Open Forum Infectious Diseases

MAJOR ARTICLE

Performance of Nanopore and Illumina Metagenomic Sequencing for Pathogen Detection and Transcriptome Analysis in Infantile Central Nervous System Infections

Kazuhiro Horiba,[1,2,3,6] Yuka Torii,[3] Yuta Aizawa,[4] Makoto Yamaguchi,[3] Kazunori Haruta,[3] Toshihiko Okumura,[3] Takako Suzuki,[3] Yoshihiko Kawano,[5] Jun-ichi Kawada,[3] Shinya Hara,[5] Akihiko Saitoh,[4] Christian G. Giske,[6] Tomoo Ogi,[1,2] and Yoshinori Ito[3,7]

[1]Department of Genetics, Research Institute of Environmental Medicine, Nagoya University, Nagoya, Japan, [2]Department of Human Genetics and Molecular Biology, Nagoya University Graduate School of Medicine, Nagoya, Japan, [3]Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan, [4]Department of Pediatrics, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan, [5]Department of Pediatrics, TOYOTA Memorial Hospital, Toyota, Japan, [6]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden, and [7]Department of Pediatrics and Child Health, Nihon University School of Medicine, Tokyo, Japan

Background. Infantile central nervous system infections (CNSIs) can be life-threatening and cause severe sequelae. However, the causative microorganism remains unknown in >40% of patients with aseptic infections. This study aimed to analyze the metagenome for detection of pathogens and the transcriptome for host immune responses during infection in a single cerebrospinal fluid (CSF) sample using 2 different next-generation sequencing (NGS) platforms, Nanopore and Illumina.

Methods. Twenty-eight CNSIs patients (<12 months) were enrolled, and 49 clinical samples (28 CSF and 21 blood) were collected. The DNA extracted from all 49 samples was sequenced using the Illumina sequencer for the detection of pathogens. Extracted RNA was obtained in sufficient quantities from 23 CSF samples and subjected to sequencing on both Nanopore and Illumina platforms. Human-derived reads subtracted during pathogen detection were used for host transcriptomic analysis from both Nanopore and Illumina sequencing.

Results. RNA metagenomic sequencing using both sequencing platforms revealed putative viral pathogens in 10 cases. DNA sequencing using the Illumina sequencer detected 2 pathogens. The results of Nanopore and Illumina RNA sequencing were consistent; however, the mapping coverage and depth to the detected pathogen genome of Nanopore RNA sequencing were greater than those of Illumina. Host transcriptomic analysis of Nanopore sequencing revealed highly expressed genes related to the antiviral roles of innate immunity from pathogen-identified cases.

Conclusions. The use of Nanopore RNA sequencing for metagenomic diagnostics of CSF samples should help to elucidate both pathogens and host immune responses of CNSI and could shed light on the pathogenesis of these infections.

Keywords. central nervous system infections; metagenomics; Nanopore sequencing; next-generation sequencing; transcriptomics.

Infantile central nervous system infections (CNSIs) are particularly frequent for those under the age of 1 year and can be life-threatening and cause severe sequelae in encephalitis and bacterial meningitis [1–3]. Early diagnosis and prompt treatment with appropriate antimicrobials are highly important. The gold standard for diagnosing CNSI is the detection of pathogens from cerebrospinal fluid (CSF) [4]. However, this approach can be challenging for the limited availability and volume of CSF samples, particularly in infants. Conventional microbiological methods, such as polymerase chain reaction (PCR), have failed to detect microorganisms in ~60% of aseptic meningitis cases with negative bacterial and fungal cultures [5] and in 40%–50% of encephalitis [6, 7]. This is because the microorganisms that cause CNSIs are diverse, and assays that can deal with all of them have not been put into practical use. Currently, there is a need for more sensitive and comprehensive assays to diagnose CNSIs.

Next-generation sequencing (NGS) has enabled the simultaneous decoding of a large number of nucleotide sequences present within a sample. Currently, there are 2 major platforms for this technology: short-read sequencing and long-read sequencing [8]. Metagenomic NGS can detect a wide variety of pathogens using a single technique [9]. This approach is effective for identifying unrevealed pathogens in cases that cannot be diagnosed using conventional microbiological methods. Metagenomic NGS using short-read sequencing has been used

Received 30 May 2022; editorial decision 26 September 2022; accepted 28 September 2022; published online 30 September 2022

Correspondence: Kazuhiro Horiba, MD, PhD, Department of Genetics, Research Institute of Environmental Medicine, Nagoya University, Nagoya 464-8601, Japan (khoriba@med.nagoya-u.ac.jp).

Open Forum Infectious Diseases®

https://doi.org/10.1093/ofid/ofac504

clinically for the diagnosis of infectious diseases [10–12]. We have also previously reported the detection of pathogens by metagenomic NGS using short-read sequencing of pediatric CSF samples [13, 14], blood samples [15, 16], and bronchoalveolar lavage fluid samples [17]. In the field of clinical infectious diseases, Nanopore sequencing, one of the long-read sequencing methods, is also attracting attention because of its simplicity and rapidity for real-time sequencing [18]. However, the number of reports on the application of Nanopore sequencing for the detection of pathogens in clinical samples is limited, compared with reports using short-read sequencing [19, 20].

This study aimed to detect the causative pathogens in CSF and blood samples of patients with infantile CNSI using metagenomic NGS analysis. The extracted RNA from CSF samples was analyzed on 2 different sequencing platforms, Nanopore sequencing for long reads and Illumina sequencing for short reads, to compare their performance with regard to pathogen detection. Finally, host transcriptome analysis associated with metagenomic NGS was performed using Nanopore and Illumina sequencing.

METHODS

Patient Consent

The study design and methods were approved by the Institutional Review Board of Nagoya University Hospital (No. 9069). The methods were carried out in accordance with approved guidelines. Written informed consent was obtained from the guardians of every patient.

Patients and Samples

Twenty-eight pediatric patients, treated between June 2012 and April 2020, were enrolled in this study. Eligible patients <1 year of age underwent evaluation for CNS infection. CNS infection was identified in patients with fever (>37.5°C) during the presenting illness and 1 or more of the following: a depressed or altered level of consciousness, seizures and/or focal neurological findings, CSF pleocytosis, abnormal results of an electroencephalogram, and abnormal neuroimaging results. A total of 49 clinical samples (28 CSF and 21 blood samples) from 28 pediatric patients were collected as residual samples after routine clinical testing in the microbiology laboratory. Five CSF samples were unavailable for RNA extraction. All patients underwent CSF and blood culture. Blood culture was performed as previously described [15]. CSF cultures were performed manually with CO2 culture on sheep blood agar and chocolate agar, and with aerobic culture on bromothymol blue (BTB) lactate agar and Sabouraud agar. Blood and CSF samples were obtained in the acute phase. PCR screening for herpes simplex virus using previously described methods [21] yielded negative results. The multiplex PCR assays were not performed in the participating hospitals. All sequencing assays were performed at the Research Institute of Environmental Medicine, Nagoya University.

Illumina Sequencing Library Preparation and Sequencing

The Illumina sequencing libraries were prepared from DNA extracted from 200 µL of CSF and blood samples using a QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). RNA was extracted from 200 µL of CSF using a NucleoSpin Blood Kit (Macherey-Nagel, Düren, Germany). Thereafter, cDNA was synthesized and amplified as previously described [22]. DNA and cDNA sequencing libraries were prepared using a Nextera XT Library Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions with slight modifications [15]. TapeStation high-sensitivity DNA ScreenTape assays (Agilent, CA, USA) were used for the sample quality control of the libraries. Digital PCR with primers specific to the Illumina library adapter sequence for the QX200 Droplet Digital PCR System (Bio-Rad, Richmond, CA, USA) was used for quantification of the libraries. Illumina sequencing libraries were then sequenced on the HiSeq × system (Illumina) using the 2×150-bp paired-end protocol.

Nanopore Sequencing Library Preparation and Sequencing

In all samples, extracted DNA was not available in sufficient quantities to prepare DNA libraries for both Nanopore and Illumina sequencing; thus, only Illumina sequencing was done. Synthesized cDNA was simultaneously used to construct the Nanopore sequencing libraries. Nanopore sequencing library preparation was performed according to the manufacturer's instructions for a Ligation Sequencing Kit (SQK-LSK109; Oxford Nanopore Technologies, Oxford, UK) and a Native Barcoding Expansion (EXP-NBD104 and EXP-NBD114; Oxford Nanopore Technologies). TapeStation genomic DNA ScreenTape assays (Agilent) were used to control the quality control of cDNA and the libraries. Sequencing was performed on a PromethION platform (Oxford Nanopore Technologies) using R9.4.1 flow cells. MinKNOW, version 20.06.9 (Oxford Nanopore Technologies), was used to collect and demultiplex the raw sequencing data and simulate the output of sequencing reads at sequence time. Guppy, version 4.0.11 (Oxford Nanopore Technologies), was used for base calling of the raw data after the sequencing runs were completed.

Metagenomic NGS Data Processing of Nanopore and Illumina Sequencing

The DNA sequencing data of blood and CSF samples generated by the Illumina sequencing platform were processed with the metagenomic pipeline PATHDET as described previously, and pathogen candidates were identified [16]. The National Center for Biotechnology Information (NCBI) nucleotide collection (nt) database [23] was used for this process. RNA sequencing data from the Nanopore sequencing platform were processed as follows. Quality filtering was performed using

NanoFilt [24], and quality control was performed using NanoPlot [24]. Next, the remaining sequence reads, mapped to the human genome database using Minimap2 [25] and Kraken2 [26], were removed. Kraken2 and BLAST [27] were used to search for highly similar species in the nt database from sequences that could not be classified by Kraken2. RNA sequencing data from the Illumina sequencing platform were processed using PATHDET as well as the DNA sequencing data. From both Nanopore and Illumina RNA sequencing analysis data, we collected RNA viruses as pathogen candidates as follows. The sequencing reads assigned to viruses were stripped off for reagent-derived contamination [28], and subtracted with 0.056% (Nanopore sequencing) [29] and 0.47% (Illumina sequencing) [30] for the most detected reads in each sample to avoid cross-contamination due to multiplex sequencing. To avoid making calls based on potentially spurious alignments, the following criteria were used for identification of pathogen candidates: (1) at least 3 different reads specific to a particular viral species, (2) reads distributed over the whole genome, and (3) detection of >10 reads of viral sequences in total [13, 31].

Hybrid metagenome assemblies on Nanopore and Illumina RNA sequencing read data were performed using metaSPAdes [32]. Contigs with high length (>500 bases) and high similarity to RNA viruses in the nt database by BLAST search were collected as pathogen candidates.

Validation by PCR and Sanger Sequencing

In patient N05, human parvovirus B19 was validated by PCR assay as previously described [33]. In patients N01, N13, N15, N16, N17, N18, N19, N20, N23, and N26, in whom enterovirus or parechovirus was detected, nested reverse transcription PCR assays were performed targeting the viral protein 1 (VP1) region [34] or VP4/VP2 region [35] for enterovirus and VP3/VP1 region [36] for parechovirus A. The PCR products were subsequently subjected to Sanger sequencing as previously described [37]. Genotypes were identified based on BLAST analysis [27].

Transcriptome Analysis

The Nanopore FASTQ files were aligned to the human reference sequence file (hg38) using Minimap2 with default parameters, and the Illumina FASTQ files were aligned using HISAT2 [38] after the trimming. Next, the alignments were passed to StringTie for transcript assembly, transcript abundance was estimated, and read-count values were calculated based on an established method [39]. Finally, the TCC package was used for normalization of read count by trimmed mean of M values and an exact test of edgeR and identification of DEGs between samples with default options [40]. The threshold of differential expression was set at $P < .05$, found using the TCC package. The TCC package also drew heat maps and volcano plots to visualize transcriptome analysis. Metascape was used for gene set and the protein–protein interaction enrichment analysis [41, 42].

Statistical Analysis

Nanopore and Illumina sequence read data were analyzed as follows. First, the Shapiro-Wilk test was used to test whether the Nanopores or short reads were normally distributed. If both were significant, Wilcoxon's signed-rank test was applied; otherwise, a corresponding *t* test was applied. All statistical analyses were performed using R, version 4.0.4, and the R package "Rcmdr" [43]. A receiver operating characteristic (ROC) analysis was constructed for each of the genes, followed by area under curve (AUC) with sensitivity and specificity for positive NGS result using the R package "pROC" [44]. Statistical significance was set at $P < .05$.

RESULTS

Patients and Sequencing

In total, 28 CSF samples and 21 blood samples were collected from 28 pediatric patients. Patient characteristics are shown in Table 1 and Supplementary Table 1. None of the patients received antimicrobial treatment before CSF and blood samples were collected. No causative pathogens were found in 26 patients. One patient had a positive CSF culture for *Proteus mirabilis* (N14). Two patients had positive blood cultures (N08,

Table 1. Clinical Characteristics of 28 Patients

Characteristic	Infantile Central Nervous System Infections (n = 28)	
Age, median (range), d	23.5	(3–311)
Sex, male, No. (%)	13	(46)
Length of stay, median, d	8	(5–113)
Clinical signs and symptoms, No. (%)	…	…
Depressed or altered level of consciousness	16	(57)
Poor feeding	13	(46)
Vomiting	2	(7)
Seizure	3	(11)
Upper respiratory symptom	3	(11)
Diarrhea	3	(11)
Bulging fontanelle	5	(18)
Systemic inflammatory response syndrome[a]	8	(29)
Pleocytosis[b]	23	(82)
Cerebrospinal fluid test, median	…	…
Cell count,/μL	255	(1–1301)
Neutrophil's count,/μL	46.5	(0–875)
Protein, mg/dL	84	(14–271)
Glucose, mg/dL	52.5	(19–203)
Blood tests, median	…	…
White blood cell count,/μL	11 250	(1100–47 500)
C-reactive protein, mg/dL	0.22	(0–6.8)

[a]Systemic inflammatory response syndrome was defined by 4 parameters: body temperature, tachycardia, hyperventilation, and white blood cell count. Specifically, it was diagnosed when body temperature was >38.5°C or <36.0°C, leukocytosis or leukocytopenia was present, and 2 or more parameters were present.

[b]Pleocytosis was defined as the following: cerebrospinal fluid cell count >30/μL for newborns (0–8 weeks), >5/μL for infants (>8 weeks).

Streptococcus agalactiae; N14, *P. mirabilis*); however, 1 of these patients (N08) had no blood sample available for this study.

To allow for comprehensive microbial detection and comparison of the Nanopore and Illumina sequencing platforms, each CSF sample was subjected to multiple metagenomic approaches (Figure 1*A*). From the collected CSF samples, 28 DNA and 23 RNA samples were extracted in sufficient quantities to allow for NGS library preparation. From the blood samples, 21 samples of DNA were also extracted. All the extracted DNA was sequenced using Illumina sequencing. All extracted RNA was sequenced using both Nanopore and Illumina sequencing (Supplementary Table 2).

Pathogen Candidates in CSF and Blood Samples by Metagenomic NGS

In CSF samples, pathogen candidates were reported in 12 (43%) of 28 patients (Figure 1*B*, Table 2). RNA sequencing of CSF detected enterovirus B or parechovirus A as the candidate pathogen in 10 patients. In 6 cases using RNA sequencing, >80% of the reads were aligned to 3 specific serotypes: coxsackievirus B5, echovirus E7, and human parechovirus 3. PCR products were produced in 4 of 9 enterovirus-positive samples (N01, N16, N20, and N26), and all serotypes were matched between NGS hybrid assembly and Sanger sequencing (Supplementary Table 3). In patient N17, NGS and Sanger sequencing resulted in the same genotype, human parechovirus A3.

DNA sequencing revealed 2 candidate pathogens: human parvovirus B19 and *P. mirabilis*. The presence of human parvovirus B19 in patient N05 was confirmed by PCR. *P. mirabilis*, cultured from CSF and blood in patient N14, was detected as the pathogen candidate by metagenomic NGS in CSF. No candidate pathogens were detected in the blood samples using metagenomic NGS. However, *P. mirabilis* was the most sequenced microbe in the sequencing reads from N14's blood. We could not detect *S. agalactiae* in sequencing of CSF from patient N08, whose blood culture test was positive for this pathogen.

Comparison of Nanopore Sequencing and Illumina RNA Sequencing

Both Nanopore sequencing and Illumina RNA sequencing detected the same viral species as pathogen candidates. Nanopore sequencing output contained significantly more human genome–derived sequencing reads ($P < .001$) (Figure 2*A*) and pathogen-derived reads ($P = .008$) (Figure 2*B*) than Illumina sequencing output. In the hybrid assembly, contigs of viral species were generated in 11 cases. Nine contigs were of the same viral species and serotypes as the pathogen candidates for metagenomic NGS (Table 2). For these 9 cases, the mapping statuses of Nanopore sequencing and Illumina sequencing were compared, as shown in Supplementary Figure 1. In this analysis, the viral genomes of serotypes selected from the hybrid assembly results were used for the mapping reference genome. Nanopore sequencing generated significantly greater mapping coverage ($P = .008$) (Figure 2*C*) and mapping depth ($P = .008$) (Figure 2*D*) than Illumina sequencing.

Nanopore sequencing allows for the retrieval of sequence data even while the sequencing is still running. We simulated sequence reads at several sequencing run times (1, 2, 4, 8, 12, 24, and 48 hours) to analyze the mapping coverage for the pathogen candidate of each patient. The mapping coverage reached >80% after 2 hours of sequencing run time in 5 cases (Figure 2*E*). In all cases, the mapping coverage did not change after >24 hours of sequencing. These observations suggest that Nanopore sequencing can provide the sequencing data required for pathogen detection in a relatively short time.

Transcriptome Analysis

After screening for pathogens from metagenomic NGS data, human-derived reads were used for host transcriptomic analysis from both Nanopore and Illumina sequencing. A multidimensional scaling plot based on gene counts for all RNA-sequenced samples did not produce clusters relevant for pathogen detection by metagenomic NGS (Supplementary Figure 2). Comparing the 2 patient groups (the 12 patients with identified pathogens via metagenomic NGS and the 9 patients with an unidentified pathogen), 638 differentially expressed genes (DEGs) were identified (Supplementary Figures 3A and 4, Supplementary Table 4) through Nanopore sequencing, and 107 DEGs were identified in the Illumina sequencing data comparison (Supplementary Figures 3B and 5, Supplementary Table 5). All DEGs from Nanopore and Illumina sequencing were enriched to the corresponding gene ontology (GO) categories using Metascape. Seven DEGs, *IDO1*, *MX1*, *ISG15*, *OAS1*, *WARS1*, *USP17L5*, and *SIGLEC14*, were common to both the Nanopore and Illumina sequencing. ROC curves and AUCs were also calculated for these 7 DEGs (Supplementary Figures 6 and 7). Of these, *MX1*, *ISG15*, and *OAS1* were annotated as negatively regulated viral genome replication (GO:0045071). In the group of pathogens detected in the CSF, the terms innate immune response (GO:0045087, $-\log_{10}P = 15.9$) and deubiquitination (R-HAS-5688426, $-\log_{10}P = 14.7$) were found to be enriched in Nanopore sequencing (Figure 3*A*). The term negative regulation of viral genome replication (GO:0045071, $-\log_{10}P = 12.8$) was enriched in Illumina sequencing (Supplementary Figure 8A). These GO terms were not enriched in the group negative for pathogens. For further interpretation, protein–protein interaction enrichment analysis was performed (Figure 3*B*; Supplementary Figure 8B). The resulting network was composed of densely connected network components identified from the input DEG list (Figure 3C; Supplementary Figure 8C). In the network from Nanopore sequencing, the terms innate immune response (GO:0045087, $-\log_{10}P = 18.2$), interferon signaling (R-HAS-913531, $-\log_{10}P = 15.5$), and defense response to virus (GO:0051607, $-\log_{10}P = 15.4$) were enriched. For Illumina sequencing, the terms negative regulation of viral genome

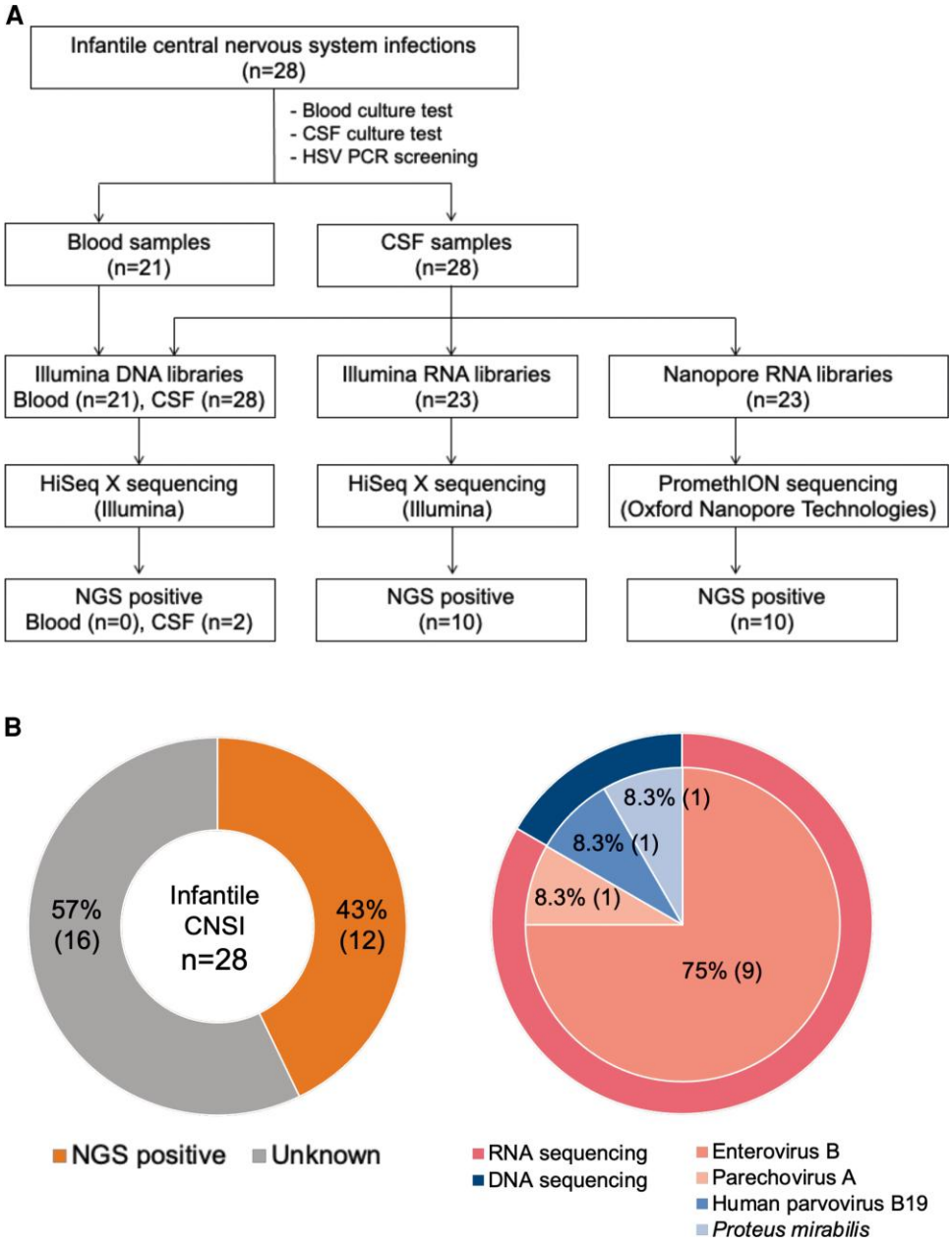

Figure 1. Flowchart and pie charts for patients with central nervous system infections (CNSIs). *A,* Flowchart of an overview of DNA and RNA library preparation from blood and cerebrospinal fluid (CSF) samples and sequencing platform selection. Nanopore DNA sequencing was not available due to insufficient extracted DNA quantities. *B,* Results of next-generation sequencing (NGS) and a pie chart of detected pathogen candidates using DNA/RNA workflows.

replication (GO:0045071, $-\log_{10}P = 15.0$), interferon alpha/beta signaling (R-HSA-909733, $-\log_{10}P = 14.5$), and interferon signaling (R-HSA-913531, $-\log_{10}P = 14.4$) were enriched. In the group negative for pathogens, GO terms that overlapped with these ones were not enriched.

The transcriptome analysis was also performed with and without pleocytosis and with and without C-reactive protein (CRP) elevation. Multidimensional scaling plots based on gene counts for all RNA-sequenced samples did not generate clusters relevant for pleocytosis or CRP elevation (Supplementary Figure 9). There were 53 genes in Nanopore sequencing and 13 genes in Illumina sequencing as upregulated DEGs in the pleocytosis group, but no DEGs were common in both sequencing platforms (Supplementary Figure 10A, Supplementary Tables 6 and 7). In the pleocytosis group, the terms negative regulation of execution phase of apoptosis (GO:1900118, $-\log10P = 6.7$) and natural killer cell–mediated cytotoxicity (hsa04650, $-\log10P = 6.1$) were found to be enriched in Nanopore sequencing (Supplementary Figure 10). There were 3160 genes in Nanopore sequencing and 26 genes in Illumina sequencing as

Table 2. Pathogen Candidates Detected in CSF Samples by Nanopore and Illumina Sequencing

Patient	Sequencing Methods	Pathogen Candidates (Species)	Reads	Major Serotype (Occupancy)	Hybrid Assembly (Accession Number)
N01	Nanopore RNA	Enterovirus B	19 242	Echovirus E7 (98%)	Echovirus E7 (KU355273.1)
	Illumina RNA	Enterovirus B	5956	Echovirus E7 (88%)	
N05	Illumina DNA	Primate erythroparvovirus 1	13 623	Human parvovirus B19 (79%)	NA
N13	Nanopore RNA	Enterovirus B	3901	Coxsackievirus B2 (52%)	Coxsackievirus B2 (KU574632.1)
	Illumina RNA	Enterovirus B	7428	Coxsackievirus B2 (58%)	
N14	Illumina DNA	*Proteus mirabilis*	49 284	NA	NA
N15	Nanopore RNA	Enterovirus B	937	Coxsackievirus B4 (26%)	Coxsackievirus B4 (MW015043.1)
	Illumina RNA	Enterovirus B	205	Coxsackievirus B4 (32%)	
N16	Nanopore RNA	Enterovirus B	6576	Coxsackievirus B5 (99%)	Coxsackievirus B5 (MW015056.1)
	Illumina RNA	Enterovirus B	2825	Coxsackievirus B5 (98%)	
N17	Nanopore RNA	Parechovirus A	1650	Human parechovirus 3 (99%)	Human parechovirus 3 (LC043127.2)
	Illumina RNA	Parechovirus A	1281	Human parechovirus 3 (90%)	
N18	Nanopore RNA	Enterovirus B	193	Coxsackievirus B5 (98%)	Coxsackievirus B5 (MW015056.1)
	Illumina RNA	Enterovirus B	67	Coxsackievirus B5 (100%)	
N19	Nanopore RNA	Enterovirus B	93	Coxsackievirus B4 (40%)	None
	Illumina RNA	Enterovirus B	71	Coxsackievirus B4 (51%)	
N20	Nanopore RNA	Enterovirus B	1007	Coxsackievirus B5 (100%)	Coxsackievirus B5 (MW015056.1)
	Illumina RNA	Enterovirus B	357	Coxsackievirus B5 (100%)	
N23	Nanopore RNA	Enterovirus B	148	Coxsackievirus B5 (96%)	Coxsackievirus B5 (MW015056.1)
	Illumina RNA	Enterovirus B	84	Coxsackievirus B5 (100%)	
N26	Nanopore RNA	Enterovirus B	92	Coxsackievirus B4 (28%)	Coxsackievirus B4 (MN590273.1)
	Illumina RNA	Enterovirus B	94	Coxsackievirus B4 (33%)	

Abbreviation: CSF, cerebrospinal fluid.

upregulated DEGs in the CRP elevation group, and 11 DEGs were common in both sequencing platforms (Supplementary Figure 11A, Supplementary Tables 8 and 9). In the CRP elevation group, the enriched terms are shown in Supplementary Figure 11. The 11 common DEGs were enriched to the term intracellular protein transport (GO:0006886, −log10P = 3.5).

DISCUSSION

In this study, we have shown that CSF samples from 28 infantile CNSIs were performed with metagenomic NGS on 2 sequencing platforms, Nanopore sequencing and Illumina sequencing, to reveal pathogenic microorganisms in 11 cases without a definite diagnosis for pathogens from a culture test and human simplex herpes virus PCR (42%). In addition, *P. mirabilis* cultured with conventional CSF testing could be detected by Illumina DNA sequencing. Seventy-five percent of the newly detected microorganisms were enteroviruses. Human parechovirus 3 and human parvovirus B19 were also detected. Enteroviruses are the most common cause of pediatric viral meningitis [3]. All patients in whom enteroviruses were detected by metagenomic NGS developed CNSI between June and August (data not shown), which coincides with the enterovirus epidemic period in Japan [45]. Case N17, in which human parechovirus 3 was detected by metagenomic NGS, was 6 days old, consistent with the epidemiology of parechovirus CNSI, which occurs mainly in young infants [46]. Human parvovirus B19 is also considered a potential CNSI-causing virus, although less frequently [47]. Thus, all the viruses detected by metagenomic NGS were clinically plausible. In 5 enterovirus-positive samples, PCR confirmation was not possible despite relatively high mapping coverage in the Nanopore and Illumina sequencing. The degradation of the stored RNA was a possible reason for negative results. Enterovirus samples with a low number of sequencing reads might have viral loads below the detection level of the PCR assay.

We also compared the performance of Nanopore and Illumina sequencing in 23 cases of RNA sequencing. We found the pathogen candidates in 10 cases, and those from Nanopore and Illumina sequencing were matched at the taxonomic rank of species. Gu et al. also reported that the sensitivities and specificities for bacterial and fungal detection across Illumina and Nanopore sequencing are comparable in DNA sequencing in independent Illumina (n = 127) and Nanopore (n = 43) validation sets [19]. Hybrid metagenome assembly using Nanopore and Illumina sequencing, which reconstructs highly accurate pathogen genome sequences, could infer viral serotypes. In our study, the criteria for the identification of viral candidates in RNA sequencing are based on the number of sequencing reads and mapping status to the reference genome. Sequencing of more definitive cases would make these cutoff value settings even more precise. By reanalyzing both sequencing data using the serotype genomes from the hybrid assembly as a reference, we found that the Nanopore sequence had more

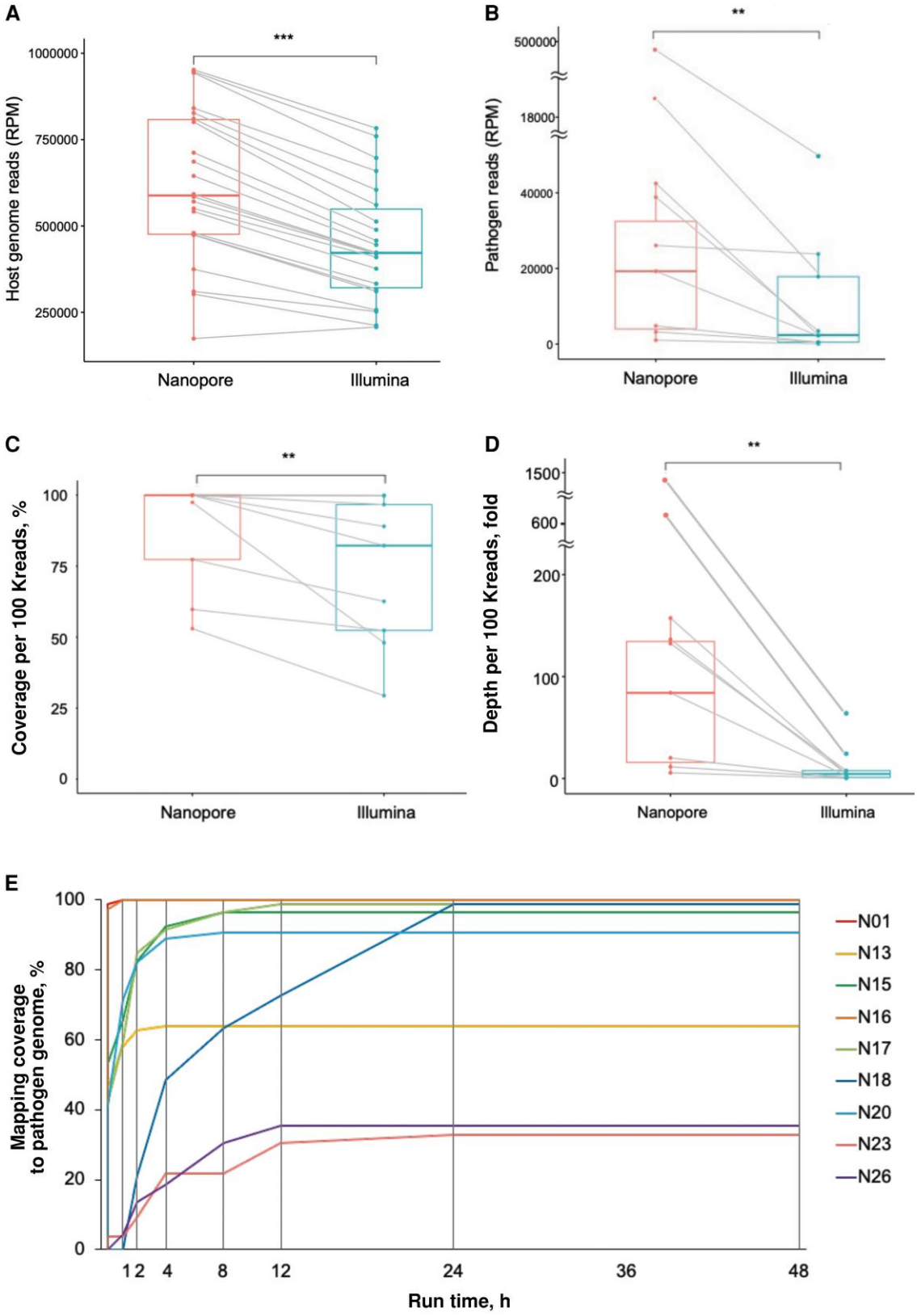

Figure 2. Performance of Nanopore and Illumina sequencing for pathogen detection. *A*, The number of reads of human genome origin per million reads of sequencing output (reads per million [RPM]) was compared. *B*, The number of reads mapped to the pathogen genome per million reads of sequencing output. The reference genome for the mapping was selected based on hybrid metagenome assembly. Additionally, 100 000 reads randomly extracted from the sequencing output were mapped to the reference genome to determine (C) mapping coverage and (D) mapping depth. E, Based on the sequencing recordings of PromethION, we simulated the sequence output for each run time and calculated the mapping coverage. $^{**}P < .01$ (B–D); $^{***}P < .001$ (A).

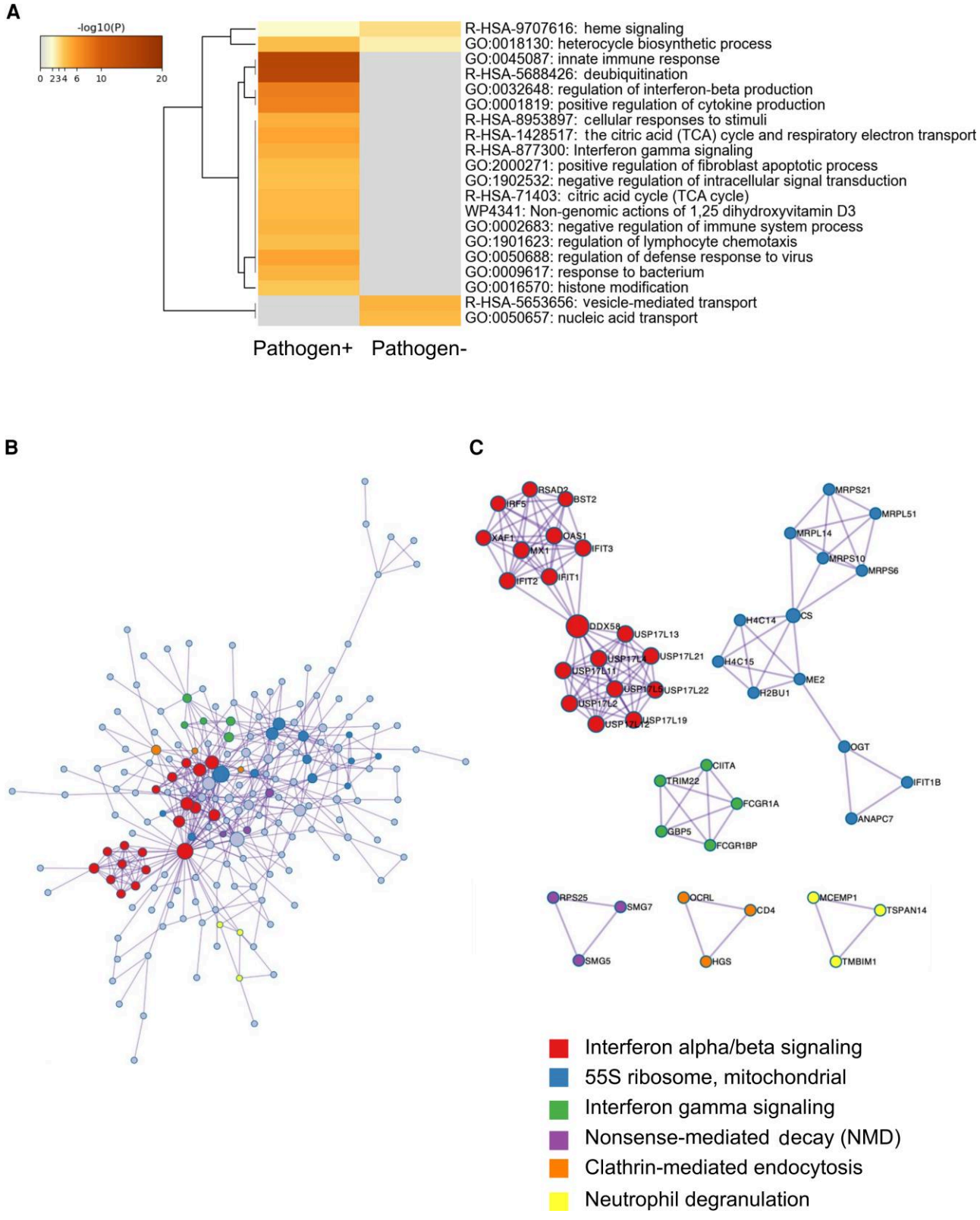

Figure 3. Enrichment analysis of differentially expressed genes from Nanopore sequencing. *A*, Heatmap of enriched terms across differentially expressed gene lists from Nanopore sequencing, colored based on *P* values. *B*, Protein–protein interaction network and (C) molecular complex detection (MCODE) components are identified from the gene lists. Pathway and process enrichment analysis has been applied to each MCODE component independently. Functional description of the lowest *P* values of the corresponding components is shown.

reads of both host and pathogenic microbes than the Illumina sequence. Furthermore, the mapping coverage and depth normalized by the sequencing total output were also greater in Nanopore sequencing. It is reasonable to assume that this occurred because Nanopore sequencing, which is a long-read sequencing technology, has many bases per read; thus, more bases of Nanopore sequencing aligned to the reference genome in comparison with Illumina short-read sequencing. It is known that Illumina sequencing can not only identify a pathogenic virus in the CSF of a patient but can also identify the single nucleotide polymorphisms of the viral genome [48]. Nanopore sequencing as well as Illumina sequencing may provide useful information for clinical diagnostics, such as the genotype and specific gene regions of microorganisms. In addition, Nanopore sequencing can acquire output data during sequencing. Our simulations showed that a maximum of 24 hours of sequencing was sufficient to obtain enough data for pathogen genome analysis. This is equivalent to Illumina sequencing, which takes ~1 day [11], and is faster than culture-based pathogen identification, which can take several days. More importantly, Nanopore sequencing has the potential to identify pathogens in several hours, as has been shown in previous studies [19, 49]. It has been suggested that the time required for microbial diagnosis could be shortened by setting the sequencing time to several hours for use in clinical settings.

Transcriptome analysis was performed using the human-derived data leftover from metagenomic NGS analysis. Patients with identified pathogens via metagenomic NGS were found to have a distinct gene expression pattern compared with those without it. *MX1*, *ISG15*, and *OAS1* were found to be differentially expressed in both the Nanopore and Illumina data and were associated with antiviral roles in innate immunity. MX1 is induced by type I and II interferons and antagonizes the replication of viruses [50]. ISG15 is a ubiquitin-like protein that is conjugated to intracellular target proteins by activation by type I interferons and that inhibits viral replication [51]. OAS1 activates endoribonuclease L, which inhibits viral replication and limits the spread of infection [52]. Most of the pathogens detected by metagenomic NGS in this study were RNA viruses, and the expression of these genes would be a reasonable result of the impact of their infections. We also identified the GO term of deubiquitination by enrichment analysis for the set of Nanopore sequencing genes in patients with identifiable for pathogens. The GO term for deubiquitination was derived from 35 genes, including *XBP1*, which is involved in the inhibition of enterovirus entry [53]. Protein–protein interaction enrichment analysis also identified this GO term, suggesting that it is part of the host response to viral infection. Therefore, transcriptome analysis supports the results of metagenomic NGS and elucidates the host immune response [54]. These results may be useful for understanding the pathophysiology of various pediatric CNSIs. However, the transcriptome analysis alone would not be sufficient to observe an actual innate immunity activity in CNSIs. Thus, it is necessary to confirm the actual product using measurement of the cytokine protein profiles of spinal fluids [48].

The limitations of this study are that it was a retrospective study, and the inclusion only of infants under 1 year of age did not allow for extra clinical specimens for more analysis. Thus, DNA sequencing of the blood and CSF could not be performed on the Nanopore sequencing platform. However, we found that *P. mirabilis* derived sequencing reads from Nanopore and Illumina RNA sequencing data in N14; Conversely, we did not find sequencing data in N05 for human parvovirus B19 (data not shown). These results might suggest that CNSIs in N05 were no longer metabolically active but had occurred at some time in the past. For Nanopore sequencing, PrometION was selected in this study to handle multiple samples. However, the low-end platforms MinION and Flongle may be useful in clinical practice, allowing on-demand sequencing by using 1 sample per flow cell. Further research is needed to optimize sequencing settings using these platforms. In this study, only 23 patients with infantile CNSIs were eligible for transcriptome analysis. Therefore, additional studies with a larger number of patients with confirmed CNSIs are needed to obtain robust results.

In conclusion, both Nanopore sequencing and Illumina sequencing were able to detect pathogens in infant CNSI, and transcriptome analysis could be performed simultaneously. In addition, there were more host- and pathogen-derived reads obtained through Nanopore sequencing than with Illumina sequencing; Nanopore sequencing also showed the potential to analyze pathogen and host immune responses with a lower output than Illumina sequencing. Because Nanopore sequencers can process data in real time, they may be attractive for the field of clinical infectious diseases, where results need to be known in a short time. The use of Nanopore sequencing would help to elucidate both pathogens and host immune responses for basic infectious disease research and may help clarify the pathogenesis of many cases of CNSI.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We thank Dr. Shotaro Ando from Okazaki City Hospital, Dr. Yuichiro Sugiyama from Japanese Red Cross Aichi Medical Center Nagoya Daiichi Hospital, Dr. Shunsuke Kato from Kasugai Municipal Hospital, Dr. Jun Okamura from Ichinomiya Municipal Hospital, Dr. Kenichi Iwata from Nagoya Memorial Hospital, and Dr. Manabu Ino from Ogaki Municipal Hospital for their contributions to this study.

Financial support. This work was supported by a Grant-in-Aid for Young Scientists 20 K17464 and the Takeda Science Foundation to K.Ho.

Potential conflicts of interest. All authors: no reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Author contributions. K.Ho., T.O., and Y.I. designed the study. Y.T., M.Y., K.Ha., T.O., and T.S. collected clinical samples. K.Ha., Y.K., J.K., and S.H. reviewed medical records for clinical assessment of the patients. K.Ho., Y.T., M.Y., T.O., T.S., and J.K. performed NGS experiments. K.Ho. performed bioinformatic analysis for detection pathogen and transcriptome analysis. Y.A. and A.S. performed validation by PCR and Sanger sequencing. K.Ho., C.G., T.O., and Y.I. drafted the manuscript. All authors read and approved the final manuscript.

Data availability. The full data supporting the findings of this study are available from the corresponding author upon reasonable request. Sequencing reads generated in this study were deposited in the DDBJ Sequence Read Archive under experiment number DRX307218-307312 (https://ddbj.nig.ac.jp/resource/bioproject/PRJDB12113).

References

1. Okike IO, Ribeiro S, Ramsay ME, Heath PT, Sharland M, Ladhani SN. Trends in bacterial, mycobacterial, and fungal meningitis in England and Wales 2004–11: an observational study. Lancet Infect Dis **2014;** 14:301–7.
2. Britton PN, Dale RC, Blyth CC, et al. Causes and clinical features of childhood encephalitis: a multicenter, prospective cohort study. Clin Infect Dis **2020**; 70: 2517–26.
3. Martin NG, Iro MA, Sadarangani M, Goldacre R, Pollard AJ, Goldacre MJ. Hospital admissions for viral meningitis in children in England over five decades: a population-based observational study. Lancet Infect Dis **2016;** 16: 1279–87.
4. Miller JM, Binnicker MJ, Campbell S, et al. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2018 update by the Infectious Diseases Society of America and the American Society for Microbiology. Clin Infect Dis **2018**; 67:e1–94.
5. Aldriweesh MA, Shafaay EA, Alwatban SM, et al. Viruses causing aseptic meningitis: a tertiary medical center experience with a multiplex PCR assay. Front Neurol **2020**; 11:602267.
6. Granerod J, Ambrose HE, Davies NWS, et al. Causes of encephalitis and differences in their clinical presentations in England: a multicentre, population-based prospective study. Lancet Infect Dis **2010**; 10:835–44.
7. Erickson TA, Muscal E, Munoz FM, et al. Infectious and autoimmune causes of encephalitis in children. Pediatrics **2020**; 145:e20192543.
8. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet **2016**; 17:333–51.
9. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet **2019;** 20:341–55.
10. Wilson MR, Sample HA, Zorn KC, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med **2019**; 380: 2327–40.
11. Blauwkamp TA, Thair S, Rosen MJ, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat Microbiol **2019**; 4:663–74.
12. Graff K, Dominguez SR, Messacar K. Metagenomic next-generation sequencing for diagnosis of pediatric meningitis and encephalitis: a review. J Pediatric Infect Dis Soc **2021**; 10:S78–87.
13. Kawada JI, Okuno Y, Torii Y, et al. Identification of viruses in cases of pediatric acute encephalitis and encephalopathy using next-generation sequencing. Sci Rep **2016**; 6:33452.
14. Horiba K, Suzuki M, Tetsuka N, et al. Pediatric sepsis cases diagnosed with group B streptococcal meningitis using next-generation sequencing: a report of two cases. BMC Infect Dis **2021**; 21:531.
15. Horiba K, Kawada JI, Okuno Y, et al. Comprehensive detection of pathogens in immunocompromised children with bloodstream infections by next-generation sequencing. Sci Rep **2018**; 8:3784.
16. Horiba K, Torii Y, Okumura T, et al. Next-generation sequencing to detect pathogens in pediatric febrile neutropenia: a single-center retrospective study of 112 cases. Open Forum Infect Dis **2021**; 8:XXX–XX.
17. Takeuchi S, Kawada JI, Horiba K, et al. Metagenomic analysis using next-generation sequencing of pathogens in bronchoalveolar lavage fluid from pediatric patients with respiratory failure. Sci Rep **2019**; 9:12909.
18. Charalampous T, Kay GL, Richardson H, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol **2019**; 37:783–92.
19. Gu W, Deng X, Lee M, et al. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. Nat Med **2020**; 27:115–24.
20. Hong NTT, Anh NT, Mai NTH, et al. Performance of metagenomic next-generation sequencing for the diagnosis of viral meningoencephalitis in a resource-limited setting. Open Forum Infect Dis **2020**; 7:XXX–XX.
21. Ito Y, Kimura H, Yabuta Y, et al. Exacerbation of herpes simplex encephalitis after successful treatment with acyclovir. Clin Infect Dis **2000**; 30:185–7.
22. Suzuki T, Kawada JI, Okuno Y, et al. Comprehensive detection of viruses in pediatric patients with acute liver failure using next-generation sequencing. J Clin Virol **2017**; 96:67–72.
23. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res **2013;** 41:D8–20.
24. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. Nanopack: visualizing and processing long-read sequencing data. Bioinformatics **2018;** 34: 2666–9.
25. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **2018;** 34:3094–100.
26. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol **2014;** 15:R46.
27. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. Bioinformatics **2008;** 24:1757–64.
28. Wally N, Schneider M, Thannesberger J, et al. Plasmid DNA contaminant in molecular reagents. Sci Rep **2019**; 9:1652.
29. Xu Y, Lewandowski K, Lumley S, et al. Detection of viral pathogens with multiplex Nanopore MinION sequencing: be careful with cross-talk. Front Microbiol **2018**; 9:2225.
30. van der Valk T, Vezzi F, Ormestad M, Dalén L, Guschanski K. Index hopping on the illumina HiseqX platform and its consequences for ancient DNA studies. Mol Ecol Resour **2020;** 20:1171–81.
31. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Arch Pathol Lab Med **2017;** 141:776–86.
32. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: a new versatile metagenomic assembler. Genome Res **2017**; 27:824–34.
33. Shibata Y, Kitajima N, Kawada JI, et al. Association of cytomegalovirus with infantile hepatitis. Microbiol Immunol **2005**; 49:771–7.
34. Nix WA, Oberste MS, Pallansch MA. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. J Clin Microbiol **2006**; 44:2698–704.
35. Ishiko H, Shimada Y, Yonaha M, et al. Molecular diagnosis of human enteroviruses by phylogeny-based classification by use of the VP4 sequence. J Infect Dis **2002**; 185:744–54.
36. Harvala H, Robertson I, McWilliam Leitch EC, et al. Epidemiology and clinical associations of human parechovirus respiratory infections. J Clin Microbiol **2008**; 46:3446–53.
37. Aizawa Y, Watanabe K, Oishi T, Hirano H, Hasegawa I, Saitoh A. Role of maternal antibodies in infants with severe diseases related to human parechovirus type 3. Emerg Infect Dis **2015;** 21:1966–72.
38. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol **2019;** 37: 907–15.
39. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol **2019;** 20:278.
40. Sun J, Nishiyama T, Shimizu K, Kadota K. TCC: an R package for comparing tag count data with robust normalization strategies. BMC Bioinformatics **2013**; 14: 219.
41. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun **2019**; 10:1523.
42. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics **2003;** 4:2.
43. Fox J. The R commander: a basic-statistics graphical user interface to R. J Stat Softw **2005;** 14:1–42.
44. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics **2011**; 12:77.
45. Koh WM, Bogich T, Siegel K, et al. The epidemiology of hand, foot and mouth disease in Asia: a systematic review and analysis. Pediatr Infect Dis J **2016**; 35: e285–300.
46. Felsenstein S, Yang S, Eubanks N, Sobrera E, Grimm JP, Aldrovandi G. Human parechovirus central nervous system infections in Southern California children. Pediatr Infect Dis J **2014;** 33:e87–91.
47. Barah F, Whiteside S, Batista S, Morris J. Neurological aspects of human parvovirus B19 infection: a systematic review. Rev Med Virol **2014**; 24:154–68.

48. Ramachandran PS, Wilson MR, Catho G, et al. Meningitis caused by the live varicella vaccine virus: metagenomic next generation sequencing, immunology exome sequencing and cytokine multiplex profiling. Viruses **2021**; 13:2286.
49. Mitsuhashi S, Kryukov K, Nakagawa S, et al. A portable system for rapid bacterial composition analysis using a Nanopore-based sequencer and laptop computer. Sci Rep **2017**; 7:5657.
50. Haller O, Staeheli P, Schwemmle M, Kochs G. Mx GTPases: dynamin-like antiviral machines of innate immunity. Trends Microbiol **2015**; 23:154–63.
51. Perng YC, Lenschow DJ. ISG15 In antiviral immunity and beyond. Nat Rev Microbiol **2018;** 16:423–39.
52. Schwartz SL, Park EN, Vachon VK, Danzy S, Lowen AC, Conn GL. Human OAS1 activation is highly dependent on both RNA sequence and context of activating RNA motifs. Nucleic Acids Res **2020;** 48:7520–31.
53. Jheng JR, Lin CY, Horng JT, Lau KS. Inhibition of enterovirus 71 entry by transcription factor XBP1. Biochem Biophys Res Commun **2012**; 420: 882–7.
54. Ramchandar N, Coufal NG, Warden AS, et al. Metagenomic next-generation sequencing for pathogen detection and transcriptomic analysis in pediatric central nervous system infections. Open Forum Infect Dis **2021**; 8:XXX–XX.