

Appendix

Yield of metagenomics in suspected central nervous system infections with negative cerebrospinal fluid cultures

Chaowen Deng et al.

Supplementary Methods

1. Procedure of CSF mNGS

Cerebrospinal fluid (CSF) was collected by lumbar puncture. The samples were stored in a sterile container, then preserved and transported in dry ice. The whole process of sequencing and pathogen detection pipeline were carried out in the laboratory of Guangzhou Kingmed Medical Test Center Co.Ltd..

The nucleic acid of samples was extracted by QIAasympohony Circulating NA Kit (Cus.48), together with a negative control (RNase-free water) and a positive control. Next, the nucleic acid libraries were prepared using TruePrep™ Library Prep Kit V2 for Illumina® (Vazyme Biotech Co., Ltd, TD503). The nucleic acid libraries concentration was measured by Qubit. Then the sequencing was carried out by Illumina nextseq 500 system with 75 cycles Reagent Kit. High-quality sequencing data were obtained by filtering out reads of low-quality. Then, reads of human were removed by mapping reads to human reference genome. The remaining data were aligned to the microbial genome database.

The procedure of filtering, mapping and alignment are carried out by the widely used software Sequence-Based Ultra-Rapid Pathogen Identification (SURPI), which is generally considered to be highly accurate. The microbial genome database is MetagenomicX for clinical application. Reads with multiple locus alignments within the same genus were excluded in secondary analysis. Only reads mapped to the genome within the same species were considered.

Sequence reads are defined a normalized sequence number in 200,000 primary sequences detected and derived from an internal control. To estimate the concentration of the target pathogen, exogenous plasmids of known concentrations were utilized as the internal control. The number of normalized sequence reads for the target pathogen can be calculated using the amplification efficiency ratio of the internal control.

$$\text{Sequence reads} = \frac{20 \text{ millions} \times \text{Number of reads only mapped within same taxon}}{\text{Total reads of this sample}}$$

In order to construct the microbial genome database, pathogens and their genomes or assemblies were selected following the Kraken2 criteria for selecting representative assemblies for microorganisms (bacteria, viruses, fungi, protozoa, and other multicellular eukaryotic pathogens) from kraken2 database (<https://benlangmead.github.io/aws-indexes/k2>). Pathogens from Johns Hopkins ABX Guide (https://www.hopkinsguides.com/hopkins/index/Johns_Hopkins_ABX_Guide/Pathogens), Manual of Clinical Microbiology, and clinical case reports or research articles published in current peer-reviewed journals are included in this in-house database. The whole database contains 22934 microorganisms, which covers most of the microbial genomes which have been sequenced. 8704 of the total microorganisms in the first-grade database have integral sequence of whole genome and detailed clinical analysis, which covers most of the known pathogenic bacteria, viruses, fungi and parasites. This first-grade database with high-quality genomes is used as first choice.

2. Quality control of procedure

During the process of sampling, storage, transportation and sequencing, aseptic procedures are strictly followed to ensure that qualified BALF and CSF specimens are collected. In the lab, the specimens were kept in the -80 degree refrigerator to prevent nucleic acid degradation. All detection procedures are finished within 48 hours after receiving specimen ensuring timely guidance for clinical treatment.

Both Nucleic acid extraction and library preparation were conducted in parallel with quality control samples. The results of this whole process with the results analyzed by software of Burrows Wheeler Alignment (BWA). The results of the two pipelines are highly consistent. To eliminate background interference, a minimum threshold of 10 RPM-r (RPM defined as Reads per million, RPM-r defined as $\text{RPM}_{\text{sample}}/\text{RPM}_{\text{no-template-control}}$) is designated for reporting the detection of a microorganism as “detected”.

3. Criteria for a Positive mNGS Result

1. Virus, bacteria (mycobacteria excluded), and parasites: according to Langelier’s study, mNGS identified microbes as confirmed pathogens following the microbes which were identified by the clinical test. Microbes were considered as potential pathogens if literature has reported the score or the pathogenicity were as least tentimes greater than that of any other microbes of the same type identified in the patient.

2. Fungi: mNGS identified microbes (species level) as confirmed pathogens when the coverage rate scored 5 times higher than that of any other fungus, since low absolute abundance appears in contaminated fungal DNA and many samples.

3. Mycobacteria: Mycobacterium tuberculosis (MTB) was identified positive when at least 1 read mapped to species or genus level, since the Mycobacteria cell wall is hard to disrupt the organism to release nucleic acid as well as low possibility for contamination. The efficient methods of extraction are crucial to achieve unbiased sequencing. Non-tuberculous mycobacteria (NTM) were considered as positive when the number of mapped reads (genus or species level) was top 10 in the bacteria list due to the low yield rate and balance of environmental contamination from hospital to laboratory.

4. List of common pathogenic microorganisms detected by mNGS

Bacteria	Virus	Fungi	Parasite	Mycoplasma/Chlamydia/Rickettsia/Spirochaeta
<i>Staphylococcus aureus</i>	Influenza A virus	<i>Epidermophyton floccosum</i>	<i>Plasmodium falciparum</i>	<i>Mycoplasma pneumoniae</i>
<i>Staphylococcus epidermidis</i>	Influenza B virus	<i>Trichophyton gypseum</i>	<i>Plasmodium malariae</i>	<i>Mycoplasma hominis</i>
<i>Staphylococcus haemolyticus</i>	Parainfluenza virus	<i>Trichophyton rubrum</i>	<i>Plasmodium ovale</i>	<i>Mycoplasma genitalium</i>
<i>Staphylococcus hominis</i>	Respiratory syncytial virus	<i>Microsporum ferrugineum</i>	<i>Plasmodium vivax</i>	<i>Ureaplasma Urealyticum</i>
<i>Staphylococcus luden</i>	Measles virus	<i>Microsporum canis</i>	<i>Toxoplasma gondii</i>	<i>Rickettsia prowazekii</i>
<i>Staphylococcus warneri</i>	Mumps virus	<i>Microsporum gypseum</i>	<i>Strongyloides stercoralis</i>	<i>Rickettsia typhi</i>
<i>Staphylococcus saprophyticus</i>	Human metapneumovirus	<i>Malassezia fufur</i>	<i>Gnathostoma siamense</i>	<i>Orientia tsutsugamushi</i>
<i>Staphylococcus capitis</i>	Rubella virus	<i>Malassezia globosa</i>	<i>Baylisascaris procyonis</i>	<i>Chlamydia trachomatis</i>
<i>Micrococcus luteus</i>	Rhinovirus	<i>Sporothrix schenckii</i>	<i>Amebic protozoa</i>	<i>Chlamydia pneumoniae</i>
<i>Streptococcus pneumoniae</i>	Adenovirus	<i>Fonsecaea pedrosoi</i>	<i>Amoeba dysenteriae</i>	<i>Chlamydia psittaci</i>
<i>Streptococcus pyogenes</i>	SARS-Cov-2	<i>Cladosporium</i>	<i>Cyclospora</i>	<i>leptospira interrogans</i>

		<i>carrionii</i>	<i>cayetanensis</i>	
<i>Streptococcus agalactis</i>	SARS-Cov-1	<i>Phialophora verrucosa</i>	<i>Cryptosporidium baileyi</i>	<i>Borrelia recurrentis</i>
<i>Streptococcus mitis</i>	Poliovirus	<i>Exophiala jeanselmei</i>	<i>Isospora belli</i>	<i>Treponema pallidum</i>
<i>Streptococcus viridans</i>	Coxsackievirus	<i>Alternaria alternata</i>	<i>Giardiasis lamblia</i>	<i>Borrelia burgdorferi</i>
<i>Streptococcus oralis</i>	Enterovirus	<i>Histoplasma capsulatum</i>	<i>Taenia solium</i>	<i>Borrelia vincentii</i>
<i>Streptococcus salivarius</i>	Rotavirus	<i>Coccidioides immitis</i>
<i>Streptococcus sanguis</i>	Norovirus	<i>Oidium dermatitidis</i>		
<i>Streptococcus bovis</i>	Astrovirus	<i>Paracoccidioides brasiliensis</i>		
<i>Streptococcus suis</i>	Hepatitis virus A	<i>Penicillium marneffeii</i>		
<i>Enterococcus faecium</i>	Hepatitis virus B	<i>Candida albicans</i>		
<i>Enterococcus faecalis</i>	Hepatitis virus C	<i>Candida tropicalis</i>		
<i>Neisseria meningitidis</i>	Hepatitis virus D	<i>Candida glabrata</i>		
<i>Neisseria gonorrhoeae</i>	Hepatitis virus E	<i>Candida krusei</i>		
<i>Escherichia coli</i>	severe fever with thrombocytopenia syndrome virus	<i>Cryptococcus neoformans</i>		
<i>Shigella dysenteriae</i>	Crimean-Congo hemorrhagic fever virus	<i>Aspergillus fumigatus</i>		
<i>Shigella flexneri</i>	Human herpesvirus-1	<i>Fusarium</i>		

		<i>oxysporum</i>		
<i>Shigella bogdii</i>	Human herpesvirus-2	<i>Mucor circinelloides</i>		
<i>Shigella sonnei</i>	Varicella-zoster virus	<i>Pneumocystis jiroveci</i>		
<i>Salmonella typhi</i>	Epstein-Barr virus		
<i>Salmonella paratyphi</i>	Human cytomegalovirus			
<i>Salmonella schottmuelleri</i>	Human herpesvirus-6			
<i>Salmonella hirschfeldii</i>	Human herpesvirus-7			
<i>Salmonella enteritidis</i>	Human herpesvirus-8			
<i>Klebsiella pneumoniae</i>	Japanese encephalitis virus			
<i>Klebsiella oxytoca</i>	Dengue virus			
<i>Proteus mirabilis</i>	Russian spring-summer encephalitis virus			
<i>Proteus vulgaris</i>	West nile virus			
<i>Enterobacter cloacae</i>	Hantavirus			
<i>Serratia marcescens</i>	Poxvirus			
<i>Morganella morganii</i>	Ebola virus			
<i>Vibrio cholerae</i>	Human immunodeficiency virus			
<i>Vibrio parahaemolyticus</i>	Human T-cell lymphotropic virus			
<i>Vibrio mimicus</i>	Rabies virus			

<i>Vibrio fluvialis</i>	Human papilloma virus			
<i>Vibrio vulnificus</i>	Parvoviruses			
<i>Helicobacter pylori</i>	Zika virus			
<i>Clostridium tetani</i>			
<i>Clostridium perfringens</i>				
<i>Clostridium botulinum</i>				
<i>Clostridium difficile</i>				
<i>Bacteroides fragilis</i>				
<i>Fusobacterium nucleatum</i>				
<i>Veillonella parvula</i>				
<i>Propionibacterium acnes</i>				
<i>Bifidobacterium dentium</i>				
<i>Eubacterium lentum</i>				
<i>Prevotella melaninogenica</i>				
<i>Mycobacterium tuberculosis</i>				
<i>Mycobacterium leprae</i>				

<i>Mycobacterium kansasii</i>				
<i>Mycobacterium avium</i>				
<i>Mycobacterium intracellulare</i>				
<i>Mycobacterium abscessus</i>				
<i>Mycobacterium chelonae</i>				
<i>Mycobacterium fortuitum</i>				
<i>Haemophilus influenzae</i>				
<i>Mycobacterium marinum</i>				
<i>Brucella melitensis</i>				
<i>Brucella abortus</i>				
<i>Yersinia pestis</i>				
<i>Yersinia enterocolitica</i>				
<i>Bacillus anthracis</i>				
<i>Bacillus cereus</i>				
<i>Coxiella burnetii</i>				
<i>Bartonella henselae</i>				

<i>Bartonella quintana</i>				
<i>Pasteurella multocida</i>				
<i>Corynebacterium diphtheriae</i>				
<i>Bordetella pertussis</i>				
<i>Legionella pneumophila</i>				
<i>Pseudomonas aeruginosa</i>				
<i>Campylobacter jejuni</i>				
<i>Campylobacter fetus</i>				
<i>Stenotrophomonas maltophilia</i>				
<i>Acinetobacter baumannii</i>				
<i>Moraxella catarrhalis</i>				
<i>Aeromonas hydrophila</i>				
<i>Aeromonas caviae</i>				
<i>Listeria monocytogenes</i>				
<i>Actinomyces israeli</i>				
<i>Actinomyces bovis</i>				
<i>Actinomyces naeslundii</i>				

<i>Actinomyces viscosus</i>				
<i>Actinomyces odontolyticus</i>				
<i>Nocardia asteroides</i>				
<i>Nocardia brasiliensis</i>				
<i>Nocardia farcinica</i>				
... ..				

5. Conventional microbiological tests performed for CSF in HKU-SZH

Types of Tests	Conventional microbiology tests (CMTs)
Smear	CSF for AFB smear and India Ink staining
Culture	CSF culture for Bacteria and fungi
PCR	Real-time PCR for Epstein-Barr virus (EBV), cytomegalovirus (CMV), herpes simplex virus (HSV), varicella-zoster virus (VZV), Enterovirus (EV) and <i>Mycobacterium tuberculosis</i> (TB)
Antigen antibody reaction	Cryptococcus antigen, <i>Streptococcus pneumoniae</i> antigen, <i>Angiostrongylus cantonensis</i> IgG

CSF, cerebrospinal fluid, PCR, polymerase chain reaction

6. List of ≤ 5 sequence reads organisms

Organism	Sequence reads	Action taken	Clinical data confirm the infection	Medical opinion

Enterovirus	1	Report	Coincidence	True positive
Cytomegalovirus	3	Report	Coincidence	True positive
Epstein-Barr virus	2	Report	Coincidence	True positive
Epstein-Barr virus	3	Report	Coincidence	True positive
Pseudomonas aeruginosa	1	Report	Inconsistency	False positive
Mycobacterium kansasii	1	Report	Inconsistency	False positive
Epstein-Barr virus	1	Report	Inconsistency	False positive
Epstein-Barr virus	2	Report	Inconsistency	False positive
Aeromonas hydrophila	1	Report	Inconsistency	False positive
Pseudomonas aeruginosa	3	Report	Inconsistency	False positive
Epstein-Barr virus	2	Report	Inconsistency	False positive
Cytomegalovirus	1	Report	Inconsistency	False positive
Stenotrophomonas maltophilia	2	Report	Inconsistency	False positive
Epstein-Barr virus	1	Report	Inconsistency	False positive
Pneumocystis jiroveci	1	Report	Inconsistency	False positive
Cytomegalovirus	2	Report	Inconsistency	False positive
Mycobacterium tuberculosis complex	2	Report	Coincidence	True positive
Adenovirus	1	Report	Coincidence	True positive

Herpes simplex virus-2	5	Report	Coincidence	True positive
Pseudomonas aeruginosa	2	Report	Inconsistency	False positive
Streptococcus pneumoniae	4	Report	Coincidence	True positive

Reference:

1. Langelier, C., et al., Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. Am J Respir Crit Care Med, 2018. 197(4): p. 524-528.
2. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 2014; 30(15): 2114-20.
3. Bittinger, K., et al., Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome Biol, 2014. 15(10): p. 487.
4. Schlager, R., et al., Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. Arch Pathol Lab Med, 2017. 141(6): p.776-786.
5. Doughty, E.L., et al., Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer. PeerJ, 2014. 2: p. e585.

6. Simner, P.J., S. Miller, and K.C. Carroll, Understanding the Promises and Hurdles of Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Infectious Diseases. *Clin Infect Dis*, 2018. 66(5): p. 778-788.
7. Ozcolpan, O.O., et al., Distribution of nontuberculous mycobacteria isolated from clinical specimens and identified with DNA sequence analysis. *Mikrobiyol Bul*, 2015. 49(4): p. 484-93.
8. Van Ingen, J., et al., Global outbreak of severe *Mycobacterium chimaera* disease after cardiac surgery: a molecular epidemiological study. *Lancet Infect Dis*, 2017. 17(10): p. 1033-1041.
9. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*. 2009;25:1754–60.
11. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome research*. 2019;29:831–42.
12. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome research*. 2014;24:1180–92.