospital Association, Professional Committee of Microbial Toxins Chinese Society for Microbiology, Professional Committee of Critical Care Medicine Shenzhen Medical Association, *Zhonghua Weizhongbing Jijiu Yixue* **2020**, 32, 531.
- [93] K. Mongkolrattanothai, S. N. Naccache, J. M. Bender, E. Samayoa, E. Pham, G. Yu, J. Dien Bard, S. Miller, G. Aldrovandi, C. Y. Chiu, *J. Pediatr. Infect. Dis. Soc.* **2017**, 6, 393.
- [94] F. P. Breitwieser, S. L. Salzberg, *Bioinformatics* **2020**, 36, 1303.
- [95] M. Maabar, A. J. Davison, M. Vucak, F. Thorburn, P. R. Murcia, R. Gunson, M. Palmarini, J. Hughes, *Virus Evol.* **2019**, 5, vez033.
- [96] G. S. Watts, J. E. Thornton Jr., K. Youens-Clark, A. J. Ponsero, M. J. Slepian, E. Menashi, C. Hu, W. Deng, D. G. Armstrong, S. Reed, L. D. Cranmer, B. L. Hurwitz, *PLoS Comput. Biol.* **2019**, 15, e1006863.
- [97] M. J. Strong, G. Xu, L. Morici, S. Splinter Bon-Durant, M. Baddoo, Z. Lin, C. Fewell, C. M. Taylor, E. K. Flemington, *PLoS Pathog.* **2014**, 10, e1004437.
- [98] B. D. Ondov, N. H. Bergman, A. M. Phillippy, *BMC Bioinformatics* **2011**, 12, 385.
- [99] Q. Miao, Y. Ma, Q. Wang, J. Pan, Y. Zhang, W. Jin, Y. Yao, Y. Su, Y. Huang, M. Wang, B. Li, H. Li, C. Zhou, C. Li, M. Ye, X. Xu, Y. Li, B. Hu, *Clin. Infect. Dis.* **2018**, 67, S231.
- [100] N. Mancini, S. Carletti, N. Ghidoli, P. Cichero, R. Burioni, M. Clementi, *Clin. Microbiol. Rev.* **2010**, 23, 235.
- [101] J. A. Fishman, *N. Engl. J. Med.* **2007**, 357, 2601.
- [102] M. Tomblyn, T. Chiller, H. Einsele, R. Gress, K. Sepkowitz, J. Storek, J. R. Wingard, J. A. Young, M. J. Boeckh, Center for International Blood and Marrow Research, National Marrow Donor Program, European Blood and Marrow Transplant Group, American Society of Blood and Marrow Transplantation, Canadian Blood and Marrow Transplant Group, Infectious Diseases Society of America, Society for Healthcare Epidemiology of America, Association of Medical Microbiology and Infectious Disease Canada, Centers for Disease Control and Prevention, *Biol. Blood Marrow Transplant.* **2009**, 15, 1143.
- [103] T. F. Barlam, S. E. Cosgrove, L. M. Abbo, C. MacDougall, A. N. Schuetz, E. J. Septimus, A. Srinivasan, T. H. Dellit, Y. T. Falck-Ytter, N. O. Fishman, C. W. Hamilton, T. C. Jenkins, P. A. Lipsett, P. N. Malani, L. S. May, G. J. Moran, M. M. Neuhauser, J. G. Newland, C. A. Ohl, M. H. Samore, S. K. Seo, K. K. Trivedi, *Clin. Infect. Dis.* **2016**, 62, e51.
- [104] M. Paul, V. Shani, E. Muchtar, G. Kariv, E. Robenshtok, L. Leibovici, *Antimicrob. Agents Chemother.* **2010**, 54, 4851.
- [105] C. A. Glaser, S. Honarmand, L. J. Anderson, D. P. Schnurr, B. Forghani, C. K. Cossen, F. L. Schuster, L. J. Christie, J. H. Tureen, *Clin. Infect. Dis.* **2006**, 43, 1565.
- [106] R. Schlager, C. Y. Chiu, S. Miller, G. W. Procop, G. Weinstock, The Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology, The Microbiology Resource Committee of the College of American Pathologists, *Arch. Pathol. Lab. Med.* **2017**, 141, 776.
- [107] G. S. Martin, D. M. Mannino, S. Eaton, M. Moss, *N. Engl. J. Med.* **2003**, 348, 1546.
- [108] S. Gupta, A. Sakhuja, G. Kumar, E. McGrath, R. S. Nanchal, K. B. Kashani, *Chest* **2016**, 150, 1251.
- [109] A. Kumar, P. Ellis, Y. Arabi, D. Roberts, B. Light, J. E. Parrillo, P. Dodek, G. Wood, A. Kumar, D. Simon, C. Peters, M. Ahsan, D. Chateau, Cooperative Antimicrobial Therapy of Septic Shock Database Research Group, *Chest* **2009**, 136, 1237.
- [110] I. De Vlaminck, L. Martin, M. Kertesz, K. Patel, M. Kowarsky, C. Strehl, G. Cohen, H. Luikart, N. F. Neff, J. Okamoto, M. R. Nicolls, D. Cornfield, D. Weill, H. Valentine, K. K. Khush, S. R. Quake, *Proc. Natl. Acad. Sci. USA* **2015**, 112, 13336.
- [111] W. Pan, T. T. M. Ngo, J. Camunas-Soler, C. X. Song, M. Kowarsky, Y. J. Blumenfeld, R. J. Wong, G. M. Shaw, D. K. Stevenson, S. R. Quake, *Clin. Chem.* **2017**, 63, 1695.
- [112] T. Gosiewski, A. H. Ludwig-Galezowska, K. Huminska, A. Sroka-Oleksiak, P. Radkowski, D. Salamon, J. Wojciechowicz, M. Kus-Slowinska, M. Bulanda, P. P. Wolkow, *Eur. J. Clin. Microbiol. Infect. Dis.* **2017**, 36, 329.
- [113] M. K. Abril, A. S. Barnett, K. Wegermann, E. Fountain, A. Strand, B. M. Heyman, B. A. Blough, A. C. Swaminathan, B. Sharma-Kuinkel, F. Ruffin, B. D. Alexander, C. M. McCall, S. F. Costa, M. O. Arcasoy, D. K. Hong, T. A. Blauwkamp, M. Kertesz, V. G. Fowler Jr., B. D. Kraft, *Open Forum Infect. Dis.* **2016**, 3, ofw144.
- [114] D. K. Hong, T. A. Blauwkamp, M. Kertesz, S. Bercovici, C. Truong, N. Banaei, *Diagn. Microbiol. Infect. Dis.* **2018**, 92, 210.
- [115] T. A. Blauwkamp, S. Thair, M. J. Rosen, L. Blair, M. S. Lindner, I. D. Vilfan, T. Kawi, F. C. Christians, S. Venkatasubrahmanyam, G. D. Wall, A. Cheung, Z. N. Rogers, G. Meshulam-Simon, L. Huijse, S. Balakrishnan, J. V. Quinn, D. Hollemon, D. K. Hong, M. L. Vaughn, M. Kertesz, S. Bercovici, J. C. Wilber, S. Yang, *Nat. Microbiol.* **2019**, 4, 663.
- [116] E. Crawford, J. Kamm, S. Miller, L. M. Li, S. Caldera, A. Lyden, D. Yokoe, A. Nichols, N. K. Tran, S. E. Barnard, P. M. Conner, A. Nambiar, M. S. Zinter, M. Moayeri, P. H. Serpa, B. C. Prince, J. Quan, R. Sit, M. Tan, M. Phelps, J. L. DeRisi, C. M. Tato, C. Langelier, *Clin. Infect. Dis.* **2019**, 71, 1179.
- [117] Y. Zhang, P. Cui, H. C. Zhang, H. L. Wu, M. Z. Ye, Y. M. Zhu, J. W. Ai, W. H. Zhang, *J. Transl. Med.* **2020**, 18, 199.
- [118] X. W. Xing, J. T. Zhang, Y. B. Ma, N. Zheng, F. Yang, S. Y. Yu, *J. Med. Microbiol.* **2019**, 68, 1204.
- [119] Z. Hu, X. Weng, C. Xu, Y. Lin, C. Cheng, H. Wei, W. Chen, *Ann. Clin. Microbiol. Antimicrob.* **2018**, 17, 45.
- [120] M. R. Wilson, B. D. O'Donovan, J. M. Gelfand, H. A. Sample, F. C. Chow, J. P. Betjemann, M. P. Shah, M. B. Richie, M. P. Gorman, R. A. Hajj-Ali, L. H. Calabrese, K. C. Zorn, E. D. Chow, J. E. Greenlee, J. H. Blum, G. Green, L. M. Khan, D. Banerji, C. Langelier, C. Bryson-Cahn, W. Harrington, J. R. Lingappa, N. M. Shanhag, A. J. Green, B. J. Brew, A. Soldatos, L. Strnad, S. B. Doernberg, C. A. Jay, V. Douglas, S. A. Josephson, J. L. DeRisi, *JAMA Neurol.* **2018**, 75, 947.
- [121] J. Granerod, H. E. Ambrose, N. W. Davies, J. P. Clewley, A. L. Walsh, D. Morgan, R. Cunningham, M. Zuckerman,



- K. J. Mutton, T. Solomon, K. N. Ward, M. P. Lunn, S. R. Irani, A. Vincent, D. W. Brown, N. S. Crowcroft, UK Health Protection Agency (HPA) Aetiology of Encephalitis Study Group, *Lancet Infect. Dis.* **2010**, 10, 835.
- [122] L. Y. Guo, Y. J. Li, L. L. Liu, H. L. Wu, J. L. Zhou, Y. Zhang, W. Y. Feng, L. Zhu, B. Hu, H. L. Hu, T. M. Chen, X. Guo, H. Y. Chen, Y. H. Yang, G. Liu, *J. Infect.* **2019**, 78, 323.
- [123] M. R. Wilson, D. Suan, A. Duggins, R. D. Schubert, L. M. Khan, H. A. Sample, K. C. Zorn, A. Rodrigues Hoffman, A. Blick, M. Shingde, J. L. DeRisi, *Ann. Neurol.* **2017**, 82, 105.
- [124] M. R. Wilson, L. L. Zimmermann, E. D. Crawford, H. A. Sample, P. R. Soni, A. N. Baker, L. M. Khan, J. L. DeRisi, *Am. J. Transplant.* **2017**, 17, 803.
- [125] C. Y. Chiu, L. L. Coffey, J. Murkey, K. Symmes, H. A. Sample, M. R. Wilson, S. N. Naccache, S. Arevalo, S. Somasekar, S. Federman, D. Stryke, P. Vespa, G. Schiller, S. Messenger, R. Humphries, S. Miller, J. D. Klausner, *Emerging Infect. Dis.* **2017**, 23, 1964.
- [126] J. A. Murkey, K. W. Chew, M. Carlson, C. L. Shannon, D. Sirohi, H. A. Sample, M. R. Wilson, P. Vespa, R. M. Humphries, S. Miller, J. D. Klausner, C. Y. Chiu, *Open Forum Infect. Dis.* **2017**, 4, ofx121.
- [127] V. Kufner, A. Plate, S. Schmutz, D. L. Braun, H. F. Gunthard, R. Capaul, A. Zbinden, N. J. Mueller, A. Trkola, M. Huber, *Genes* **2019**, 10, 661.
- [128] X. X. Zhang, L. Y. Guo, L. L. Liu, A. Shen, W. Y. Feng, W. H. Huang, H. L. Hu, B. Hu, X. Guo, T. M. Chen, H. Y. Chen, Y. Q. Jiang, G. Liu, *BMC Infect. Dis.* **2019**, 19, 495.
- [129] M. R. Wilson, N. M. Shanbhag, M. J. Reid, N. S. Singhal, J. M. Gelfand, H. A. Sample, B. Benkli, B. D. O'Donovan, I. K. Ali, M. K. Keating, T. H. Dunnebacke, M. D. Wood, A. Bollen, J. L. DeRisi, *Ann. Neurol.* **2015**, 78, 722.
- [130] X. W. Xing, J. T. Zhang, Y. B. Ma, M. W. He, G. E. Yao, W. Wang, X. K. Qi, X. Y. Chen, L. Wu, X. L. Wang, Y. H. Huang, J. Du, H. F. Wang, R. F. Wang, F. Yang, S. Y. Yu, *Front. Cell. Infect. Microbiol.* **2020**, 10, 88.
- [131] J. R. Brown, T. Bharucha, J. Breuer, *J. Infect.* **2018**, 76, 225.
- [132] L. Yan, W. Sun, Z. Lu, L. Fan, *Int. J. Infect. Dis.* **2020**, 96, 270.
- [133] K. T. Thakur, *Curr. HIV/AIDS Rep.* **2020**, 17, 507.
- [134] K. L. Tyler, *N. Engl. J. Med.* **2018**, 379, 557.
- [135] M. Fang, X. Weng, L. Chen, Y. Chen, Y. Chi, W. Chen, Z. Hu, *BMC Infect. Dis.* **2020**, 20, 159.
- [136] H. Xia, Y. Guan, S. D. Zaongo, H. Xia, Z. Wang, Z. Yan, P. Ma, *Front. Neurol.* **2019**, 10, 1202.
- [137] H. Guan, A. Shen, X. Lv, X. Yang, H. Ren, Y. Zhao, Y. Zhang, Y. Gong, P. Ni, H. Wu, Y. Zhu, L. Cui, *J. NeuroVirol.* **2016**, 22, 240.
- [138] J. Ai, Z. Xie, G. Liu, Z. Chen, Y. Yang, Y. Li, J. Chen, G. Zheng, K. Shen, *BMC Infect. Dis.* **2017**, 17, 494.
- [139] S. L. Salzberg, F. P. Breitwieser, A. Kumar, H. Hao, P. Burger, F. J. Rodriguez, M. Lim, A. Quinones-Hinojosa, G. L. Gallia, J. A. Tornheim, M. T. Melia, C. L. Sears, C. A. Pardo, *Neurol.: Neuroimmunol. Neuroinflammation* **2016**, 3, e251.
- [140] M. Yao, J. Zhou, Y. Zhu, Y. Zhang, X. Lv, R. Sun, A. Shen, H. Ren, L. Cui, H. Guan, H. Wu, *J. Clin. Neurol.* **2016**, 12, 446.
- [141] S. Fan, H. Ren, Y. Wei, C. Mao, Z. Ma, L. Zhang, L. Wang, Y. Ge, T. Li, L. Cui, H. Wu, H. Guan, *Int. J. Infect. Dis.* **2018**, 67, 20.
- [142] Q. Wang, J. Li, J. Ji, L. Yang, L. Chen, R. Zhou, Y. Yang, H. Zheng, J. Yuan, L. Li, Y. Bi, G. F. Gao, J. Ma, Y. Liu, *BMC Infect. Dis.* **2018**, 18, 349.
- [143] P. Liu, X. Weng, J. Zhou, X. Xu, F. He, Y. Du, H. Wu, Y. Gong, G. Peng, *BMC Infect. Dis.* **2018**, 18, 113.
- [144] S. Fan, X. Qiao, L. Liu, H. Wu, J. Zhou, R. Sun, Q. Chen, Y. Huang, C. Mao, J. Yuan, Q. Lu, Y. Ge, Y. Li, H. Ren, J. Wang, L. Cui, W. Zhao, H. Guan, *Front. Neurol.* **2018**, 9, 471.
- [145] R. He, W. Zheng, J. Long, Y. Huang, C. Liu, Q. Wang, Z. Yan, H. Liu, L. Xing, Y. Hu, H. Xie, *J. NeuroVirol.* **2019**, 25, 127.
- [146] J. Milucky, T. Pondo, C. J. Gregory, D. Iuliano, S. S. Chaves, J. McCracken, A. Mansour, Y. Zhang, M. A. Aleem, B. Wolff, B. Whitaker, T. Whistler, C. Onyango, M. R. Lopez, N. Liu, M. Z. Rahman, N. Shang, J. Winchell, M. Chittaganpitch, B. Fields, H. Maldonado, Z. Xie, S. Lindstrom, K. Sturm-Ramirez, J. Montgomery, K. H. Wu, C. A. Van Beneden, Adult TAC Working Group, *PLoS One* **2020**, 15, e0240309.
- [147] H. Li, H. Gao, H. Meng, Q. Wang, S. Li, H. Chen, Y. Li, H. Wang, *Front. Cell. Infect. Microbiol.* **2018**, 8, 205.
- [148] J. Huang, E. Jiang, D. Yang, J. Wei, M. Zhao, J. Feng, J. Cao, *Infect. Drug Resist.* **2020**, 13, 567.
- [149] A. L. van Rijn, S. van Boheemen, I. Sidorov, E. C. Carbo, N. Pappas, H. Mei, M. Feltkamp, M. Aanerud, P. Bakke, E. C. J. Claas, T. M. Eagan, P. S. Hiemstra, A. C. M. Kroes, J. J. C. de Vries, *PLoS One* **2019**, 14, e0223952.
- [150] Y. Xie, J. Du, W. Jin, X. Teng, R. Cheng, P. Huang, H. Xie, Z. Zhou, R. Tian, R. Wang, T. Feng, *J. Infect.* **2018**, 78, 158.
- [151] J. Wang, Y. Han, J. Feng, *BMC Pulm. Med.* **2019**, 19, 252.
- [152] N. Liu, J. Kan, W. Cao, J. Cao, E. Jiang, Y. Zhou, M. Zhao, J. Feng, *J. Int. Med. Res.* **2019**, 47, 4878.
- [153] H. Wang, Z. Lu, Y. Bao, Y. Yang, R. de Groot, W. Dai, M. I. de Jonge, Y. Zheng, *PLoS One* **2020**, 15, e0232610.
- [154] Y. Li, X. Deng, F. Hu, J. Wang, Y. Liu, H. Huang, J. Ma, J. Zhang, F. Zhang, C. Zhang, *J. Infect.* **2018**, 76, 311.
- [155] Y. Xie, J. Du, W. Jin, X. Teng, R. Cheng, P. Huang, H. Xie, Z. Zhou, R. Tian, R. Wang, T. Feng, *J. Infect.* **2019**, 78, 158.
- [156] Y. Li, B. Sun, X. Tang, Y. L. Liu, H. Y. He, X. Y. Li, R. Wang, F. Guo, Z. H. Tong, *Eur. J. Clin. Microbiol. Infect. Dis.* **2020**, 39, 369.
- [157] G. Li, J. Huang, Y. Li, J. Feng, *Can. Respir. J.* **2020**, 2020, 2367505.
- [158] K. N. Malizos, *J. Bone Jt. Surg., Am. Vol.* **2017**, 99, e20.
- [159] M. Thoendel, P. Jeraldo, K. E. Greenwood-Quaintance, N. Chia, M. P. Abdel, J. M. Steckelberg, D. R. Osmon, R. Patel, *Clin. Infect. Dis.* **2017**, 65, 332.
- [160] X. Fang, Y. Cai, T. Shi, Z. Huang, C. Zhang, W. Li, C. Zhang, B. Yang, W. Zhang, Z. Guan, *Int. J. Infect. Dis.* **2020**, 99, 108.
- [161] Z. Huang, W. Li, G. C. Lee, X. Fang, L. Xing, B. Yang, J. Lin, W. Zhang, *Bone Jt. Res.* **2020**, 9, 440.
- [162] T. L. Street, N. D. Sanderson, B. L. Atkins, A. J. Brent, K. Cole, D. Foster, M. A. McNally, S. Oakley, L. Peto, A. Taylor, T. E. A. Peto, D. W. Crook, D. W. Eyre, *J. Clin. Microbiol.* **2017**, 55, 2334.
- [163] M. I. Ivy, M. J. Thoendel, P. R. Jeraldo, K. E. Greenwood-Quaintance, A. D. Hanssen, M. P. Abdel, N. Chia, J. Z. Yao, A. J. Tande, J. N. Mandrekar, R. Patel, *J. Clin. Microbiol.* **2018**, 56, e00402.
- [164] Z. Huang, C. Zhang, W. Li, X. Fang, Q. Wang, L. Xing, Y. Li, X. Nie, B. Yang, W. Zhang, *J. Bone Jt. Infect.* **2019**, 4, 50.
- [165] Z. Huang, C. Zhang, X. Fang, W. Li, C. Zhang, W. Zhang, B. Yang, *J. Infect.* **2019**, 78, 158.
- [166] M. Li, F. Yang, Y. Lu, W. Huang, *BMC Infect. Dis.* **2020**, 20, 467.
- [167] V. Mouraviev, M. McDonald, *Can. J. Urol.* **2018**, 25, 9349.
- [168] M. Q. Li Na, W. Qingqing, C. Sishi, M. Yuyan, S. Yi, P. Jue, H. Bijie, *Natl. Med. J. China* **2020**, 100, 2198.
- [169] A. L. Flores-Mireles, J. N. Walker, M. Caparon, S. J. Hultgren, *Nat. Rev. Microbiol.* **2015**, 13, 269.
- [170] Y. Zhou, K. M. Wylie, R. E. El Feghaly, K. A. Mihindukulasuriya, A. Elward, D. B. Haslam, G. A. Storch, G. M. Weinstock, *J. Clin. Microbiol.* **2016**, 54, 368.
- [171] A. Ramesh, S. Nakielnny, J. Hsu, M. Kyohere, O. Byaruhanga, C. de Bourcy, R. Egger, B. Dimitrov, Y. F. Juan, J. Sheu, J. Wang, K. Kalantar, C. Langelier, T. Ruel, A. Mpimbaza, M. R. Wilson, P. J. Rosenthal, J. L. DeRisi, *PLoS One* **2019**, 14, e0218318.
- [172] M. Kujiraoka, M. Kuroda, K. Asai, T. Sekizuka, K. Kato, M. Watanabe, H. Matsukiyo, T. Saito, T. Ishii, N. Katada, Y. Saida, S. Kusachi, *Front. Microbiol.* **2017**, 8, 685.
- [173] S. Somasekar, D. Lee, J. Rule, S. N. Naccache, M. Stone, M. P. Busch, C. Sanders, W. M. Lee, C. Y. Chiu, *Clin. Infect. Dis.* **2017**, 65, 1477.



- [174] R. Yee, F. P. Breitwieser, S. Hao, B. N. A. Opene, R. E. Workman, P. D. Tamma, J. Dien-Bard, W. Timp, P. J. Simner, *Eur. J. Clin. Microbiol. Infect. Dis.* **2020**, 39, 1011.
- [175] K. V. Sukhum, L. Diorio-Toth, G. Dantas, *Clin. Pharmacol. Ther.* **2019**, 106, 512.
- [176] X. Chen, K. Cao, Y. Wei, Y. Qian, J. Liang, D. Dong, J. Tang, Z. Zhu, Q. Gu, W. Yu, *Infection* **2020**, 48, 535.
- [177] L. Gu, W. Liu, M. Ru, J. Lin, G. Yu, J. Ye, Z. A. Zhu, Y. Liu, J. Chen, G. Lai, W. Wen, *BMC Pulm. Med.* **2020**, 20, 65.
- [178] Y. Huang, Y. Ma, Q. Miao, J. Pan, B. Hu, Y. Gong, Y. Lin, *Ann. Transl. Med.* **2019**, 7, 589.
- [179] W. Jin, Q. Miao, M. Wang, Y. Zhang, Y. Ma, Y. Huang, H. Wu, Y. Lin, B. Hu, J. Pan, *Ann. Transl. Med.* **2020**, 8, 247.
- [180] M. Quan, L. Liu, T. Zhou, Y. Jiang, X. Wang, Z. Zong, *Wien. Klin. Wochenschr.* **2020**, 132, 589.
- [181] Y. Dai, L. Chen, W. Chang, H. Lu, P. Cui, X. Ma, *Front. Public Health* **2019**, 7, 379.
- [182] J. Wu, Y. Wu, M. Huang, *Int. J. Infect. Dis.* **2020**, 90, 1.
- [183] D. Kumar, S. Chaudhary, N. Lu, M. Duff, M. Heffel, C. A. McKinney, D. Bedenice, D. Marthaler, *Viruses* **2019**, 11, 701.
- [184] T. Li, P. Mbala-Kingebeni, S. N. Naccache, J. Theze, J. Bouquet, S. Federman, S. Somasekar, G. Yu, C. Sanchez-San Martin, A. Achari, B. S. Schneider, A. W. Rimoin, A. Rambaut, J. Nsio, P. Mulembakani, S. Ahuka-Mundeke, J. Kapetshi, O. G. Pybus, J. J. Muyembe-Tamfum, C. Y. Chiu, *J. Clin. Microbiol.* **2019**, 57, e00827.
- [185] S. I. Sardi, S. Somasekar, S. N. Naccache, A. C. Bandeira, L. B. Tauro, G. S. Campos, C. Y. Chiu, *J. Clin. Microbiol.* **2016**, 54, 2348.
- [186] B. Hoffmann, D. Tappe, D. Hoper, C. Herden, A. Boldt, C. Mawrin, O. Niederstrasser, T. Muller, M. Jenckel, E. van der Grinten, C. Lutter, B. Abendroth, J. P. Teifke, D. Cadar, J. Schmidt-Chanasit, R. G. Ulrich, M. Beer, *N. Engl. J. Med.* **2015**, 373, 154.
- [187] F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes, Y. Z. Zhang, *Nature* **2020**, 579, 265.
- [188] P. Parize, E. Muth, C. Richaud, M. Gratigny, B. Pilmis, A. Lamamy, J.-L. Mainardi, J. Cheval, L. de Visser, F. Jagorel, L. Ben Yahia, G. Bamba, M. Dubois, O. Join-Lambert, M. Leruez-Ville, X. Nassif, A. Lefort, F. Lanternier, F. Suarez, O. Lortholary, M. Lecuit, M. Eloit, *Clin Microbiol. Infect.* **2017**, 23, 574.e1.
- [189] M. Ye, W. Wei, Z. Yang, Y. Li, S. Cheng, K. Wang, T. Zhou, J. Sun, S. Liu, N. Ni, H. Jiang, H. Jiang, *BMC Infect. Dis.* **2016**, 16, 5.
- [190] A. Bal, C. Sarkozy, L. Josset, V. Cheynet, G. Oriol, J. Becker, G. Vilchez, P. Sesques, F. Mallet, A. Pachot, F. Morfin, B. Lina, G. Salles, F. Reynier, S. Trouillet-Assant, K. Brengel-Pesce, *Viruses* **2018**, 10, 633.
- [191] D. W. Lewandowska, P. W. Schreiber, M. M. Schuurmans, B. Ruehe, O. Zagordi, C. Bayard, M. Greiner, F. D. Geissberger, R. Capaul, A. Zbinden, J. Boni, C. Benden, N. J. Mueller, A. Trkola, M. Huber, *PLoS One* **2017**, 12, e0177340.
- [192] I. De Vlamincq, H. A. Valentine, T. M. Snyder, C. Strehl, G. Cohen, H. Luikart, N. F. Neff, J. Okamoto, D. Bernstein, D. Weissshaar, S. R. Quake, K. K. Khush, *Sci. Transl. Med.* **2014**, 6, 241ra77.
- [193] I. De Vlamincq, K. K. Khush, C. Strehl, B. Kohli, H. Luikart, N. F. Neff, J. Okamoto, T. M. Snyder, D. N. Cornfield, M. R. Nicolls, D. Weill, D. Bernstein, H. A. Valentine, S. R. Quake, *Cell* **2013**, 155, 1178.
- [194] F. M. Fusco, R. Pisapia, S. Nardiello, S. D. Cicala, G. B. Gaeta, G. Brancaccio, *BMC Infect. Dis.* **2019**, 19, 653.
- [195] K. M. Wylie, K. A. Mihindukulasuriya, E. Sodergren, G. M. Weinstock, G. A. Storch, *PLoS One* **2012**, 7, e27735.
- [196] H. Jerome, C. Taylor, V. B. Sreenu, T. Klymenko, A. D. S. Filipe, C. Jackson, C. Davis, S. Ashraf, E. Wilson-Davies, N. Jesudason, K. Devine, L. Harder, C. Aitken, R. Gunson, E. C. Thomson, *J. Infect.* **2019**, 79, 383.
- [197] J. H. Chai, C. K. Lee, H. K. Lee, N. Wong, K. Teo, C. S. Tan, P. Thokala, J. W. Tang, P. A. Tambyah, V. M. S. Oh, T. P. Loh, J. Yoong, *PLoS One* **2018**, 13, e0194648.
- [198] V. Peddu, R. C. Shean, H. Xie, L. Shrestha, G. A. Perchetti, S. S. Minot, P. Roychoudhury, M. L. Huang, A. Nalla, S. B. Reddy, Q. Phung, A. Reinhardt, K. R. Jerome, A. L. Greninger, *Clin. Chem.* **2020**, 66, 966.
- [199] P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, Z. L. Shi, *Nature* **2020**, 579, 270.
- [200] J. W. Ai, Y. Zhang, H. C. Zhang, T. Xu, W. H. Zhang, *Emerg Microbes Infect.* **2020**, 9, 597.
- [201] J. Lu, L. du Plessis, Z. Liu, V. Hill, M. Kang, H. Lin, J. Sun, S. Francois, M. U. G. Kraemer, N. R. Faria, J. T. McCrone, J. Peng, Q. Xiong, R. Yuan, L. Zeng, P. Zhou, C. Liang, L. Yi, J. Liu, J. Xiao, J. Hu, T. Liu, W. Ma, W. Li, J. Su, H. Zheng, B. Peng, S. Fang, W. Su, K. Li, R. Sun, R. Bai, X. Tang, M. Liang, J. Quick, T. Song, A. Rambaut, N. Loman, J. Raghwan, O. G. Pybus, C. Ke, *Cell* **2020**, 181, 997.
- [202] B. Allegranzi, S. Bagheri Nejad, C. Combescure, W. Graafmans, H. Attar, L. Donaldson, D. Pittet, *Lancet* **2011**, 377, 228.
- [203] A. L. Greninger, D. M. Zerr, X. Qin, A. L. Adler, R. Sampoleo, J. M. Kuypers, J. A. Englund, K. R. Jerome, *J. Clin. Microbiol.* **2017**, 55, 177.
- [204] A. L. Greninger, S. N. Naccache, K. Messacar, A. Clayton, G. Yu, S. Somasekar, S. Federman, D. Stryke, C. Anderson, S. Yagi, S. Messenger, D. Wadford, D. Xia, J. P. Watt, K. Van Haren, S. R. Dominguez, C. Glaser, G. Aldrovandi, C. Y. Chiu, *Lancet Infect. Dis.* **2015**, 15, 671.
- [205] A. L. Greninger, A. Waghmare, A. Adler, X. Qin, J. L. Crowley, J. A. Englund, J. M. Kuypers, K. R. Jerome, D. M. Zerr, *J. Pediatr. Infect. Dis. Soc.* **2017**, 6, 168.
- [206] N. J. Loman, C. Constantinidou, M. Christner, H. Rohde, J. Z. Chan, J. Quick, J. C. Weir, C. Quince, G. P. Smith, J. R. Betley, M. Aepfelbacher, M. J. Pallen, *JAMA, J. Am. Med. Assoc.* **2013**, 309, 1502.
- [207] C. A. Marotz, J. G. Sanders, C. Zuniga, L. S. Zaramela, R. Knight, K. Zengler, *Microbiome* **2018**, 6, 42.
- [208] A. Bal, M. Pichon, C. Picard, J. S. Casalegno, M. Valette, I. Schuffenecker, L. Billard, S. Vallet, G. Vilchez, V. Cheynet, G. Oriol, S. Trouillet-Assant, Y. Gillet, B. Lina, K. Brengel-Pesce, F. Morfin, L. Josset, *BMC Infect. Dis.* **2018**, 18, 537.
- [209] J. J. Jacob, B. Veeraraghavan, K. Vasudevan, *Indian J. Med. Microbiol.* **2019**, 37, 133.
- [210] G. R. Feherly, E. Yigit, S. O. Oyola, B. W. Langhorst, V. T. Schmidt, F. J. Stewart, E. T. Dimalanta, L. A. Amaral-Zettler, T. Davis, M. A. Quail, S. Pradhan, *PLoS One* **2013**, 8, e76096.
- [211] S. Miller, S. N. Naccache, E. Samayoa, K. Messacar, S. Arevalo, S. Federman, D. Stryke, E. Pham, B. Fung, W. J. Bolosky, D. Ingebrigtsen, W. Lorizio, S. M. Paff, J. A. Leake, R. Pesano, R. DeBiasi, S. Dominguez, C. Y. Chiu, *Genome Res.* **2019**, 29, 831.
- [212] S. S. Zhang, D. Chen, Q. Lu, *Genet. Mol. Res.* **2012**, 11, 1532.
- [213] O. M. de Bruin, H. C. Birnboim, *BMC Microbiol.* **2016**, 16, 197.
- [214] W. K. Aldous, J. I. Pounder, J. L. Cloud, G. L. Woods, *J. Clin. Microbiol.* **2005**, 43, 2471.
- [215] L. Chen, W. Liu, Q. Zhang, K. Xu, G. Ye, W. Wu, Z. Sun, F. Liu, K. Wu, B. Zhong, Y. Mei, W. Zhang, Y. Chen, Y. Li, M. Shi, K. Lan, Y. Liu, *Emerging Microbes Infect.* **2020**, 9, 313.
- [216] W. Gu, S. Miller, C. Y. Chiu, *Annu. Rev. Pathol.* **2019**, 14, 319.
- [217] I. E. Pereira, M. A. , Guedes, R. L. M. , **2017**.
- [218] C. S. Pareek, R. Smoczynski, A. Tretyn, *J. Appl. Genet.* **2011**, 52, 413.
- [219] The Arabidopsis Genome Initiative, *Nature* **2000**, 408, 796.

- [220] M. Ye, J. Tu, J. Jiang, Y. Bi, W. You, Y. Zhang, J. Ren, T. Zhu, Z. Cao, Z. Yu, C. Shao, Z. Shen, B. Ding, J. Yuan, X. Zhao, Q. Guo, X. Xu, J. Huang, M. Wang, *Front. Infect. Microbiol.* **2016**, 6, 165.
- [221] D. Seto, *Viruses* **2010**, 2, 2587.
- [222] M. Stano, G. Beke, L. Klucar, *Database* **2016**, 2016, baw162.
- [223] E. J. Lefkowitz, D. M. Dempsey, R. C. Hendrickson, R. J. Orton, S. G. Siddell, D. B. Smith, *Nucleic Acids Res.* **2018**, 46, D708.
- [224] C. H. Calisher, T. Biese, J. R. Brister, R. N. Charrel, R. Durrwald, H. Ebihara, C. F. Fulhorst, G. F. Gao, M. H. Groschup, A. D. Haddow, T. H. Hyndman, S. Junglen, B. Klempa, J. Klingstrom, A. M. Kropinski, M. Krupovic, A. D. LaBeaud, P. Maes, N. Nowotny, M. R. T. Nunes, S. L. Payne, S. R. Radoshitzky, D. Rubbenstroth, S. Sabanadzovic, T. Sasaya, M. D. Stenglein, A. Varsani, V. Wahl, S. C. Weaver, F. M. Zerbini, N. Vasilakis, J. H. Kuhn, *Syst. Biol.* **2019**, 68, 828.
- [225] S. Roux, E. M. Adriaenssens, B. E. Dutilh, E. V. Koonin, A. M. Kropinski, M. Krupovic, J. H. Kuhn, R. Lavigne, J. R. Brister, A. Varsani, C. Amid, R. K. Aziz, S. R. Bordenstein, P. Bork, M. Breitbart, G. R. Cochrane, R. A. Daly, C. Desnues, M. B. Duhaime, J. B. Emerson, F. Enault, J. A. Fuhrman, P. Hingamp, P. Hugenholtz, B. L. Hurwitz, N. N. Ivanova, J. M. Labonte, K. B. Lee, R. R. Malmstrom, M. Martinez-Garcia, I. K. Mizrahi, H. Ogata, D. Paez-Espino, M. A. Petit, C. Putonti, T. Rattei, A. Reyes, F. Rodriguez-Valera, K. Rosario, L. Schriml, F. Schulz, G. F. Steward, M. B. Sullivan, S. Sunagawa, C. A. Suttle, B. Temperton, S. G. Tringe, R. V. Thurber, N. S. Webster, K. L. Whiteson, S. W. Wilhelm, K. E. Wommack, T. Woyke, K. C. Wrighton, P. Yilmaz, T. Yoshida, M. J. Young, N. Yutin, L. Z. Allen, N. C. Kyrpides, E. A. Elie-Fadrosh, *Nat. Biotechnol.* **2019**, 37, 29.
- [226] G. Mahmoudabadi, R. Phillips, *eLife* **2018**, 7, e31955.
- [227] K. Y. Zhang, Y. Z. Gao, M. Z. Du, S. Liu, C. Dong, F. B. Guo, *Front. Microbiol.* **2019**, 10, 184.
- [228] N. Goodacre, A. Aljanahi, S. Nandakumar, M. Mikailov, A. S. Khan, *mSphere* **2018**, 3, e00069.
- [229] J. R. Brister, D. Ako-Adjei, Y. Bao, O. Blinkova, *Nucleic Acids Res.* **2015**, 43, D571.
- [230] M. S. Zinter, C. C. Dvorak, M. Y. Mayday, K. Iwanaga, N. P. Ly, M. E. McGarry, G. D. Church, L. E. Faricy, C. M. Rowan, J. R. Hume, M. E. Steiner, E. D. Crawford, C. Langelier, K. Kalantar, E. D. Chow, S. Miller, K. Shimano, A. Melton, G. A. Yanik, A. Sapru, J. L. DeRisi, *Clin. Infect. Dis.* **2019**, 68, 1847.
- [231] S. N. Naccache, A. L. Greninger, D. Lee, L. L. Coffey, T. Phan, A. Rein-Weston, A. Aronsohn, J. Hackett Jr., E. L. Delwart, C. Y. Chiu, *J. Virol.* **2013**, 87, 11966.
- [232] S. J. Salter, M. J. Cox, E. M. Turek, S. T. Calus, W. O. Cookson, M. F. Moffatt, P. Turner, J. Parkhill, N. J. Loman, A. W. Walker, *BMC Biol.* **2014**, 12, 87.
- [233] M. A. Malla, A. Dubey, A. Kumar, S. Yadav, A. Hashem, E. F. Abd Allah, *Front. Immunol.* **2018**, 9, 2868.
- [234] M. P. Cheng, J. Papenburg, M. Desjardins, S. Kanjilal, C. Quach, M. Libman, S. Dittrich, C. P. Yansouni, *Ann. Intern. Med.* **2020**, 172, 726.
- [235] A. L. Greninger, *Expert Rev. Mol. Diagn.* **2018**, 18, 605.
- [236] C. L. Shi, P. Han, P. J. Tang, M. M. Chen, Z. J. Ye, M. Y. Wu, J. Shen, H. Y. Wu, Z. Q. Tan, X. Yu, G. H. Rao, J. P. Zhang, *J. Infect.* **2020**, 81, 567.
- [237] S. Saha, A. Ramesh, K. Kalantar, R. Malaker, M. Hasanuzzaman, L. M. Khan, M. Y. Mayday, M. S. I. Sajib, L. M. Li, C. Langelier, H. Rahman, E. D. Crawford, C. M. Tato, M. Islam, Y. F. Juan, C. de Bourcy, B. Dimitrov, J. Wang, J. Tang, J. Sheu, R. Egger, T. R. De Carvalho, M. R. Wilson, S. K. Saha, J. L. DeRisi, *mBio* **2019**, 10, e02877.
- [238] H. Zhou, X. Chen, T. Hu, J. Li, H. Song, Y. Liu, P. Wang, D. Liu, J. Yang, E. C. Holmes, A. C. Hughes, Y. Bi, W. Shi, *Curr. Biol.* **2020**, 30, 2196.
- [239] P. Chen, W. Sun, Y. He, *J. Thorac. Dis.* **2020**, 12, 4014.
- [240] M. Zhao, K. Tang, F. Liu, W. Zhou, J. Fan, G. Yan, S. Qin, Y. Pang, *Front. Microbiol.* **2020**, 11, 2034.
- [241] J. Chen, T. He, X. Li, X. Wang, L. Peng, L. Ma, *Infect. Drug Resist.* **2020**, 13, 2829.
- [242] S. Ardui, A. Ameur, J. R. Vermeesch, M. S. Hestand, *Nucleic Acids Res.* **2018**, 46, 2159.
- [243] H. Ji, E. Enns, C. J. Brumme, N. Parkin, M. Howison, E. R. Lee, R. Capina, E. Marinier, S. Avila-Rios, P. Sandstrom, G. Van Domselaar, R. Harrigan, R. Paredes, R. Kantor, M. Noguera-Julian, *J. Int. AIDS Soc.* **2018**, 21, e25193.
- [244] J. M. Elkins, S. Kates, J. Lange, J. Lange, P. Lichstein, J. Otero, A. Soriano, C. Wagner, M. Wouthuyzen-Bakker, *J. Arthroplasty* **2019**, 34, S181.



**Na Li** is a resident in the Department of Infectious Diseases of Zhongshan Hospital, Fudan University. She received her Master's degree from Fudan University. Her research interest is focused on molecular diagnosis of infectious diseases and alternative therapeutic strategies for multiple drug resistant organisms.



**Bijie Hu**, a professor and present director of the Department of Infectious Diseases of Zhongshan Hospital, Fudan University, graduated from Shanghai Medical College of Fudan University with a Ph.D. degree. He is also an expert on multiple drug resistant organisms and nosocomial infection control at the World Health Organization, and a director of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) collaboration center. He has expertise in various bacterial and fungal infections, especially in the diagnosis of difficult pulmonary infections, as well as accurate and early determination of the pathogens.