



Detection of Pulmonary Infectious Pathogens From Lung Biopsy Tissues by Metagenomic Next-Generation Sequencing

Henan Li¹, Hua Gao¹, Han Meng¹, Qi Wang¹, Shuguang Li¹, Hongbin Chen¹, Yongjun Li² and Hui Wang^{1*}

¹ Department of Clinical Laboratory, Peking University People's Hospital, Beijing, China, ² BGI Genomics, Beijing Genomics Institute-Shenzhen, Shenzhen, China

OPEN ACCESS

Edited by:

Dongsheng Zhou,
Beijing Institute of Microbiology and
Epidemiology, China

Reviewed by:

Lili Ren,
Institute of Pathogen Biology (CAMS),
China
Adrian Ochoa-Leyva,
Universidad Nacional Autónoma de
México, Mexico
Jiazhen Chen,
Fudan University, China

*Correspondence:

Hui Wang
wanghui@pkuph.edu.cn

Received: 29 March 2018

Accepted: 05 June 2018

Published: 25 June 2018

Citation:

Li H, Gao H, Meng H, Wang Q, Li S,
Chen H, Li Y and Wang H (2018)
Detection of Pulmonary Infectious
Pathogens From Lung Biopsy Tissues
by Metagenomic Next-Generation
Sequencing.
Front. Cell. Infect. Microbiol. 8:205.
doi: 10.3389/fcimb.2018.00205

Metagenomic next-generation sequencing (mNGS) is a comprehensive approach for sequence-based identification of pathogenic microbes. However, reports on the use of mNGS in pulmonary infection applied to lung biopsy tissues remain scarce. In this study, we applied mNGS to detect the presence of pathogenic microbes in lung biopsy tissues from 20 patients with pulmonary disorders indicating possible infection. We applied a new data management for identifying pathogen species based on mNGS data. We determined the thresholds for the unique reads and relative abundance required to identify the infectious pathogens. Potential pathogens of pulmonary infections in 15 patients were identified by mNGS. The comparison between mNGS and culture method resulted that the sensitivity and specificity were 100.0% (95% CI: 31.0–100.0%) and 76.5% (95% CI: 49.8–92.2%) for bacteria, 57.1% (95% CI: 20.2–88.2%) and 61.5% (95% CI: 32.2–84.9%) for fungi. The positive predictive value (PPV) (42.9% for bacteria, 44.4% for fungi) was much lower than negative predictive value (NPV) (100% for bacteria, 72.7% for fungi) in mNGS vs. culture method. The mNGS showed the highest specificity (100.0 and 94.1%) and PPV (100.0 and 75.0%) in the evaluation of fungi and MTBC respectively, when compared with histopathology method. The study indicated that mNGS of lung biopsy tissues can be used to detect the presence (or absence) of pulmonary pathogens in patients, with potential benefits in speed and sensitivity. However, accurate data management and interpretation of mNGS are required, and should be combined with observations of clinical manifestations and conventional laboratory-based diagnostic methods.

Keywords: metagenomic next-generation sequencing (mNGS), pulmonary infection, lung biopsy tissues, data management, clinical diagnosis

INTRODUCTION

Pulmonary infection is a leading cause of death and morbidity worldwide (Magill et al., 2014). However, its diagnosis is challenging due to the multitude of possible pathogens. Hundreds of pathogens have been associated with pulmonary infections, including bacterial, viral, or fungal pathogens (Renaud and Campbell, 2011; Ruppé et al., 2016; De La Cruz and Silveira, 2017). In

immunocompromised patients, virtually any bacteria or fungus can be considered as a potential pneumonia-causing pathogen.

A rapid microbiological diagnosis of pulmonary infections facilitates the timely application of antimicrobial therapy. Smear by microscopy and culture are the conventional microbiological methods used to identify pathogens, but both methods are relatively insensitive and culture is time-consuming. Histopathological diagnosis is the diagnostic gold standard of invasive fungal infections. However, it requires time and is not pathogen-specific. Rapid advances in sequencing technology and bioinformatics have made metagenomic next-generation sequencing (mNGS) a fertile area for the development of clinical diagnostics (Hilton et al., 2016; Salzberg et al., 2016; Somasekar et al., 2017). Culture-independent screening for pathogens with mNGS only needs a small amount of DNA directly taken from the sample, and a bioinformatics tool, which identifies pathogens by linking sequencing reads to an accurate reference genome (or marker) database (Forbes et al., 2017). Eventually, if the sequence depth is sufficient enough, the antibiotic susceptibility of pathogens can be inferred. Recent work has highlighted the current interest in using mNGS for the identification and antibiotic susceptibility testing of pathogens in the diagnosis of viral acute encephalitis (Naccache et al., 2015), infective endocarditis (Fukui et al., 2015), and bacterial meningitis (Salzberg et al., 2016). However, reports on the use of mNGS in pulmonary infection applied to lung biopsy tissues remain scarce.

In particular, we applied mNGS on 20 computer tomography (CT)-guided puncture lung biopsy tissues collected from patients with suspected pulmonary infections. After mNGS, we analyzed the relative abundance, coverage, depth, and unique reads of microbial sequences mapping to fungal and bacterial reference genomes from NCBI. The results obtained by mNGS were compared with those from conventional laboratory-based diagnostic methods. Our results indicated that mNGS provided a major new opportunity to investigate the pathogens of pulmonary infections.

MATERIALS AND METHODS

Ethics Statement

Application for ethical review was submitted to the Ethical Review Committee of Peking University People's Hospital (Reference 2017PHB075). The study was considered exempt from ethical review as it was a retrospective study and patients were anonymized.

Specimen Collection and Processing

CT-guided puncture lung biopsy tissues with suspected pulmonary infections were collected from Peking University People's Hospital according to standard procedures. A total of 20 samples collected between February 2016 and October 2017 were investigated in this study. Of these, 12 samples were culture negative, and eight samples were both smear and culture positive. The lung biopsies were separately sent to clinical microbiology and histopathology laboratories within 2 h for analyses. The histopathology laboratory used standard methods for processing clinical samples. In the clinical microbiology laboratory, tissues

were homogenized in 2 mL brain heart infusion broth in a glass grinder, and used for smear and culture. Gram stain, KOH test, and Ziehl-Neelsen stain were used to identify bacteria, fungi and *Mycobacterium tuberculosis* complex (MTBC) by smear microscopy. Homogenized samples were inoculated onto blood agar, eosin methylene blue agar and chocolate agar (Oxoid, Basingstoke, Hampshire, UK) at 35°C to isolate bacteria for up to 5 days. Sabouraud dextrose agar supplemented with and without chloramphenicol (Oxoid, Basingstoke, Hampshire, UK) were used to isolate fungi at 28 and 35°C for up to 5 days, respectively. All cultured microorganisms were identified using the Vitek 2 automated system (bioMérieux, Marcy-l'Etoile, France). Filamentous fungi were identified according to colony morphology and smear results. The remaining or left over tissue homogenates were stripped of patient identification details and stored at -70°C for mNGS.

Isolation of Genomic DNA, Library Preparation, and mNGS

DNA of samples were extracted directly from the tissue homogenates with a TIANamp Micro DNA Kit (TIANGEN BIOTECH). The extracted DNA was fragmented ultrasonically to yield 200–500 bp fragments. After fragmentation by sonication or fragmentase, the DNA fragments were underwent end-repairing, phosphorylation and A-tailing reactions. BGISEQ-500 platform-specific adaptors were ligated to the A-tailed fragments, and the ligated fragments were purified, and then amplified using PCR. Finally, circularization was performed to generate single stranded DNA circles. After quantitation and qualification, the libraries were sequenced. BGI performed the DNA nanoball preparation and whole genome sequencing using the circular single stranded libraries as a template for rolling circle amplification to form DNA nanoballs. The DNA nanoballs were loaded onto a sequencing flow cell and then processed for 50 bp single end sequencing on the BGISEQ-500 platform (Fang et al., 2017). Samples were extracted in batches, with a negative control of whole blood sample from healthy donors prepared alongside each batch using this same protocol.

Bioinformatics Analyses

High quality sequencing data were generated by removing low quality reads, adapter contamination, and duplicated reads, as well as those shorter than 35 bp. Human sequence data were identified by mapping to a human reference (hg19) using Burrows-Wheeler Aligner software (Li and Durbin, 2009) and excluded. The nonhuman sequence reads from each sample were deposited at Genome Sequence Archive of Beijing Institute of Genomics, Chinese Academy of Sciences (gsa.big.ac.cn) under accession number PRJCA000880. The remaining sequence data were aligned to the current bacterial, virus, fungal, and protozoan databases (NCBI; ftp://ftp.ncbi.nlm.nih.gov/genomes). The database used for this study contained 1,428 bacterial species, 1,130 viral species, 73 fungal species, 48 parasites, four species of the MTBC (*M. tuberculosis*, *Mycobacterium canettii*, *Mycobacterium africanum*, and *Mycobacterium bovis*), and 40 mycoplasma/chlamydia related to human diseases. Unique reads were defined as reads

whose alignment length was higher than 90%, identity with reference sequence higher than 95%, and ratio of suboptimal to optimal alignment score lower than 0.8. The infectious bacteria or fungi were determined if it met any of the following thresholds: (i) >30% relative abundance at the genus level in bacteria or fungi; (ii) culture and/or histopathological examination positive and at least 50 unique reads from a single species of bacteria or fungi; (iii) at least one unique read from MTBC.

Statistical Analyses

In accordance with the extracted data, 2×2 contingency tables were derived to determine sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV). All statistics have reported as absolute values with their 95% confidence interval (95% CI). Sensitivity and specificity were calculated on the basis of the formulas TP (true positive)/TP + FN (false negative) and TN (true negative)/TN + FP (false positive), respectively. PPV is expressed by the TP/TP+FP ratio, while NPV from the TN/TN+FN.

RESULTS

Patient Demographics and Basic Sequencing Information

Of the 20 CT-guided puncture lung biopsy tissues processed using mNGS analysis, eight samples were culture positive, and fungal components or acid-fast bacteria were identified in 13 samples by histopathological examination. Potential pathogens of pulmonary infections in 15 patients were identified by mNGS (Table 1). The sequenced biopsies generated between 0.7 and 69 million reads per sample (Supplemental Table S1). The mNGS provide a wide range of microbial profiles, which were difficult to interpret. The microbial species were ranked by relative abundance (Figure 1) and the threshold was determined to identify the presence of true infectious pathogens. *Propionibacterium acnes*, *Micrococcus luteus*, *Malassezia globosa*, *Lactococcus lactis*, and *Saccharomyces*, which were known normal flora of the skin or respiratory tract, were not interpreted as pathogens (Aas et al., 2005; Byrd et al., 2018). Optimal thresholds for determining if samples contained low-level contamination or true infection were determined by numerical optimization (Figure 2). We determined the 30% relative abundance as the final threshold that maximized the sensitivity (40.0%, 95% CI: 17.5–67.1%) and specificity (100.0%, 95% CI: 67.9–100%) of mNGS for fungal infection. For culture and/or histopathology positive samples, 50 unique reads threshold from a single species of fungi was selected to maximize the sensitivity (73.3%, 95% CI: 44.8–91.1%) and reduce missed diagnosis. We determined the same thresholds for the relative abundance and unique number of bacteria required to identify the infectious bacteria according to fungi thresholds, and confirmed our findings in two bacterial culture positive samples. Meanwhile, more fastidious and anaerobic bacteria were identified by mNGS than the conventional culture method (Table 2).

MTBC include *M. tuberculosis*, *M. canettii*, *M. africanum*, and *M. bovis* in the mNGS database. The genome sequence of the four *Mycobacterium* showed high genomic similarity,

so there are few species specific unique reads. For MTBC infection, we suggested focus on matching the number of unique reads in MTBC, rather than unique reads in species. Since the clinical MTBC nucleic acid testing result was positive when only one unique read of MTBC detected by mNGS, we determined the threshold of MTBC infection as one unique read. The number of unique reads of the identified pathogens by mNGS ranged from 1 to 1,425,089. The coverage of identified pathogens ranged from 0.0033 to 84% with a depth value of 1–75, respectively.

Identification of Pathogen Species in Culture Negative Samples by mNGS

Of 12 samples which had negative culture results, mNGS analysis produced negative or nonspecific findings in five samples (P6~P8, P11, and P12) and positive pathogen identification in seven samples (P1~P5, P9, and P10; Table 1). Two samples (P4 and P5) were positive for acid-fast bacteria identification by smearing, and mNGS data confirmed laboratory-based diagnostic testing results. In P5, mNGS analysis identified the MTBC as the most abundant bacteria with 65.71% relative abundance in bacteria and 1,627 unique reads at the genus level. In P4, the number of unique reads of MTBC was 205. GeneXpert MTBC testing was consistent with mNGS result in lung biopsy tissues. Quantitative real-time PCR assay also confirmed positive amplification for MTBC-specific genomic DNA in bronchoalveolar lavage of these patient. Among the three samples which were positive for fungal hypha identification by smearing, fungal species were identified in two samples (P1 and P3) by mNGS (Table 2). Two smear negative samples (P9 and P10) were also positive for fungi by mNGS. Among these samples, *Mucor racemosus* conformed to the threshold and was considered as a possible infectious pathogen in P3 and P9. *Rhizopus microspores* (P1) and *Rhizopus oryzae* (P10), which were consistent with histopathological descriptions, were also interpreted as infectious pathogens. Both *Rhizopus* and *Mucor* belong to *Mucoraceae*, which are difficult to test in a routine microbiology laboratory. In sample P12 from a lung cancer patient, analyses at the species level detected *Enterococcus faecalis* by mNGS, which was not detected by culture. However, *E. faecalis* was not interpreted as pathogen because it was uncommon in pulmonary infection of immunocompetent patient.

Conflicting results between mNGS and conventional smear method were identified in two samples. Other laboratory-based diagnostic testing results were listed in Table 3. In P2, fungi identified by smearing were not detected by mNGS at the predetermined thresholds. A fraction of bacterial reads were assigned to *Haemophilus influenzae* and MTBC by mNGS. Several acid-fast bacteria were identified by histopathological examination in P2, which was consistent with mNGS result. MTBC positive result was also confirmed by quantitative real-time PCR. In P6, no dominant abundant bacteria species was identified by mNGS, which contained Gram negative bacilli according to the smear result. *Malassezia globosa* was dominant over other fungal species in this sample, but was

TABLE 1 | Pulmonary disorders patients for metagenomic next-generation sequencing and conventional laboratory-based diagnostic testing.

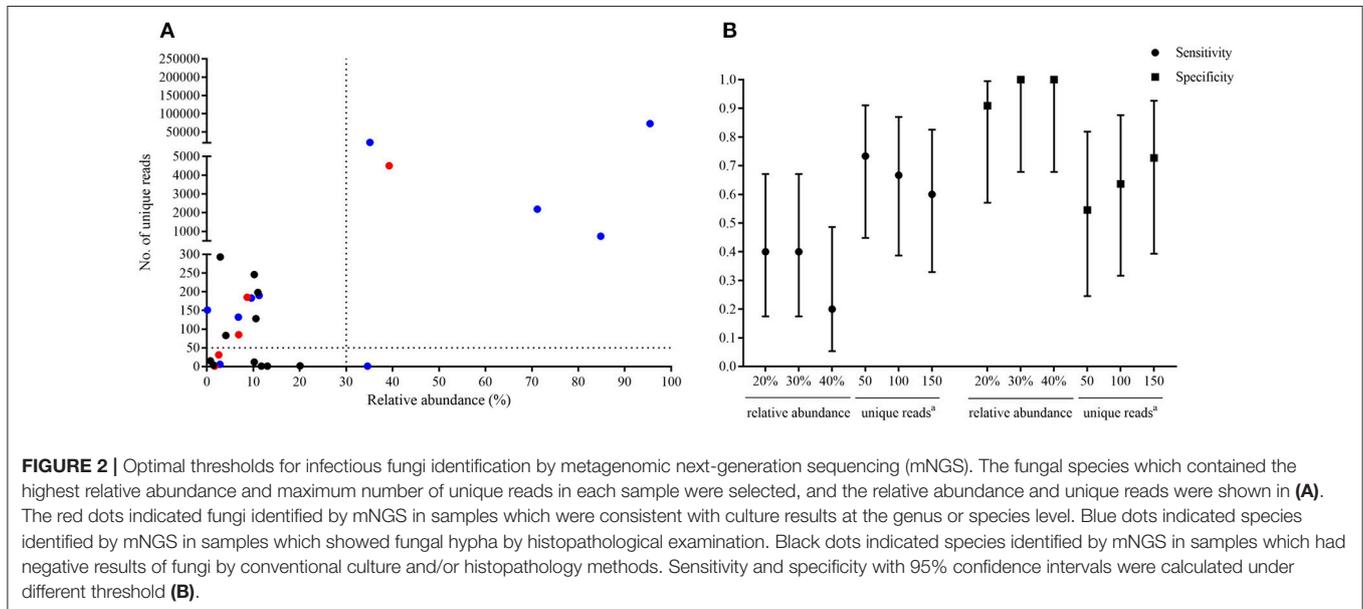
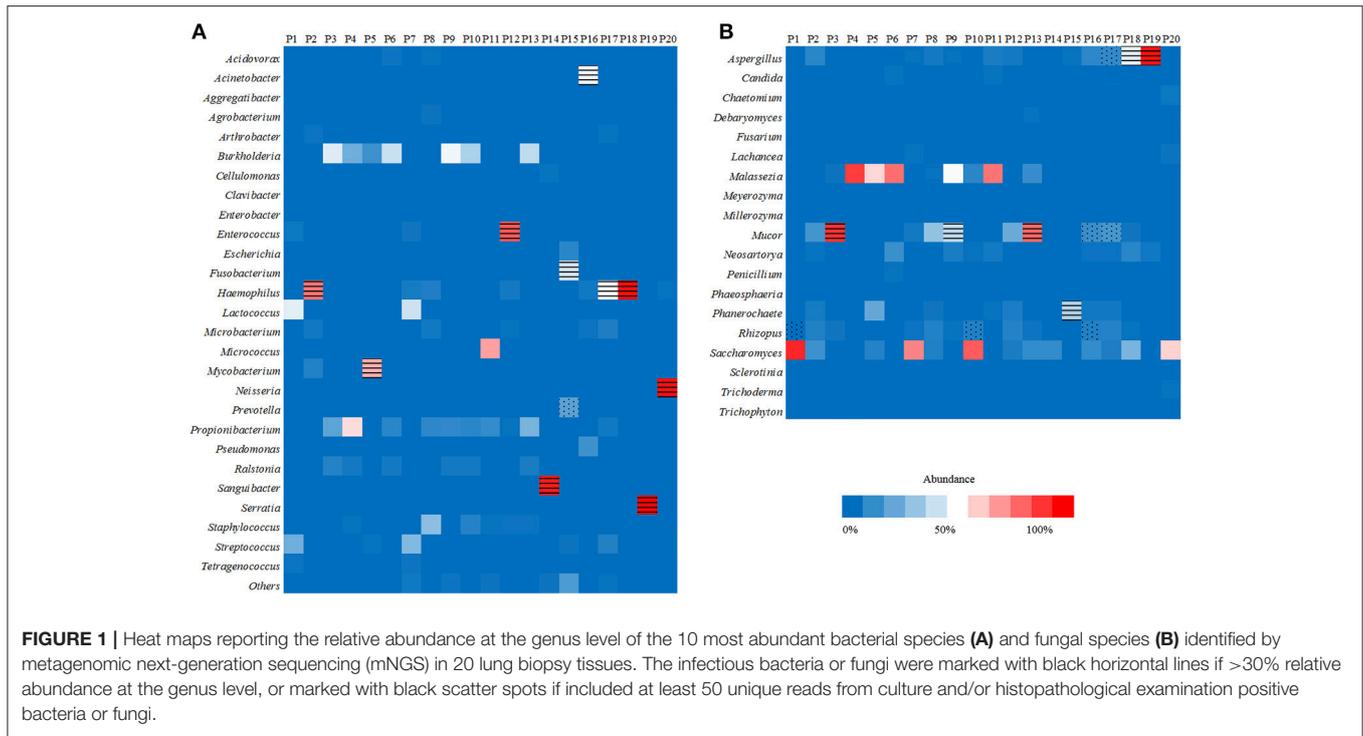
Patient ID	Sex	Age/yr	Underlying disease	Pulmonary disorders	Immunocompromised	Smear results	Culture results	Pathology results	Clinical nucleic acid testing results	mNGS based diagnosis
P1	Male	26	AML; allo-HSCT	Fungal infections	Yes	Fungal nonseptate hypha (90 degrees)	Negative	Fungal hypha (probable <i>Zygomycetes</i>)	ND	Fungal infection
P2	Female	43	Uterine leiomyoma (postoperative)	Pulmonary occupying lesion	No	Fungal hypha	Negative	Several acid-fast bacteria	MTBC positive by quantitative real-time PCR	MTBC infection, bacterial infection
P3	Female	54	ALL	Pulmonary infection	Yes	Fungal nonseptate hypha (90 degrees)	Negative	Fungal hypha	ND	Fungal infection
P4	Male	51	Type 2 diabetes mellitus	Pulmonary tuberculosis	No	Acid-fast bacteria	Negative	Single suspicious acid-fast bacterium	MTBC positive by GeneXpert	MTBC infection
P5	Male	60	Type 2 diabetes mellitus; excision of thyroid adenoma	Pulmonary tuberculosis	No	Acid-fast bacteria	Negative	Acid-fast bacteria	ND	MTBC infection
P6	Female	49	None	Lung cancer (adenocarcinoma)	No	Gram-negative bacilli	Negative	Adenocarcinoma	MTBC negative by GeneXpert	Noninfectious
P7	Female	61	None	Pulmonary inflammatory lesions; Pulmonary occupying lesion	No	Negative	Negative	Chronic nonspecific inflammatory	ND	Noninfectious
P8	Female	40	None	Pulmonary occupying lesion, pulmonary infection	No	Negative	Negative	Scattered lymphocyte infiltration	ND	Noninfectious
P9	Female	24	ALL; allo-HSCT	Pneumonia (coinfection, fungi confirmed)	Yes	Negative	Negative	Fungal hypha	ND	Fungal infection
P10	Male	49	AML; allo-HSCT	Fungal infections	Yes	Negative	Negative	Fungal hypha	ND	Fungal infection
P11	Male	58	Chronic myelogenous leukemia; COPD	Lung squamous cell carcinoma	Yes	Negative	Negative	Probable squamous cell carcinoma	ND	Noninfectious
P12	Male	66	COPD	Lung cancer	No	Negative	Negative	Probable a malignant tumor with epithelioid differentiation	MTBC negative by GeneXpert	Noninfectious

(Continued)

TABLE 1 | Continued

Patient ID	Sex	Age/yr	Underlying disease	Pulmonary disorders	Immunocompromised	Smear results	Culture results	Pathology results	Clinical nucleic acid testing results	mNGS based diagnosis
P13	Female	39	AML; allo-HSCT	Fungal infections	Yes	Negative	<i>Rhizopus</i>	Fungal hypha	MTBC negative by quantitative real-time PCR	Fungal infection
P14	Female	43	Chronic hepatitis B virus hepatitis	Pulmonary tuberculosis (high probability)	No	Negative	<i>Aspergillus fumigatus</i>	Scattered lymphocyte infiltration; granuloma formation	MTBC positive by GeneXpert	MTBC infection
P15	Male	54	Hypertension	Pulmonary inflammatory lesions	No	Fungal hypha	<i>Aspergillus fumigatus</i> , <i>Prevotella intermedia</i>	Chronic nonspecific suppurative inflammation	ND	Bacterial infection
P16	Male	29	AML; allo-HSCT	Lung infection (fungi, bacteria)	Yes	Fungal hypha	<i>Rhizopus</i>	Fungal components (probable <i>Zygomycetes</i>)	ND	Fungal infection, bacterial infection
P17	Female	73	Coronary atherosclerosis	Fungal infections	No	Fungal hypha	<i>Aspergillus fumigatus</i>	Fungal components	ND	Fungal infection, bacterial infection
P18	Female	39	None	Fungal infection	No	Fungal hypha	<i>Aspergillus fumigatus</i> , <i>Staphylococcus haemolyticus</i>	Fungal components	ND	Fungal infection, bacterial infection
P19	Male	46	None	Fungal infections	No	Gram-negative bacilli	<i>Serratia marcescens</i>	Fungal hypha	ND	Fungal infection, bacterial infection
P20	Male	6	ALL; post-right lower lung resection	Pulmonary infection	Yes	Fungal septate hypha	<i>Penicillium</i>	Fungal hypha (probable filamentous fungi except for <i>Aspergillus</i> and <i>Zygomycetes</i>)	MTBC negative by quantitative real-time PCR	Bacterial infection

AML, Acute myeloid leukemia; ALL, Acute lymphocytic leukemia; Allo-HSCT, allogeneic hematopoietic stem cell transplantation; COPD, chronic obstructive pulmonary disease; MTBC, *Mycobacterium tuberculosis* complex; PCR, polymerase chain reaction; ND, no data.



not interpreted as a pathogen because it was known to normal flora of human skin (Byrd et al., 2018). We speculated P6 was noninfectious, which agreed with the clinical diagnosis (lung cancer) and the levels of serological infection markers procalcitonin (PCT) and C-reactive protein (CRP). The possible reasons for not detecting these pathogens by mNGS might be explained by predetermined thresholds or ambiguous smear results.

Identification of Pathogen Species in Culture Positive Samples by mNGS

The results from the mNGS analysis were consistent with three fungi culture results and two bacteria culture results at the species level in eight samples (P13–P20). Moreover, the mNGS identified more infectious pathogens than that by culture method (Table 2). The possible reasons for not detecting these pathogens in culture were that: *Fusobacterium nucleatum* is an obligate anaerobic

TABLE 2 | Metagenomic next-generation sequencing results of 20 lung biopsy samples.

Patient ID	Culture results	Metagenomic results				
		Species	No. of unique reads	Coverage, %	Depth	mNGS based diagnosis
P1	Negative	<i>Rhizopus microsporus</i> ^a	151	0.0425	1.31	Fungal infection
P2	Negative	<i>Mycobacterium canettii</i> (MTBC)	1	0.038	1	MTBC infection, bacterial infection
	Negative	<i>Haemophilus influenzae</i>	1,500	3.1	1	
P3	Negative	<i>Mucor racemosus</i>	743	0.21	1	Fungal infection
P4	Negative	<i>Mycobacterium canettii</i> (MTBC)	5	0.021	1	MTBC infection
P5	Negative	<i>Mycobacterium tuberculosis</i> (MTBC)	5	0.9	1	MTBC infection
P6	Negative	Negative	/	/	/	Noninfectious
P7	Negative	Negative	/	/	/	Noninfectious
P8	Negative	Negative	/	/	/	Noninfectious
P9	Negative	<i>Mucor racemosus</i>	21,548	0.14	1	Fungal infection
P10	Negative	<i>Rhizopus oryzae</i> ^a	132	0.073	1	Fungal infection
P11	Negative	Negative	/	/	/	Noninfectious
P12	Negative	<i>Enterococcus faecalis</i> ^b	453	3.5	1	Noninfectious
P13		<i>Mucor racemosus</i>	2,187	0.27	1	Fungal infection
	<i>Rhizopus</i>	<i>Rhizopus oryzae</i> ^{ad}	31	0.011	1	
P14	<i>Aspergillus fumigatus</i>	<i>Sanguibacter keddiei</i> ^b	1,425,089	50	75	MTBC infection
		<i>Mycobacterium tuberculosis</i> (MTBC)	0 (23 from MTBC)	0.027	1.5	
P15		<i>Phanerochaete chrysosporium</i> ^b	1	0.011	1.3	Bacterial infection
	<i>Aspergillus fumigatus</i>	<i>Aspergillus fumigatus</i> ^d	1	0.00041	1.7	
		<i>Fusobacterium nucleatum</i> ^c	69	0.2478	1	
	<i>Prevotella intermedia</i>	<i>Prevotella intermedia</i> ^a	57	0.1153	1	
P16	<i>Rhizopus</i>	<i>Rhizopus microsporus</i> ^a	85	0.0106	9	Fungal infection, bacterial infection
		<i>Mucor indicus</i> ^a	183	0.0036	36.35	
		<i>Acinetobacter lwoffii</i>	660	1.6	1.1	
P17	<i>Aspergillus fumigatus</i>	<i>Aspergillus fumigatus</i> ^a	185	0.0343	2.91	Fungal infection, bacterial infection
		<i>Mucor indicus</i> ^a	190	0.0033	39.42	
		<i>Haemophilus influenzae</i>	242	0.66	1	
P18	<i>Aspergillus fumigatus</i> ,	<i>Aspergillus fumigatus</i>	4,508	1.1149	1.01	Fungal infection, bacterial infection
	<i>Staphylococcus haemolyticus</i>	<i>Haemophilus influenzae</i>	378,078	84	12	
P19	<i>Serratia marcescens</i>	<i>Serratia marcescens</i>	227,748	82	8.1	Fungal infection, bacterial infection
		<i>Aspergillus fumigatus</i>	73,043	32	1.2	
P20	<i>Penicillium</i>	<i>Chaetomium globosum</i> ^{abd}	6	0.0036	1.1	Bacterial infection
		<i>Neisseria meningitidis</i>	4,247	15	3.9	

^aInferred pathogen combined with culture and/or histopathological examination results.

^bRare pathogen in pulmonary infection, which was not interpreted as pathogen.

^cObligate anaerobe.

^dBelow the thresholds.

bacterium, *H. influenzae* and *Neisseria meningitidis* are fastidious organisms, *Aspergillus fumigatus* requires longer culture period. The grind processing may affect the isolation of *Zygomycetes* (such as *Rhizopus* and *Mucor*). Moreover, pathogens isolation may not be possible especially if antibiotic treatment has already been initiated.

In P13, P15, P16, and P17, the fungi identified by culture were not the most abundant species in mNGS results. In P13, the fungal culture indicated *Rhizopus*, but *M. racemosus* had

the highest abundance in fungi by mNGS (71.16%). *Rhizopus* was also identified in P13 with the low relative abundance (2.56%), which ranked fifth in species of fungi. The mNGS data showed the number of unique reads of *R. oryzae* was 31. *Mucor indicus* was identified in P16 and P17 as the most abundant species in fungi, but relative abundance were lower than 15%. In P15, *Phanerochaete chrysosporium* is a species of wood-rotting fungus and generally regarded as unable to induce diseases in human.

TABLE 3 | The detail information of samples with conflicting results.

Patient ID	Smear results	Culture results	Histopathology results	Potential pathogens by mNGS	Other laboratory-based diagnostic testing results	Clinical diagnosis of pulmonary disorders
P2	Fungal hypha	Negative	Several acid-fast bacteria	MTBC <i>Haemophilus influenzae</i>	MTBC positive by quantitative real-time PCR WBC $14.80 \times 10^9/L$ (↑)	Pulmonary occupying lesion
P6	Gram-negative bacilli	Negative	Adenocarcinoma	Negative	MTBC negative by GeneXpert CRP 1.5mg/L PCT <0.02ng/ml	Lung cancer (adenocarcinoma)
P14	Negative	<i>Aspergillus fumigatus</i>	Scattered lymphocyte infiltration; granuloma formation	MTBC	MTBC positive by GeneXpert	Pulmonary tuberculosis (high probability)
P20	Fungal septatehypha	<i>Penicillium</i>	Fungal hypha (probable filamentous fungi except for <i>Aspergillus</i> and <i>Zygomycetes</i>)	<i>Neisseria meningitidis</i>	WBC $11.06 \times 10^9/L$ (↑) CRP 16.62 mg/L (↑) MTBC negative by quantitative real-time PCR	Pulmonary infection

MTBC, *Mycobacterium tuberculosis* complex; PCR, polymerase chain reaction; WBC, white blood cell count; CRP, C-reactive protein; PCT, procalcitonin. Up arrow (↑) indicated that the result was higher than normal reference range.

Conflicting data from mNGS and culture techniques were obtained in two samples (P14 and P20). Other laboratory-based diagnostic testing results were listed in **Table 3**. *Sanguibacter keddiei* was identified in P14 with 94.38% relative abundance in all reads assigned to bacteria by mNGS. Because pulmonary infection due to *Sanguibacter* is not common (Jones, 2010), *S. keddiei* was not interpreted as pathogen. Furthermore, mNGS identified the MTBC in P14, which was not detected by conventional smear methods. GeneXpert MTBC testing was consistent with mNGS result in P14. The possible reasons for not detecting these pathogens by conventional laboratory-based diagnostic methods could be explained by the fact that direct smear tests for MTBC have a low sensitivity. The P20 was a 6-year-old child, who had been admitted due to acute lymphocytic leukemia with central nervous system abnormalities. *Penicillium*, detected in fungal culture, was not specifically detected by mNGS. The mNGS identified *N. meningitidis* and *Saccharomyces cerevisiae* among the top most abundant species in bacteria and fungi respectively. *N. meningitidis* is a fastidious organism which lives in the nasopharynx, and *S. cerevisiae* is a known reagent contaminant. Although pulmonary infection due to *Neisseria* is uncommon (Hirai et al., 2016), physicians should consider *N. meningitidis* as possible pathogens in an immunocompromised patient. Histopathological examination indicated filamentous fungi except for *Aspergillus* and *Zygomycetes*. Further analysis of mNGS data showed *Chaetomium globosum* was ranked second by unique reads in species of fungi. However, *C. globosum* was not interpreted as a pathogen since it was only 6 unique reads.

Comparison of Conventional Laboratory-Based Diagnostic Methods and mNGS

Pulmonary disorder patients were classified into three groups according to the clinical diagnosis (**Figure 3**). In 12 patients diagnosed with pulmonary infections, two fungi positive and two MTBC positive samples detected by the conventional smear

method were identified correctly by mNGS. Among six fungi positive samples detected by smearing, five were called fungi positive by mNGS. It's worth noting that three samples with a negative smear result were indicated as fungi positive by mNGS. Bacteria and fungi coinfections were identified in four cases, in which only one etiology was detected by smearing. Three samples with non-infectious etiologies were confirmed by mNGS. Importantly, one sample positive for bacteria positive by smearing was not detected by mNGS, which agreed with clinical diagnosis. The mNGS confirmed two of five inconclusive etiology cases, failing to detect fungi identified by smearing in two samples. Moreover, previously unrecognized MTBC infections and/or bacterial infections were found in two cases by mNGS.

The comparison of mNGS results with conventional microbiology methods (smear and culture) and histopathological method were shown in **Table 4**. The sensitivity and specificity of mNGS compared with smear method were 50.0% (95% CI: 2.7–97.3%) and 66.7% (95% CI: 41.2–85.6%) for bacteria, 62.5% (95% CI: 25.9–89.8%) and 66.7% (95% CI: 35.4–88.7%) for fungi, 100.0% (95% CI: 19.8–100.0%) and 88.9% (95% CI: 63.9–98.1%) for MTBC, respectively. The comparison between mNGS and culture method resulted that the sensitivity and specificity were 100.0% (95% CI: 31.0–100.0%) and 76.5% (95% CI: 49.8–92.2%) for bacteria, 57.1% (95% CI: 20.2–88.2%) and 61.5% (95% CI: 32.2–84.9%) for fungi. The PPV (42.9% for bacteria, 44.4% for fungi) was much lower than NPV (100% for bacteria, 72.7% for fungi) in mNGS vs. culture method. The mNGS showed the highest specificity (100.0 and 94.1%) and PPV (100.0 and 75.0%) in the evaluation of fungi and MTBC, respectively, when compared with histopathology method.

DISCUSSION

We reported a retrospective study of the application of mNGS in the diagnosis of infectious pathogens in lung biopsy tissues. The mNGS offered the advantage of a less biased pathogen detection

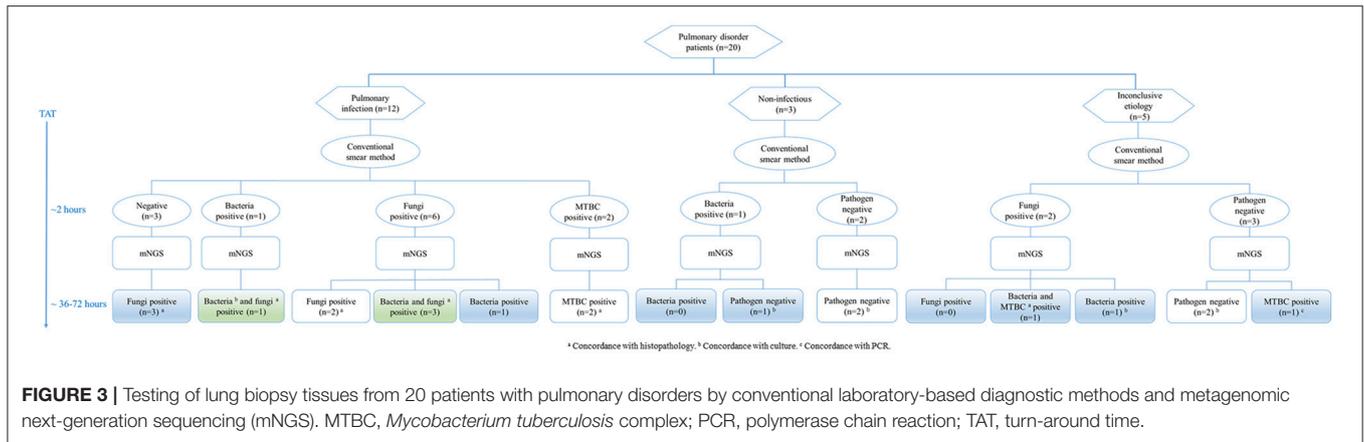


FIGURE 3 | Testing of lung biopsy tissues from 20 patients with pulmonary disorders by conventional laboratory-based diagnostic methods and metagenomic next-generation sequencing (mNGS). MTBC, *Mycobacterium tuberculosis* complex; PCR, polymerase chain reaction; TAT, turn-around time.

TABLE 4 | Diagnostic performance of metagenomic next-generation sequencing (mNGS) when compared with conventional laboratory-based diagnostic methods in the detection of bacteria, fungi, and *Mycobacterium tuberculosis* complex (MTBC).

Pathogen	Conventional laboratory-based diagnostic method	mNGS			
		Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Bacteria	Smear	50.0 (2.7–97.3)	66.7 (41.2–85.6)	14.3 (0.8–58.0)	92.3 (62.1–99.6)
	Culture	100.0 (31.0–100.0)	76.5 (49.8–92.2)	42.9 (11.8–79.8)	100.0 (71.7–100)
Fungi	Smear	62.5 (25.9–89.8)	66.7 (35.4–88.7)	55.6 (22.7–84.7)	72.7 (39.3–92.7)
	Culture	57.1 (20.2–88.2)	61.5 (32.2–84.9)	44.4 (15.3–77.4)	72.7 (39.3–92.7)
	Histopathology	90.0 (54.1–99.5)	100.0 (65.6–100.0)	100.0 (62.9–100.0)	90.9 (57.1–99.5)
MTBC	Smear	100.0 (19.8–100.0)	88.9 (63.9–98.1)	50.0 (9.2–90.8)	100.0 (76.0–100.0)
	Histopathology	100.0 (31.0–100.0)	94.1 (69.2–99.7)	75.0 (21.9–98.7)	100.0 (75.9–100.0)

PPV, positive predictive value; NPV, negative predictive value; CI: Confident Intervals.

methodology through direct sequencing of the sample's extracted DNA. In the current study, mNGS successfully identified the infectious pathogens in 15 out of 20 patients. It covered a wider range of fastidious and anaerobic pathogens than conventional culture method. In P15, two obligate anaerobic bacteria were identified. *F. nucleatum* and *P. intermedia* are often isolated from periodontal lesions associated with various forms of periodontal disease (Ximénez-Fyvie et al., 2000; Baek et al., 2018). Several studies reported the pathogenic potential of *P. intermedia* in the respiratory tract and demonstrated that extracellular toxins of *P. intermedia* are cytotoxic for human alveolar type II cells and neutrophils (Ulrich et al., 2010). The presence of *P. intermedia* in the oral cavity or lower respiratory tract may be a risk factor for severe pneumococcal pneumonia (Nagaoka et al., 2014).

It is noteworthy that our research indicated that the application of mNGS improved the diagnosis of pulmonary invasive fungal infections. In the clinical microbiology laboratory, tissues were homogenized in a glass grinder, and used for smear and culture. The grind processing may affect the isolation of *Zygomycetes* (such as *Rhizopus* and *Mucor*). The mNGS identified more *Zygomycetes* than that by culture method. Four samples with negative culture results were indicated as *Rhizopus* or *Mucor* by mNGS. In another three samples, *Mucor*

was identified by mNGS, which were absent in culture results. The PPV of mNGS relative to histopathological examination of fungi was high at 100.0%, with 90.0% sensitivity and 100.0% specificity in our study. The mNGS combined with smear analyses could be used as a routine diagnostic tool in invasive fungal infections, which can reduce the turn-around time and provide accurate identification of fungi species (Figure 3).

The mNGS could provide a wide range of organism and microbial profiles, which were difficult to interpret. The lung microbiome should be considered when interpreting mNGS results. The microbiome of the lung has been particularly difficult to characterize due to prior assumptions about the community composition of the lung, the diversity of pathogens causing diseases, and sampling concerns (Pragman et al., 2018). Several genera were proposed as possible core genera of the lung microbiome, including *Pseudomonas*, *Streptococcus*, *Prevotella*, *Fusobacterium*, *Haemophilus*, *Veillonella*, and *Porphyromonas* (Morris et al., 2013; Yu et al., 2016). However, samples of most lung microbiome studies were bronchoalveolar lavages rather than lung tissues. The composition of the microbial genera in the lung biopsies detected in our study showed some similarities with other lung microbiome studies, but many of these genera were found in relatively smaller proportions. In our study, we

applied a new data management pipeline for the identification of pathogenic species from mNGS data. The relative abundance of 30% at the genus level and at least fifty unique reads could be used as thresholds to indicate the presence of potential pathogens and background lung microbiome. However, this should not apply to the MTBC, the common pathogen in pulmonary infections (Zumla et al., 2015). MTBC included *M. tuberculosis*, *M. canettii*, *M. africanum*, and *M. bovis* in our mNGS database. The genome sequence of the four *Mycobacterium* species showed high genomic similarity (Gutierrez et al., 2005), so there are few species-specific sequences. In P2 and P4, *M. canettii* contained more unique reads than other species in MTBC, but the number of unique reads to different species was not more than 5. Among the MTBC, *M. canettii* is a peculiar member specifically diagnosed in dozens of tuberculosis patients with reported contacts to the Horn of Africa (Aboubaker et al., 2015). *M. canettii* differs from the other members of the MTBC by processing a larger 4.48 ± 0.05 Mb mosaic genome and producing cordless and smooth-looking mycobacteria (Gutierrez et al., 2005). We speculated that the identification of specific species of MTBC may not be accurate, due to the high genomic similarity between different species of MTBC. Moreover, not all sequences in difference among the reference genomes can be used for species classification in MTBC. To avoid misunderstanding, we recommended reporting the results of MTBC, rather than specific species of MTBC. Based on our research, whenever the MTBC was detected by mNGS, pulmonary tuberculosis should be considered. The mNGS of lung biopsies yielded reads from bacteria from the oropharyngeal or skin flora, as well as viruses and yeasts that were not considered causes of pulmonary infections in immunocompetent patients. However, in immunocompromised patients, interpretation of the normal flora or the presence of environmental microbes should be guided by clinical manifestations and related laboratory examination results. Hence, from our mNGS data, we propose to compile an extensive list of pathogens for which the matching reads should be kept and which can be considered as potential pathogens in lung infections in future mNGS analyses. Meanwhile, combined with the results of conventional laboratory-based diagnostic methods, the interpretation of mNGS data could be improved.

This study had limitations, such as the small sample size, the depth of mNGS and the high amount of human DNA fragments in lung biopsy tissues. The lung biopsy tissues were collected by invasive procedures, so the sample size was less than other respiratory specimens. Identification of pathogens by

mNGS in lung biopsy tissues was challenging and meaningful. The samples selected in our study included patients with different infectious pathogens and noninfectious etiology. Further, we will increase the sample size for verification the optimal thresholds for pathogens identification. Meanwhile, the clinical prospective study with patient follow-up and treatment would have to be conducted to ensure that mNGS can accurately characterize the lung microbiome of samples and highlight the presence of lung pathogens. In addition, the sequenced biopsies generated varied largely, from 0.7 to 69 million reads per sample (Supplemental Table S1). The P9 sample has only 0.7 million reads but contains high proportions of microbes, which means that the microbe-human ratio was relative high in the original sample. We believed that the low sequencing depth can reflect the real microbes' constitutions at this condition. All the samples were planned to have a depth about 20 million reads, but it was affect by a lot of factors, such as the DNA concentration of each sample, the accuracy of pooling operation, and the cyclization efficiency of the library. If the DNA concentration was too low, a larger volume was used for pooling, but it may still affect the reads counts. Standardized procedures need to be developed in the further study to eliminate or reduce biases between samples.

Although there are some limitations, our study firstly indicated that mNGS could offer an improved detection of pulmonary infectious pathogens (or absence) in lung biopsy tissues, with potential benefits in speed and sensitivity.

AUTHOR CONTRIBUTIONS

HW conceived and designed the study. HL and HG analyzed the data. HM collected the related clinical information. HM, QW, SL, and HC conducted clinical work associated with the study. YL provided technical support. HL wrote the draft, and HW revised it. All authors approved the final version.

FUNDING

This work was partly supported by the National Natural Science Foundation of China (81625014).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2018.00205/full#supplementary-material>

REFERENCES

- Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., and Dewhirst, F. E. (2005). Defining the normal bacterial flora of the oral cavity. *J. Clin. Microbiol.* 43, 5721–5732. doi: 10.1128/JCM.43.11.5721-5732.2005
- Aboubaker, O. D., Bouzid, F., Canaan, S., and Drancourt, M. (2015). Smooth tubercle bacilli: neglected opportunistic tropical pathogens. *Front. Public Health* 3:283. doi: 10.3389/fpubh.2015.00283
- Baek, K., Ji, S., and Choi, Y. (2018). Complex intratissue microbiota forms biofilms in periodontal lesions. *J. Dent. Res.* 97, 192–200. doi: 10.1177/0022034517732754
- Byrd, A. L., Belkaid, Y., and Segre, J. A. (2018). The human skin microbiome. *Nat. Rev. Microbiol.* 16, 143–155. doi: 10.1038/nrmicro.2017.157
- De La Cruz, O., and Silveira, F. P. (2017). Respiratory fungal infections in solid organ and hematopoietic stem cell transplantation. *Clin. Chest Med.* 38, 727–739. doi: 10.1016/j.ccm.2017.07.013

- Fang, C., Zhong, H., Lin, Y., Chen, B., Han, M., Ren, H., et al. (2017). Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *Gigascience* 7, 1–8. doi: 10.1093/gigascience/gix133
- Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., and Reimer, A. (2017). Metagenomics: the next culture-independent game changer. *Front. Microbiol.* 8:1069. doi: 10.3389/fmicb.2017.01069
- Fukui, Y., Aoki, K., Okuma, S., Sato, T., Ishii, Y., and Tateda, K. (2015). Metagenomic analysis for detecting pathogens in culture-negative infective endocarditis. *J. Infect. Chemother.* 21, 882–884. doi: 10.1016/j.jiac.2015.08.007
- Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., et al. (2005). Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog.* 1:e5. doi: 10.1371/journal.ppat.0010005
- Hilton, S. K., Castro-Nallar, E., Perez-Losada, M., Toma, I., McCaffrey, T. A., Hoffman, E. P., et al. (2016). Metataxonomic and metagenomic approaches vs. culture-based techniques for clinical pathology. *Front. Microbiol.* 7:484. doi: 10.3389/fmicb.2016.00484
- Hirai, J., Kinjo, T., Tome, T., Hagihara, M., Sakashashi, D., Nakamura, H., et al. (2016). Meningococcal pneumonia in Japan: a case report and literature review. *J. Infect. Chemother.* 22, 833–836. doi: 10.1016/j.jiac.2016.07.014
- Jones, R. N. (2010). Microbial etiologies of hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia. *Clin. Infect. Dis.* 51(Suppl. 1), S81–S87. doi: 10.1086/653053
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Magill, S. S., Edwards, J. R., Bamberg, W., Beldavs, Z. G., Dumyati, G., Kainer, M. A., et al. (2014). Multistate point-prevalence survey of health care-associated infections. *N. Engl. J. Med.* 370, 1198–1208. doi: 10.1056/NEJMoa1306801
- Morris, A., Beck, J. M., Schloss, P. D., Campbell, T. B., Crothers, K., Curtis, J. L., et al. (2013). Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am. J. Respir. Crit. Care Med.* 187, 1067–1075. doi: 10.1164/rccm.201210-1913OC
- Naccache, S. N., Peggs, K. S., Mattes, F. M., Phadke, R., Garson, J. A., Grant, P., et al. (2015). Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clin. Infect. Dis.* 60, 919–923. doi: 10.1093/cid/ciu912
- Nagaoka, K., Yanagihara, K., Morinaga, Y., Nakamura, S., Harada, T., Hasegawa, H., et al. (2014). *Prevotella intermedia* induces severe bacteremic pneumococcal pneumonia in mice with upregulated platelet-activating factor receptor expression. *Infect. Immun.* 82, 587–593. doi: 10.1128/IAI.00943-13
- Pragman, A. A., Lyu, T., Baller, J. A., Gould, T. J., Kelly, R. F., Reilly, C. S., et al. (2018). The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease. *Microbiome* 6:7. doi: 10.1186/s40168-017-0381-4
- Renaud, C., and Campbell, A. P. (2011). Changing epidemiology of respiratory viral infections in hematopoietic cell transplant recipients and solid organ transplant recipients. *Curr. Opin. Infect. Dis.* 24, 333–343. doi: 10.1097/QCO.0b013e3283480440
- Ruppé, E., Baud, D., Schicklin, S., Guigon, G., and Schrenzel, J. (2016). Clinical metagenomics for the management of hospital- and healthcare-acquired pneumonia. *Future Microbiol.* 11, 427–439. doi: 10.2217/fmb.15.144
- Salzberg, S. L., Breitwieser, F. P., Kumar, A., Hao, H., Burger, P., Rodriguez, F. J., et al. (2016). Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system. *Neurol. Neuroimmunol. Neuroinflamm.* 3:e251. doi: 10.1212/NXI.0000000000000251
- Somasekar, S., Lee, D., Rule, J., Naccache, S. N., Stone, M., Busch, M. P., et al. (2017). Viral surveillance in serum samples from patients with acute liver failure by metagenomic next-generation sequencing. *Clin. Infect. Dis.* 65, 1477–1485. doi: 10.1093/cid/cix596
- Ulrich, M., Beer, I., Braitmaier, P., Dierkes, M., Kummer, F., Krismer, B., et al. (2010). Relative contribution of *Prevotella intermedia* and *Pseudomonas aeruginosa* to lung pathology in airways of patients with cystic fibrosis. *Thorax* 65, 978–984. doi: 10.1136/thx.2010.137745
- Ximénez-Fyvie, L. A., Haffajee, A. D., and Socransky, S. S. (2000). Microbial composition of supra- and subgingival plaque in subjects with adult periodontitis. *J. Clin. Periodontol.* 27, 722–732. doi: 10.1034/j.1600-051x.2000.027010722.x
- Yu, G., Gail, M. H., Consonni, D., Carugno, M., Humphrys, M., Pesatori, A. C., et al. (2016). Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome Biol.* 17:163. doi: 10.1186/s13059-016-1021-1
- Zumla, A., George, A., Sharma, V., Herbert, R. H., Oxley, A., and Oliver, M. (2015). The WHO 2014 global tuberculosis report—further to go. *Lancet Glob. Health* 3, e10–e12. doi: 10.1016/S2214-109X(14)70361-4

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

Copyright © 2018 Li, Gao, Meng, Wang, Li, Chen, Li and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.