Supplementary methods

NGS sequencing

Library preparation and sequencing. The NGS Automatic Library Preparation System was used for both DNA extraction and library preparation. Different reagents were used based on the type of sample, including those from the Nucleic Acid Extraction Kit (cat. no. MD013; Hangzhou Matridx Biotechnology Co., Ltd.) and Total DNA Library Preparation Kit (body fluid samples) (cat. no. MD001T; Hangzhou Matridx Biotechnology Co., Ltd.). After preparing the libraries, they were pooled and sequenced on an Illumina NextSeq500 system (Illumina, Inc.) using a 75-cycle sequencing kit. Each sample yielded between 10-20 million reads (1,2).

Bioinformatic pipeline. Raw sequencing data were analyzed by a bioinformatic pipeline, which included the following steps: i) Unnecessary adapter sequences and low-quality bases (Q-score cutoff, 20) were trimmed off in the pipeline; ii) sequencing reads were aligned with the human reference genome (GRCh38.p13; https://www.ncbi.nlm.nih.gov/assembly/2334371) and only unique, mapped reads were selected for subsequent analysis (3); iii) features of data were noted during a mapping step, and copy number variation information was obtained from the calling method. Specifically, after mapping, the waviness (standard deviation of the read fold-change of each bin) of data and normalized read counts were noted as features. Next, information on copy number variation was obtained using the calling method (4); iv) low-complexity reads were removed from the data. Next, the remaining sequencing data were aligned using BWA to reference databases such as NCBI nt (https://www.ncbi.nlm.nih.gov/nucleotide/) and GenBank (https://www.ncbi.nlm.nih.gov/genbank/) (5). This alignment process helped to identify the microbial species present in the sequencing data. The purpose of this method was to obtain information about which microorganisms were present in the sample being analyzed.

Immunophenotyping by flow cytometry

Instrumentation and reagents. The BD FACSCanto II flow cytometer (BD Biosciences) was utilized with dual lasers for four-color fluorescence excitation. The utilized reagents included four-color-labeled monoclonal antibodies specific to myeloid markers like CD13, CD14, CD15, CD33, CD11b, and MPO. Additionally, T cells were marked by CD4, CD8, CD5, CD2, CD3, and CD7, while the B-cell lineage was marked by CD19, CD10, CD20, CD22, and cCD79a. Stem cells and non-specific markers such as CD34, HLA-DR, CD38, CD1a, CD117, CD16, CD56, CD58, CD64, CD71, CD123, CD45, Ccd3, CD123, TCRα/β, TCRγ/δand TdT were also included. All the reagents, lysis solution and PBS used in the study were procured from BD Biosciences.

Pre-processing. The single monoclonal antibody, four-color fluorescence-labeled direct method was used, in which fluorescent markers, including fluorescein isothiocyanate, phycoerythrin, allophycocyanin and CD45 PerCP were used to label antibodies. Suspended pleural fluid sample (50 μl) and 5 μl of each of the four directly labeled fluorescent antibodies were added to each tube. The mixture was thoroughly mixed, allowed to stand in the dark at 4˚C for 30 min, and, if necessary, part of the CD antigen was appropriately increased based on the test results. Next, 1,000 μl lysis solution was added to each tube. The tubes were left standing for 15 min at room temperature before centrifugation at 1,036 x g for 5 min at room temperature. The upper supernatant was discarded and 1,000 μl PBS was added, centrifuged at 1,036 x g for 5 min at room temperature, and the centrifugation was repeated once more. The upper supernatant was discarded, and 500 μl PBS was added to resuspend the cells. After being thoroughly mixed, the cells were analyzed using flow cytometry.

Flow cytometry analysis. A total od 20,000 cells were collected each time and analyzed using BD FACSDiva software (BD Biosciences). CD45/SSC was utilized for gating analysis. The positive threshold for membrane antigen expression was set at >20% and for cytoplasmic antigen expression at >10%.

References

1. Zhang R, Zhuang Y, Xiao ZH, Li CY, Zhang F, Huang WQ, Zhang M, Peng XM and Liu C: Diagnosis and surveillance of neonatal infections by metagenomic next-generation sequencing. Front Microbiol 13: 855988, 2022.
2. Shen H, Shen D, Song H, Wu X, Xu C, Su G, Liu C and Zhang J: Clinical assessment of the utility of metagenomic next-generation sequencing in pediatric patients of hematology department. Int J Lab Hematol 43: 244-249, 2021.
3. Jiang H, Wu C, Xu J, Wang Q, Shen L, Ou X, Liu H, Han X, Wang J, Ding W, *et al*: Bacterial and fungal infections promote the bone erosion progression in acquired cholesteatoma revealed by metagenomic next-generation sequencing. Front Microbiol 12: 761111, 2021.
4. Su J, Han X, Xu X, Ding W, Li M, Wang W, Tian M, Chen X, Xu B, Chen Z, *et al*: Simultaneous detection of pathogens and tumors in patients with suspected infections by next-generation sequencing. Front Cell Infect Microbiol 12: 892087, 2022.
5. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW: GenBank. Nucleic Acids Res 41(Database issue): D36-D42, 2013.