Supplementary Material. Methods of mNGS

The methods of mNGS in Guangxi KingMed Diagnostics were same as that described in a previous published mNGS article which was conducted in Guangzhou KingMed Diagnostics [1].

1. Sample Processing

Pleural effusion was collected and transported with dry ice. Blood was stored in EDTA tubes, from which plasma was separated by centrifuging at 1600 g for 10 min at 4℃.

2. Sequencing and Quality Control

In this study, nucleic acid (DNA and RNA) extraction and library preparation were performed on samples through the lab's self-built process. Both nucleic acid extraction and library preparation were conducted in parallel with quality control samples. Qubit was used to measure the concentration of the library. Single-end 75bp sequencing was carried out using Illumina nextseq 550 system with 75 cycles Reagent Kit. Then, the low-quality sequencing data were filtered out. After the removal of the sequences mapped to human reference genome, the remaining data were aligned to the microbial genome database.

3. Bioinformatic Analysis

The adapter sequences, low-quality data, and polyG tails were remove to generate the clean data using fastp v0.20.0 [2]. Then sequences that can be mapped to human reference genome were filtered using bwa v0.7.10-r789 [3]. The alignment of the remaining microbial data was carried out using bwa v0.7.10-r789 [3] and Sequence-Based Ultra-Rapid Pathogen Identification (SURPI v1.0.18) pipeline (UCSF), a previously published research pipeline for pathogen identification [4, 5]. For clinical usage, a self-built microorganisms database "MetagenomicX" was used to aligned the sequencing data. The database contains 36497 microorganisms' genomes, which covers most of the microorganisms 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.

References:

[1] Shi Y, Chen J, Shi X, Hu J, Li H, Li X, et al. A case of chlamydia psittaci caused severe pneumonia and meningitis diagnosed by metagenome next-generation sequencing and clinical analysis: a case report and literature review. BMC Infect Dis. 2021;21(1):621. doi:10.1186/s12879-021-06205-5.

[2] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90. doi:10.1093/bioinformatics/bty560.

[3] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi:10.1093/bioinformatics/btp324.

[4] 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 Res. 2019;29(5):831-42. doi:10.1101/gr.238170.118.

[5] 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 Res. 2014;24(7):1180-92. doi:10.1101/gr.171934.113.