Heliyon 9 (2023) e19854

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Untargeted metagenomics protocol for the diagnosis of infection from CSF and tissue from sterile sites

Laura Atkinson [a,*], Jack CD. Lee [a], Alexander Lennon [a], Divya Shah [a], Nathaniel Storey [a], Sofia Morfopoulou [a,b], Kathryn A. Harris [c], Judy Breuer [a,b], Julianne R. Brown [a]

[a] Great Ormond Street Hospital for Children NHS Foundation Trust, Department of Microbiology, Virology and Infection Control, London, UK
[b] Department of Infection, Immunity and Inflammation, Institute of Child Health, University College London, London, UK
[c] Royal London Hospital, Barts Health NHS Trust, Department of Virology, London, UK

A B S T R A C T

Metagenomic next-generation sequencing (mNGS) is an untargeted technique capable of detecting all microbial nucleic acid within a sample. This protocol outlines our wet laboratory method for mNGS of cerebrospinal fluid (CSF) specimens and tissues from sterile sites. We use this method routinely in our clinical service, processing 178 specimens over the past 2.5 years in a laboratory that adheres to ISO:15189 standards. We have successfully used this protocol to diagnose multiple cases of encephalitis and hepatitis.

1. Before you begin

1. CSF for mNGS should be stored at −80 °C within 24 h of collection to minimise RNA degradation. CSF can also be collected into RNA*later* (1:4 RNA*later*:CSF), allowing storage for up to one week at ambient temperature or at +4 °C. Collection without RNA*later* is preferred to avoid diluting the CSF.
2. The optimum specimen collection method for tissue biopsies for mNGS is directly into RNA*later* at the point of collection. The sample is then stable for one week at room temperature, one month at +4 °C or indefinitely at −80 °C. Alternatively, tissue can be collected in a sterile container directly on to dry ice and stored at −80 °C until processing with RNA*later*-ICE.
   a. Fresh tissue in RNA*later* is held at +4 °C overnight to allow tissue penetration. The supernatant can then be removed, and the sample can be stored at −80 °C indefinitely.
   b. Frozen tissue is combined with pre-chilled (at −80 °C) RNA*later*-ICE. The RNA*later*-ICE must remain cold and the tissue must remain frozen, so work quickly on dry ice. Soak tissue in RNA*later* ICE at −20 °C for at least 16 h. The sample is then stable for up to 30 min at room temperature or overnight at +4 °C.
3. 2 × 20 μm or 4 × 10 μm rolled sections of FFPE tissue are required for processing for mNGS, stored at +2–8 °C. DNA and RNA purified from FFPE sections is likely to be highly degraded which may limit the ability to detect low level pathogens.

* Corresponding author.
E-mail address: laura.atkinson@gosh.nhs.uk (L. Atkinson).

https://doi.org/10.1016/j.heliyon.2023.e19854
Received 23 February 2023; Received in revised form 15 August 2023; Accepted 4 September 2023
Available online 9 September 2023

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Cowpox virus DNA (strain Brighton Red) (unquantified) (Stock solution may need diluting to give an approximate Ct value of 26 when tested by targeted real-time PCR. In our hands this is a 10^{-1} dilution).	National Collection of Pathogenic Viruses, operated by UKHSA	0310121v
FCV RNA (unquantfified) (Stock solution may need diluting to give an approximate Ct value of 25 when tested by targeted real-time PCR. In our hands this is a 10^{-3} dilution).	National Collection of Pathogenic Viruses, operated by UKHSA	0101148v
MS2 RNA (unquantified) (Stock solution may need diluting to give an approximate Ct value of 25–26 when tested by targeted real-time PCR. In our hands this is a 10^{-6} dilution)	Roche	10165948001
Biological samples		
FirstChoice Total brain RNA (1 μg/μl)	Life Technologies	AM7962
Human genomic DNA	Promega	G1471
Chemicals, peptides, and recombinant proteins		
Buffer ATL	Qiagen	939011
Buffer AVE	Qiagen	1010371
Buffer PKD	Qiagen	1034963
Deparaffinisiation solution	Qiagen	19093
DTT	Sigma	DTT-RO
PhiX	Illumina	FC-110-3002
Proteinase K	Qiagen	19131
Reagent Dx	Qiagen	19088
RNA*later* Stabilisation solution	Life Technologies	AM7022
RNA*later*-ICE Frozen Tissue Transition Solution	Life Technologies	AM7030
Critical commercial assays		
AllPrep DNA/RNA Mini Kit	Qiagen	80204
EZ1 DNA Blood 350 μl Kit	Qiagen	951054
EZ1 RNA Tissue Mini Kit	Qiagen	959034
EZ1&2 Virus Mini Kit	Qiagen	955134
High sensitivity D1000 reagents	Agilent Technologies	5067–5585
High sensitivity D1000 screen tape	Agilent Technologies	5067–5584
High sensitivity RNA ladder	Agilent Technologies	5067–5581
High sensitivity RNA sample buffer	Agilent Technologies	5067–5580
High sensitivity RNA screen tape	Agilent Technologies	5067–5579
KAPA RNA HyperPrep Kit	Roche	8098093702
KAPA RNA HyperPrep Kit with RiboErase	Roche	8098131702
KAPA Unique Dual Indexed Adaptors	Roche	8861919702
NEBNext Microbiome DNA Enrichment Kit	New England BioLabs	E2612L
NEBNext Multiplex Oligos for Illumina (Index Primers Set 1)	New England BioLabs	E7335L
NEBNext Ultra II FS DNA Library Prep Kit for Illumina	New England BioLabs	E7805L
NextSeq High-Output Kit v2.5 (150 cycles)	Illumina	20024907
NextSeq Mid-Output Kit v2.5 (150 cycles)	Illumina	20024904
Qubit dsDNA HS assay	Invitrogen	Q32854
Qubit RNA HS assay	Invitrogen	Q32852
Other		
EZ1 Advanced XL	Qiagen	9001875
Qubit	Life Technologies	Q33238
4200 TapeStation System	Agilent Technologies	G2991AA
NextSeq 500 System	Illumina	Has been discontinued, 550 system can be purchased, model number: SY-415-1002
0.1 mm silica beads (lysing matrix B)	MP Biomedicals	116911050-CF
1 mm silica beads (lysing matrix C)	MP Biomedicals	116912050-CF
Lysing matrix E	MP Biomedicals	116914050-CF
AMPure XP Beads	Beckman Coulter	A63881
Optical tube 8x caps (TapeStation)	Agilent Technologies	401425
Optical tube 8x strip (TapeStation)	Agilent Technologies	401428
TapeStation loading tips	Agilent Technologies	5067–5599
TapeStation plates	Agilent Technologies	5042–8502

2. Step-by-step method details

2.1. Nucleic acid extraction

Timing: 1 h–4.5 h depending on sample type (see sub-steps 1–3)..

2.1.1. Cerebrospinal fluid (CSF)

Timing:1 h.

a) Add 500 μl of whole CSF to 0.1 mm and 1 mm glass beads. If less than 500 μl is available, top up the volume to 500 μl using DNA-free water.
b) Create a mock positive CSF control as per Table 1. The mock positive will be processed alongside clinical samples from extraction to sequencing to control for contamination, therefore acting as both a positive and negative control.
c) To lyse fungal and bacterial cells bead beat the samples and control for 2 min at 50 Hz, then pulse spin at 13,000 rpm. Transfer 400 μl lysate to a 2 ml sample tube.
d) Make up the carrier RNA/internal positive control (IPC) as per Table 2.
e) Load samples, control and carrier RNA on to a Qiagen EZ1 instrument using a virus mini kit according to the manufacturer's instructions, with a sample volume of 400 μl and an elution volume of 60 μl.
f) Assess the RNA quantity (ng/μl) using the Qubit RNA HS assay, and DNA quantity (ng/μl) using the Qubit dsDNA DNA HS assay. Proceed with library preparation even if yield is below the Qubit limit of detection
g) Store purified nucleic acid at −80 °C until library preparation.

2.1.2. Fresh/frozen tissue

Timing: 1.5 h.

a) Remove tissue from RNA*later* or RNA*later*-ICE.
b) Place the tissue on a sterile petri dish for weighing. A maximum of 20 mg of tissue is processed for extraction. If the tissue is more than 20 mg, cut to 20 mg using a sterile disposable scalpel and return excess to storage at −80 °C. If the tissue is less than 7 mg process as one piece, if 7–14 mg cut into two pieces and if 15–20 mg cut into three pieces.
c) Prepare RLT Plus buffer as per Table 3.
d) Add tissue (one, two or three pieces depending on amount received) to a single tube of MP Biomedicals Lysing matrix E containing 4 mm, 1.4 mm and 0.1 mm beads with 600 μl of RLT Plus buffer.
e) Create a positive control as per Table 4. This control should be made into a tube of Lysing matrix E. The mock positive will be processed alongside clinical samples from extraction to sequencing to control for contamination, therefore acting as both a positive and negative control.
f) To lyse the tissue, fungal and bacterial cells, bead beat samples and control for 2 min at 50 Hz. If the tissue is not completely lysed, repeat this step.
g) Centrifuge the lysate for 3 min at 13,000 rpm and transfer the supernatant to an AllPrep DNA spin column and proceed according to manufacturers instructions. Use a new collection tube after every centrifugation step to minimise the risk of cross-contamination.
h) Add 20 μl of MS2 RNA to the RNA flow-through from the DNA column
i) Elute the RNA in 30 μl of RNase-free water and DNA in 50 μl of buffer EB
j) Assess RNA quality (RIN score and DV200) using the TapeStation High Sensitivity RNA assay.

NOTE: RIN score is the RNA Integrity Number and DV200 is the percentage of RNA with a size profile greater than 200 nt..

k) Assess the RNA quantity (ng/μl) using the Qubit RNA HS assay, and DNA quantity (ng/μl) using the Qubit dsDNA DNA HS assay. Record the RIN score from the TapeStation analysis, as this is used to determine the fragmentation conditions during library preparation.
l) Store purified RNA at −80 °C and DNA at +4 °C until library preparation.

Table 1
Preparation of mock positive control for nucleic acid extraction from CSF.

Reagent	Amount (μl)
DNA-free water	414
Total Brain RNA (0.001 μg/μl)	3.75
Human genomic DNA (100 ng/μl)	7.6
Cowpox DNA	37.5
FCV RNA	37.5

Table 2
Carrier RNA/Internal Positive Control for nucleic acid extraction from CSF. This controls for successful nucleic acid extraction, library preparation and sequencing. Volumes given are for one reaction.

Reagent	Amount (μl)
Buffer AVE	37
Total Brain RNA (0.01 μg/μl)	3
MS2 RNA	20

Table 3
Preparation of RLT Plus buffer for nucleic acid extraction from fresh tissue in RNA*later* and frozen tissue. Volumes given are for one reaction.

Reagent	Amount (μl)
Buffer RLT Plus	1000
DTT (2 M)	20
Reagent Dx	5

2.1.3. Formalin-fixed paraffin embedded tissue (FFPE)
Timing: 4.5 h.

2.1.3.1. Deparaffinization.

a) If the tissue is clearly visible, cut away excess wax from the ends of the rolled sections. If the tissue is not clearly visible, leave rolled sections as they are.

b) Add 640 μl of deparaffinization solution to the rolled sections in a 2 ml tube.

c) Vortex vigorously for 10 s, then pulse centrifuge.

d) Incubate at 56 °C for 3 min then cool to room temperature

e) Centrifuge at 13,000 rpm for 2 min (mark the outside of the tube to indicate which side the pellet will be on)

f) Remove and discard supernatant and any residual deparaffinization solution. Do not disturb the pellet – if the pellet is not visible, aspirate from the opposite side to where the pellet should be. If excess wax remains, remove the supernatant and repeat steps ii to vi.

g) Incubate at 37 °C for 10 min with the lid open to dry the pellet.

h) Re-suspend the pellet in 150 μl of buffer PKD.

i) Create an RNA positive control as per Table 5. The mock positive will be processed alongside clinical samples from extraction to sequencing to control for contamination, therefore acting as both a positive and negative control.

j) Add 10 μl of proteinase K, vortex and incubate at 56 °C for 15 min then incubate on ice for 3 min, until samples have cooled completely.

k) Centrifuge for 15 min at 20,000 g.

l) Transfer supernatant to a new tube for RNA extraction, and store pellet for DNA extraction.
NOTE: The pellet can be stored at room temperature for 2 h, for one day at +4 °C or longer at −20 °C.

2.1.3.2. Purification of total RNA.

m) Incubate supernatant at 80 °C for exactly 15 min then pulse centrifuge

Table 4
Preparation of mock positive control for nucleic acid extraction from fresh/frozen tissue. Preparation of a positive control for nucleic acid extraction from fresh tissue (in RNA*later* or frozen).

Reagent	Amount (μl)
RLT Plus/DTT/Reagent Dx	460
Total Brain RNA (1 μg/μl)	3
Human genomic DNA (100 ng/μl)	76
Cowpox DNA	37.5
FCV RNA	30

Table 5
Preparation of mock positive RNA control for nucleic acid extraction from formalin fixed paraffin embedded (FFPE) tissue.

Reagent	Amount (μl)
PKD	150
Total Brain RNA (1 μg/μl)	3
FCV RNA	30
MS2 RNA	20

n) Add 200 μl of buffer RLT and 1 mm and 0.1 mm glass beads. Mix by vortexing.

o) To lyse fungal and bacterial cells, bead beat for 2 min at 50 Hz, then pulse centrifuge.

p) Load 300 μl of lysate on the Qiagen EZ1 RNA Tissue Mini kit using the EZ1 Advanced XL RNA card. Use the total RNA, RNA mini tissue programme without DNAse, with a 50 μl elution volume.

2.1.3.3. Purification of total DNA.

q) Bring the pellet to room temperature, and re-suspend in 180 μl of buffer ATL.

r) Add 40 μl of proteinase K, vortex, then incubate at 56 °C for 1 h followed by 90 °C for 1 h.

s) Add 250 μl of buffer AL and vortex.

t) Prepare a DNA positive control as per Table 6. The mock positive will be processed alongside clinical samples from extraction to sequencing to control for contamination, therefore acting as both a positive and negative control.

u) Add 1 mm and 0.1 mm glass beads, and bead beat for 2 min at 50 Hz to lyse bacterial and fungal cells. Pulse spin.

v) Load samples and control on to EZ1 DNA blood kit (without ethanol wash) using the 350 μl protocol and 50 μl elution volume.

w) Assess RNA quality (RIN score and DV200) using the TapeStation High Sensitivity RNA assay.

x) Assess the RNA quantity (ng/μl) using the Qubit RNA HS assay, and DNA quantity (ng/μl) using the Qubit dsDNA DNA HS assay. Record the RIN score from the TapeStation analysis, as this is used to determine the fragmentation conditions during library preparation.

y) Store purified RNA at −80 °C and DNA at +4 °C until library preparation.

Table 6
Preparation of mock positive DNA control for nucleic acid extraction from formalin fixed paraffin embedded (FFPE) tissue.

Reagent	Amount (μl)
Buffer AL	355
Human genomic DNA (100 ng/μl)	76
Cowpox DNA	37.5

2.2. Library preparation for RNA sequencing

Timing: 6.5 h if including ribosomal RNA depletion (for tissue samples including FFPE), or 4 h if performing RNA library prep without ribosomal RNA depletion (CSF).

2.2.1. Ribosomal RNA depletion

Ribosomal RNA depletion is performed using the KAPA RNA HyperPrep Kit with RiboErase (Roche) as per the manufacturer's instructions. This step is performed on samples with an RNA input between 25 ng and 1 μg. These samples are typically tissue (including FFPE) samples only. For CSF samples, proceed straight to RNA Library Prep without ribodepletion.

2.2.2. RNA library prep

RNA library preparation is performed using the KAPA RNA HyperPrep Kit (Roche) as per the manufacturer's instructions, with the following modifications:

a) Use DNA-free water (Molzyme) for steps prior to adapter ligation to minimise microbial DNA contamination
b) RNA fragmentation conditions are decided using the RIN score as per Table 7.
c) KAPA unique dual-indexed adaptors are diluted to 1 μM for CSFs or 1.5 μM for tissues.
d) The number of cycles in the denaturation and annealing/extension step of the library amplification PCR is dependent on the total RNA input (ng) of each sample. This is summarised in Table 8. The number of cycles for CSF samples is capped at a maximum of 12 cycles, regardless of RNA input, to minimise the over-amplification of contaminating DNA.

2.2.3. RNA library quality control (QC)

a) Quantify the amplified libraries using 2 μl library (measured with a P2 pipette) with the Qubit dsDNA HS Assay Kit as per the manufacturer's instructions.
b) Check the size profile of the amplified libraries using the TapeStation High Sensitivity D1000 ScreenTape as per the manufacturer's instructions. Use 'add region' to determine the average library size in base pairs. The average library size is expected to be approximately 300 base pairs. **NOTE:** a peak at 120–140 base pairs indicates the presence of adapter-dimers. If this is seen, perform a second 1X post-amplification clean-up and repeat the QC.

2.3. Library preparation for DNA sequencing

Timing: 4 h if including microbial DNA enrichment (for tissue samples including FFPE), or 2.5 h if performing DNA library prep without microbial DNA enrichment (CSF).

2.3.1. Microbial DNA enrichment

NOTE: This step is performed on tissue and FFPE samples only. CSFs sample do not undergo any human DNA depletion due to the low amount of host genome present in samples; in our experience host depletion of CSFs results in a poor quality final library due to insufficient DNA for the library preparations. For CSF samples, proceed straight to DNA Library Prep unless DNA yield is > 1 ng/μl.

Microbial DNA Enrichment is performed using the NEBNext® Microbiome DNA Enrichment Kit as per the manufacturer's instructions, with the following modifications:

a) Use 1 μl MBD2-Fc-bound magnetic beads for every 3.125 ng of input DNA (this is double the manufacturer recommended quantity of MBD2-Fc-bound magnetic beads).
b) Use DNA-free water (Molzyme) throughout to minimise contamination with microbial DNA

NOTE: The DNA quantity (ng/μl) is re-assessed prior to DNA library preparation using the Qubit dsDNA DNA HS assay.

Table 7
Conditions for the fragmentation of RNA for RNA library preparation. RNA fragmentation conditions are decided using the RIN score which is determined through TapeStation analysis (see steps 2, j and 3, x).

Input RNA type	RIN score	Fragmentation
Intact	≥5	8 min at 94 °C
Partially degraded	5 and all CSFs	6 min at 85 °C
	4	3 min at 85 °C
	3	1 min at 85 °C
Degraded (inc. FFPE)	≤3	1 min at 65 °C

Table 8
Summary of the number of PCR cycles required for the denaturation and annealing/extension step of RNA library amplification. The number of RNA library amplification PCR cycles is dependent on the initial RNA input (ng). The number of cycles for CSF samples is capped at a maximum of 12 cycles, regardless of RNA input, to minimise the over-amplification of contaminating DNA.

RNA input (ng)	Number of cycles
25 < 75	14
75 < 100	13
100 < 150	12
150 < 200	11
200 < 250	10
250 < 500	9
500 < 750	8
750 < 1000	7
1000	6

2.3.2. DNA library prep

DNA library preparation is performed using the NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina® as per the manufacturer's instructions, with the following modifications:

a) Use DNA-free water (Molzyme) for steps prior to adapter ligation
b) DNA fragmentation is performed at 37 °C for 20 min, followed by 30 min at 65 °C, with a target fragment size of 150–350 bp
c) For samples with input >100 ng, size selection is performed during bead clean-up for a target insert size distribution of 150–250 bp. This is achieved with a 0.4X ratio of sample purification beads to sample (e.g. 40 μl beads to 100 μl sample) followed by a 0.2X ratio of sample purification beads to supernatant.
d) The number of cycles in the denaturation and annealing/extension step of the library amplification PCR is dependent on the total DNA input (ng) of each sample. This is summarised in Table 9. The cycle number is capped at maximum 8 cycles, regardless of DNA input, to avoid over-amplification of contaminating sequences.

2.3.3. DNA library quality control (QC)

a) Quantify the amplified libraries using the Qubit dsDNA HS Assay Kit as per the manufacturer's instructions using 2 μl library (measured with a P2 pipette).
b) Check the size profile of the amplified libraries using the TapeStation High Sensitivity D1000 ScreenTape as per the manufacturer's instructions. Use 'add region' to determine the average library size in base pairs. **NOTE:** A peak at around 80 base pairs indicates primers, and a peak at around 128 base pairs indicates adapter-dimer. If either of these are seen, bring up the sample volume to 50 μl with 0.1X TE buffer and repeat the clean-up of the library amplification PCR reaction, then repeat the QC.

2.4. Library denature and dilute

Timing: 45 min.

Libraries are pooled according to Table 10 with a final pool concentration of 4 nM. Pooled libraries are diluted and denatured prior to sequencing using the Illumina NextSeq 500/550 kit as per manufacturer's instructions.

Table 9
Summary of the number of PCR cycles required for the denaturation and annealing/extension step of DNA library amplification. The number of DNA library amplification PCR cycles is dependent on the initial DNA input (ng). The cycle number is capped at maximum 8 cycles to avoid over-amplification of contaminating sequences.

DNA input (ng)	Number of cycles
>0.5 ≤ 1	8
>1 ≤ 5	7–8
>5 ≤ 10	6–7
>10 ≤ 50	4–5
>50 ≤ 200	3–4
>200 ≤ 500	3

Table 10
Desired share of reads for pooling of libraries in preparation for sequencing on the Illumina NextSeq platform. Pooled libraries include RNA and DNA libraries for each sample and their respective control.

	NextSeq Mid Output kit		NextSeq High Output kit	
Sample type	Minimum proportion of run assigned to each library	Maximum number of libraries per run	Minimum proportion of run assigned to each library	Maximum number of libraries per run
Tissue	25%	4	8.30%	12
CSF	12.50%	8	4.17%	24

2.5. Loading and sequencing on the NextSeq

Timing: 30 min to load the NextSeq, 15 h for the sequencing run.

Samples are sequenced on the Illumina NextSeq instrument, using a NextSeq 500/550 Mid or High Output Kit v2.5, 150 cycles as per the manufacturer's instructions. Load the NextSeq 500 instrument as per the manufacturer's instructions, using a 76 base-pair paired end run with Mid or High Output mode.

2.6. Analysis

Timing: Variable depending on sample complexity, approximately 3–4 h per sample.

Following demultiplexing, analyse data using analysis protocols that adhere to recommendations for clinical metagenomics [1]. We recommend the use of metaMix [2,3] which has previously been shown to perform with superior sensitivity and specificity [4].

A detailed protocol for the bioinformatics analysis is beyond the scope of this wet lab protocol, however we recommend that for results to be valid, MS2 internal positive control must be detected in all RNA libraries and Cowpox and Feline Calicivirus must be detected in the positive control. Any organisms identified by metaMix are confirmed by mapping non-human reads to a relevant reference genome. For a positive result to be valid, the following criteria must be fulfilled:

- >10 reads mapping to viral genome [5].
  - o If a DNA organism is identified in tissue, consider results with <5 reads. Due to the low number of reads, confirm or refute this result with targeted PCR
- Reads distributed across genome
- ≥3 non-overlapping regions > read length [6].
- Viruses only: detected in clinical specimen must be absent in the controls
- Bacteria only: reads per million (RPM) ratio ≥10 (*RPM ratio = RPM sample RPM negative control*) [6].

3. Expected Outcomes

We present four clinical cases which were processed in our laboratory using the protocol we describe. These are representative of those seen in our clinical service and are not an exhaustive description of all samples processed using this protocol; instead, they serve to illustrate common expected outcomes in the wet lab.

Patient A is a one-year-old with severe combined immunodeficiency who presented with symptoms of neurodevelopmental regression. A brain biopsy collected into RNA*later* was sent to us for mNGS. Post RNA extraction QC demonstrated the isolation of high-quality RNA from this sample. This is shown in Fig. 1A, alongside examples of partially degraded RNA (Fig. 1B) and highly degraded RNA (Fig. 1C) for comparison. The results of post-amplification QC are summarised in Table 11 and Fig. 2A. The specimen gave a high library yield of 8.53 ng/μL with an average library size in the expected region and no adapter dimer present. Sequence analysis identified 4,860,141 reads of Astrovirus VA1, which was confirmed by targeted real-time PCR.

Patient B is a 64-year-old presenting with neurological symptoms. A neat CSF sample was sent for mNGS and underwent RNA and DNA sequencing. The results of the DNA library post-amplification QC are summarised in Table 11 and Fig. 2B. The specimen gave a library yield of 2.67 ng/uL with an average library size in the expected region and no adapter dimer present. Sequence analysis identified 2048 reads of torque teno virus.

Patient C is a 15-year-old with acute onset seizures and ongoing cognitive and behavioural difficulties. A neat CSF sample was collected and sent to us for mNGS analysis. The results of the RNA library post-amplification QC revealed adapter dimer, indicated by an additional peak at 128 base pairs (Fig. 3A). To rectify this, an additional 1X post-amplification bead clean-up was performed and the QC was repeated. Repeat QC confirmed the removal of all adapter dimer (Fig. 3B). Sequence analysis did not detect any pathogens.

Patient D is an 83-year-old with altered behaviour, and MRI results in keeping with possible encephalitis. A neat CSF sample was collected and sent to us for mNGS analysis. The sample was initially stored at +4 °C for six days, prior to shipping on dry ice and storage at −80 °C. The results of the RNA library post-amplification QC showed a very low library yield (Table 11). Analysis by TapeStation showed no detectable sequencing library (Fig. 4). We were therefore unable to perform RNAseq on this sample, due to poor RNA quality which was likely a result of improper storage prior to processing.

We assessed the range of detection compared to targeted real-time PCR for model organisms using DNA or RNA virus spiked into control material, as described in Supplementary Data. The detection range for DNA and RNA viruses in CSF samples is comparable to

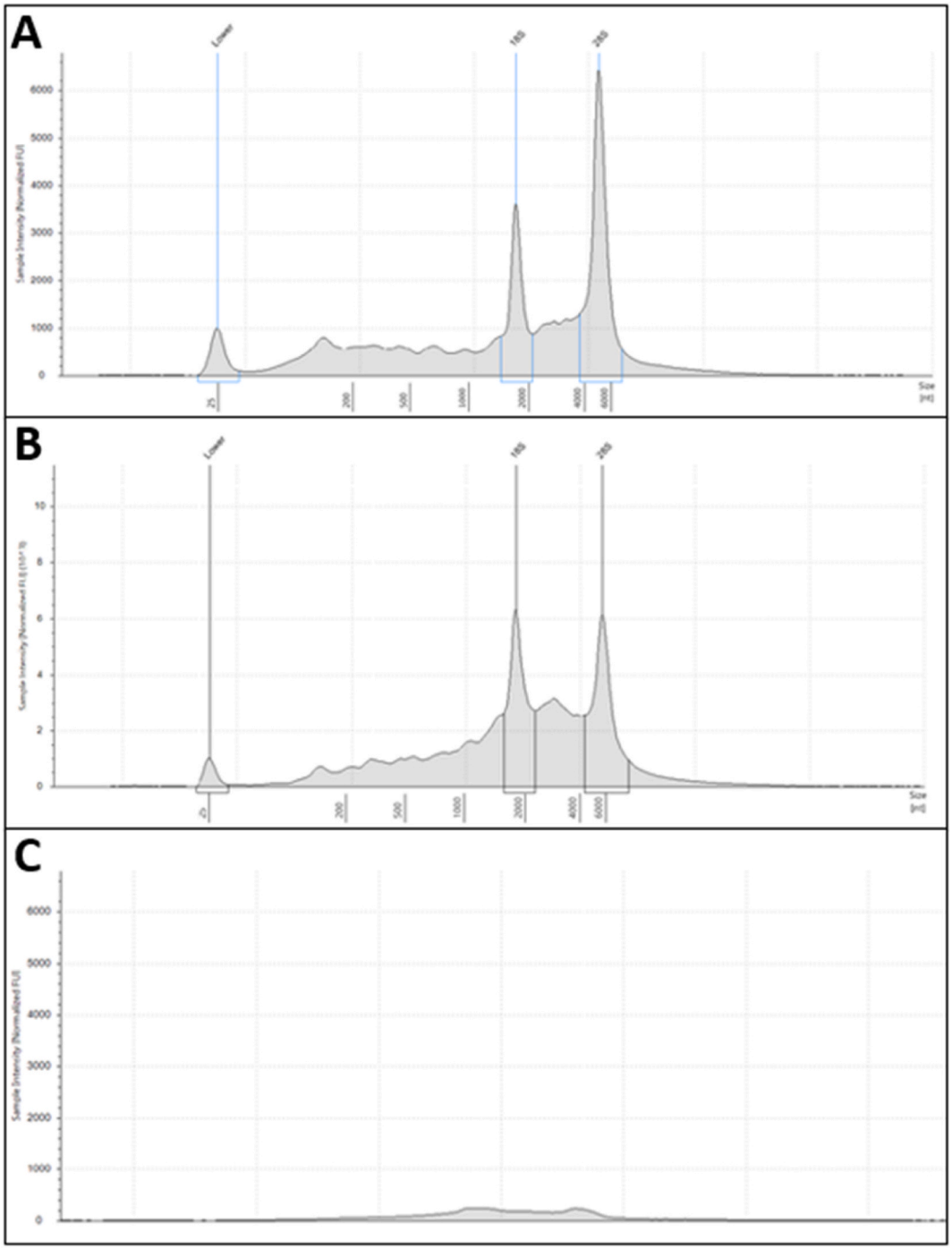

Fig. 1. RNA extraction TapeStation results showing high quality, partially degraded and highly degraded RNA. (A) Post RNA extraction QC of the brain biopsy collected from patient A. Distinct peaks for 18S and 28S rRNA indicate fully intact RNA. (B) An example of post RNA extraction QC of a sample with partially degraded RNA. Distinct peaks can be seen for the 18S and 28S rRNA, but with a greater proportion of RNA in the 200–1000 nucleotide range and decreasing height of the 28S rRNA peak. (C) An example of post extraction QC of a sample with highly degraded RNA. No discernible peaks can be seen for the 18S or 28S rRNA.

Table 11
RIN score (post RNA extraction), post-amplification quality control and sequencing results for patients A-D. RIN score, post library amplification quality control, molarity and sequencing results for patients A-D, whose samples were processed by our metagenomics service between 2018 and 2021. RIN score is determined using the TapeStation High Sensitivity RNA assay and is assessed after RNA extraction for tissues only. N/A = not applicable.

Patient	Sample	RIN score	Library yield (ng/uL)	Library size (base pairs)	Molarity (nM)	Result
A	Brain biopsy in RNALater	7.9	8.53	334	38.6	4,860,141 reads of Astrovirus VA1 detected, confirmed by PCR (Ct 31.2).
B	Neat CSF	N/A	2.67	303	13.5	2048 reads of torque teno virus detected.
C	Neat CSF	N/A	0.92	441	3.14	No pathogens detected.
D	Neat CSF	N/A	0.12	487	0.37	Unable to perform RNAseq, no RNA recovered.

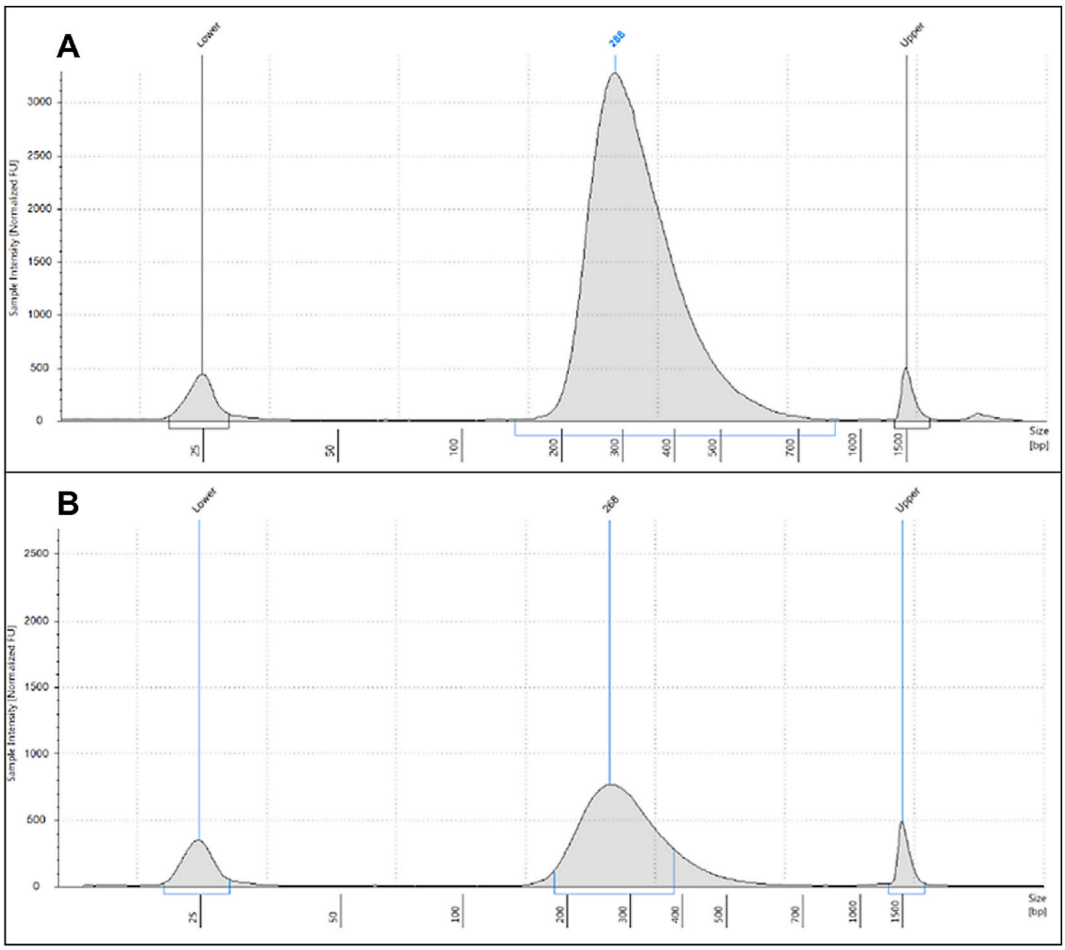

Fig. 2. Post library amplification TapeStation results for patients A and B showing successful library preparation. (A) RNA library of a brain biopsy from patient A collected into RNA*later*. The specimen gave a high RNA library yield and an average library size in the expected region of approximately 300 base pairs. No adapter dimer is present in the sample. (B) DNA library of a neat CSF sample collected from patient B. The specimen gave a high DNA library yield and an average library size in the expected region of approximately 300 base pairs. No adapter dimer is present in the sample.

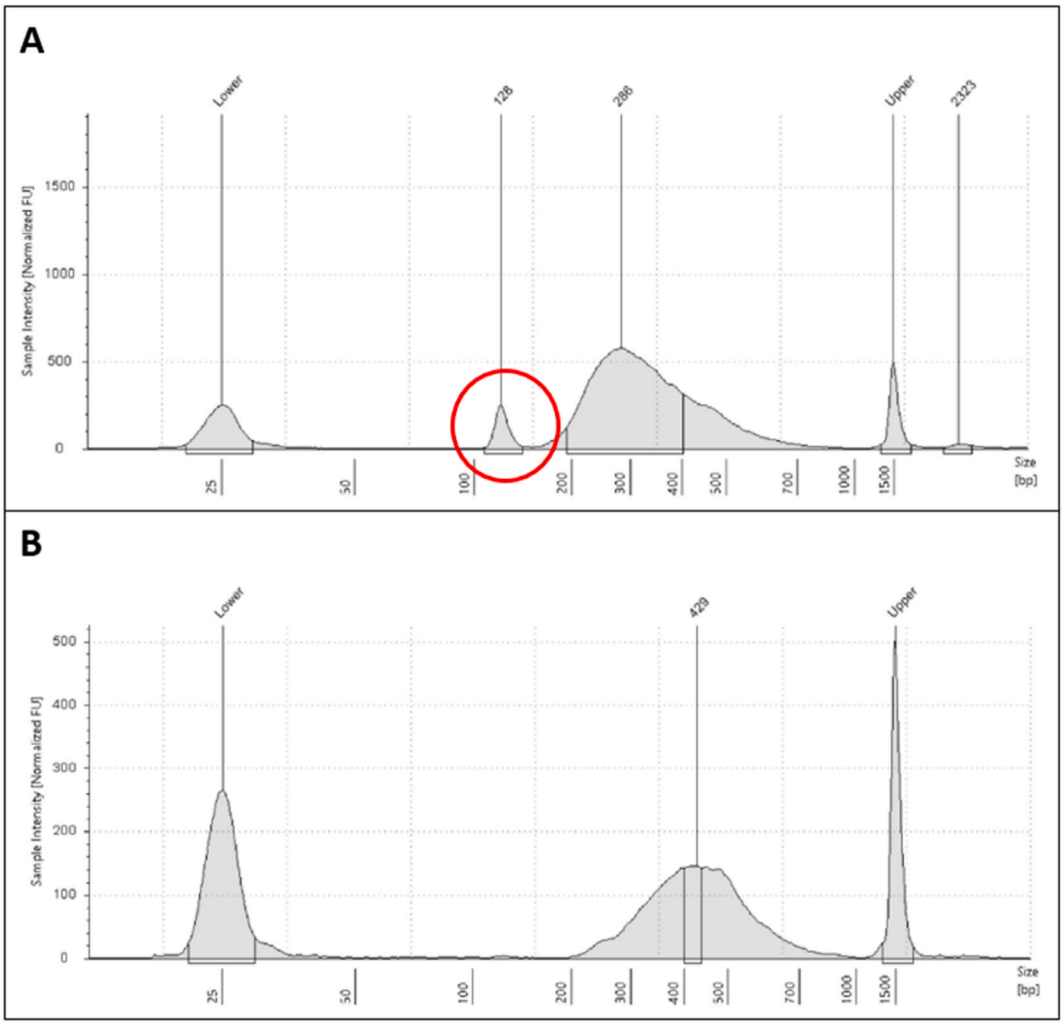

Fig. 3. Post RNA library amplification TapeStation results for patient C, indicating the presence and subsequent removal of adapter dimer. (A) RNA library of a neat CSF. A peak representing adapter-dimer is seen at a size of 128 base pairs (circled in red). An additional 1X post-amplification bead clean-up was performed to rectify this. (B) Repeat TapeStation analysis of the RNA library after an additional 1X bead clean-up shows there is no longer adapter dimer present in the library.

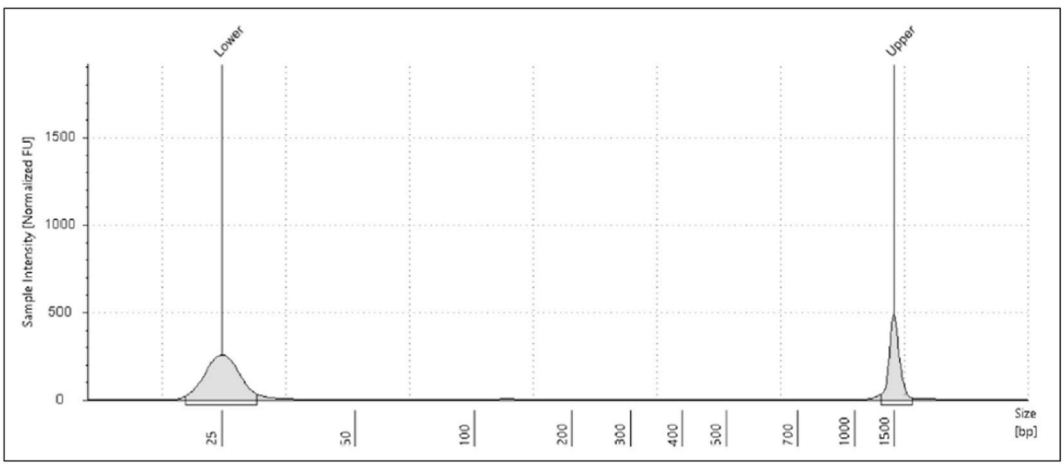

Fig. 4. Post library amplification TapeStation results for patient D, showing unsuccessful library preparation. RNA library of a neat CSF which was stored incorrectly prior to processing. No detectable sequencing library is present, therefore we were unable to proceed with RNA sequencing.

targeted real-time PCR. In tissues, the detection range for RNA viruses is also comparable to targeted real-time PCR, however for DNA viruses in tissue the detection range is approximately 100-fold less using metagenomics. The specificity of our mNGS protocol was confirmed by testing residual known positive specimens (see Supplementary Data).

4. Limitations

Metagenomic next-generation sequencing (mNGS) is an untargeted technique capable of detecting all microbial DNA and RNA within a clinical sample, allowing definitive, unbiased detection of pathogens[7]

The detection limits in CSF are comparable to targeted real-time PCR. Nonetheless where there is strong suspicion of a particular organism causing infection it is still worthwhile performing targeted real-time PCR in parallel to metagenomics due to the faster turn-around-time of PCR. In tissues, we postulate the reduced detection limit for DNA organisms compared to targeted real-time PCR is due to insufficient depletion of host DNA prior to sequencing, therefore there is a need for better human DNA depletion methods to maximise sensitivity. For this reason we recommend interrogating the metagenomics results in tissue for viruses with <5 reads, which should be confirmed with a targeted PCR prior to clinical reporting. It is worthwhile noting that whilst the sensitivity of metagenomics in tissue for DNA organisms is suboptimal, it is equivalent to the sensitivity of broad-range bacterial PCR widely used in clinical practice [8] and has been used successfully by our laboratory to diagnose multiple case of encephalitis [9–14] and hepatitis [2,15].

Our protocol depends on a high degree of specimen integrity to preserve nucleic acid in clinical samples, particularly RNA. Improper storage of samples prior to extraction results in being unable to proceed with RNA sequencing due to little or no RNA being recovered post library amplification (see Fig. 4 in Expected Outcomes) and poor quality DNA will reduce the effectiveness of host DNA depletion. To address this, we have outlined specific specimen requirements in our protocol which maintain the integrity of nucleic acid (see Before you begin, steps 1–3).

5. Troubleshooting

Problem 1

No RNA and/or DNA recovered after library amplification.

Potential solution to Problem 1

Ensure specimens are collected according to the instructions specified in Before you begin, steps 1–3.

Problem 2

Post-extraction QC of FFPE tissue shows highly degraded RNA.

Potential solution to Problem 2

- Formalin fixation is known to degrade nucleic acid, therefore fresh tissue specimens are preferred where possible
- If fresh tissue is not available, proceed with the RNA library prep amending the RNA fragmentation conditions as per the RIN score obtained from post-extraction QC (see Table 8 in Library Preparation for RNA Sequencing).

Problem 3

An organism has been detected in both clinical specimen and mock positive control.

Potential solution to Problem 3

There is a risk of contamination associated with this protocol due to the broad range nature of the assay. The method is theoretically able to detect any organism present in a clinical sample, therefore contamination of reagents, equipment or the laboratory workspace risks detecting these organisms in samples and interfering with the analysis and interpretation of sequencing results or leading to false positive results. To address this, we recommend cleaning the workspace and equipment thoroughly between sequencing runs. The laboratory should also be organised into a pre-PCR area for nucleic acid extraction and library preparation, and a post-PCR area for library amplification, pooling and sequencing [16]. The use of a control sample processed in parallel to all clinical specimens mitigates the risk of false positive results, as gross contamination of the laboratory or reagents is expected to also contaminate the control (see Nucleic Acid Extraction for preparation of controls). If contamination is suspected, mNGS of the affected specimen should be repeated from extraction.

Problem 4

Results of mNGS vary depending on the bioinformatics pipeline used.

Potential solution to Problem 4

The method chosen for data analysis is critical and can greatly influence the final result [1,17]. In our clinical service we use metaMix [2,3] as it has previously been shown to have superior performance for both sensitivity and specificity [4]. Further discussion of analysis methods and approaches is beyond the scope of this wet lab protocol.

6. Resource availability

6.1. Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Laura Atkinson (laura.atkinson@gosh.nhs.uk).

6.2. Materials availability

This study did not generate new unique reagents.

6.3. Data and code availability

There are restrictions to the availability of data generated by this protocol due to patient confidentiality. The R package for our recommended analysis method, metaMix, has been previously described [2,3].

Author contribution statement

Laura Atkinson: Performed the experiments; Wrote the paper.
Jack CD Lee; Alexander Lennon; Divya Shah: Performed the experiments.
Kathryn A Harris; Judy Breuer: Conceived and designed the experiments; Analyzed and interpreted the data.
Julianne R Brown: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
Nathaniel Storey: Analyzed and interpreted the data.
Sofia Morfopoulou: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The development, optimisation and validation of the protocol described was supported by a paediatric research grant from the Great Ormond Street Hospital Children's Charity ("Diagnosis of encephalitis by deep sequencing", V4317).
All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre.
The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

[1] de Vries JJC, J.R. Brown, N. Couto, M. Beer, P. Le Mercier, I. Sidorov, A. Papa, N. Fischer, B.B. Oude Munnink, C. Rodriquez, M. Zaheri, A. Sayiner, M. Hönemann, A. Perez Cataluna, E.C. Carbo, C. Bachofen, J. Kubacki, D. Schmitz, K. Tsioka, S. Matamoros, D. Höper, M. Hernandez, E. Puchhammer-Stöckl, A. Lebrand, M. Huber, P. Simmonds, C.J. Claas ECJ, López-labrador FX on behalf of the ESCV network on next-generation sequencing, Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting, Journal of Clinical Virology 138 (2021), 104812.
[2] S. Morfopoulou, , et al.J.R. Brown, J. Breuer, Genomic investigations of acute hepatitis of unknown aetiology in children, Nature 617 (2022) 564–573.
[3] S. Morfopoulou, V. Plagnol, Bayesian mixture analysis for metagenomic community profiling, Bioinformatics 31 (18) (2015) 2930–2938.
[4] de Vries JJC, J.R. Brown, N. Fischer, I.A. Sidorov, S. Morfopoulou, J. Huang, B.B. Oude Munnink, A. Sayiner, A. Bulgurcu, C. Rodriguez, G. Gricourt, E. Keyaerts, L. Beller, C. Bachofen, J. Kubacki, C. Samuel, L. Florian, S. Dennis, M. Beer, Dl Hoeper, M. Huber, V. Kufner, M. Zaheri, A. Lebrand, A. Papa, S. van Boheemen, A. C.M. Kroes, J. Breuer, F.X. Lopez-Labrador, E.C.J. Claas, Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples, J. Clin. Virol. 141 (2021), 104908.
[5] J. Kawada, Y. Okuno, Y. Torii, R. Okada, S. Hayano, A. Ando, Y. Kamiya, A. Kojima, Y. Ito, Identification of viruses in cases of pediatric acute encephalitis and encephalopathy using next-generation sequencing, Sci. Rep. 6 (2016), 33452.
[6] K. Mongkolrattanothai, S.N. Naccache, J.M. Bender, E. Samayoa, E. Pham, G. Yu, J. Dien Bard, S. Miller, F.G. Aldrovandi, C.Y. Chiu, Neurobrucellosis: unexpected answer from metagenomics next generation sequencing, J. Pediatr. Infect. Dis. Soc. 6 (4) (2017) 393–398.
[7] S.M. Prashanth, M.R. Wilson, Metagenomics for neurological infections – expanding our imagination, Nat. Rev. Neurol. 16 (2020) 547–556.

[8] K.A. Harris, J.R. Brown, Diagnostic yield of broad-range 16S rRNA gene PCR varies by sample type and is improved by the addition of qPCR panels targeting the most common causative organisms, J. Med. Microbiol. 71 (12) (2022).
[9] J.R. Brown, S. Morfopoulou, J. Hubb, W.A. Emmett, W. Ip, D. Shah, T. Brooks, S.M.L. Paine, G. Anderson, A. Virasami, C.Y.W. Tong, D.A. Clark, V. Plagnol, T. S. Jacques, W. Qasim, M. Hubank, J. Breuer, Astrovirus VA1/HMO-C: an increasingly recognised neurotropic pathogen in immunocompromised patients, Clin. Infect. Dis. 60 (6) (2015) 881–888.
[10] J.A.D. Duncan, S.M.B. Mohamad, D.F. Young, A.J. Skelton, T.R. Leahy, D.C. Munday, K.M. Butler, S. Morfopoulou, J.R. Brown, M. Hubank, J. Connell, P. J. Gavin, C. McMahon, E. Dempsey, N.E. Lynch, T.S. Jacques, M. Valappil, A.J. Cant, J. Beuer, K.R. Engelhardt, R.E. Randall, S. Hambleton, Human IFNAR2 deficiency: lessons for antiviral immunity, Sci. Transl. Methods 7 (307) (2015), 307RA154.
[11] S.H. Lum, A. Turner, M. Guiver, D. Bonney, T. Martland, E. Davies, M. Newbould, J.R. Brown, S. Morfopoulou, J. Breuer, R. Wynn, An emerging opportunistic infection: fatal astrovirus (VA1/HMO-C) encephalitis in a pediatric stem cell transplant recipient, Transpl. Infect. Dis. 18 (6) (2016) 960–964.
[12] S. Morfopoulou, E.T. Mee, S.M. Connaughton, J.R. Brown, K. Gilmour, W.K. Chong, W.P. Duprex, D. Ferguson, M. Hubank, C. Hutchinson, M. Kaliakatsos, S. McQuaid, S. Paine, V. Plagnol, C. Ruis, A. Virasami, H. Zhan, T.S. Jacques, S. Schepelmann, W. Qasim, J. Breuer, Deep sequencing reveals persistence of cell-associated mumps vaccine virus in chronic encephalitis, Acta Neuropathol. 133 (1) (2016) 139–147.
[13] S. Morfopoulou, J.R. Brown, E.G. Davies, G. Anderson, A. Virasami, W. Qasim, W.K. Chong, M. Hubank, V. Plagnol, M. Desforges, T.S. Jacques, P.J. Talbot, J. Breuer, Coronavirus HCoV-OC43 associated with fatal encephalitis, N. Engl. J. Med. 375 (5) (2016) 497–498.
[14] Penner J. et al., Breuer J and Morfopoulou S (2023). Translating metagenomics into clinical practice for complex paediatric neurological presentations, *Journal of Infection*, In Press.
[15] G. Bucciol, L. Moens, K. Payne, E. Wollants, D. Mekahli, E. Levtchenko, F. Vermeulen, T. Tousseyn, P. Gray, C.S. Ma, S.G. Tangye, M. Van Ranst, J.R. Brown, J. Breuer, I. Meyts, Chronic Aichi virus infection in a patient with X-linked Agammaglobulinemia, J. Clin. Immunol. 38 (7) (2018) 748–752.
[16] F.X. López-Labrador, J.R. Brown, N. Fischer, H. Harvala, S. Van Boheemen, O. Cinek, A. Sayiner, T.V. Madsen, E. Auvinen, V. Kufner, M. Huber, C. Rodriguez, M. Jonges, M. Hönemann, P. Susi, H. Sousa, P.E. Klapper, A. Pérez-Cataluña, M. Hernandez, R. Molenkamp, L. van der Hoek, R. Schuurman, N. Couto, K. Leuzinger, P. Simmonds, M. Beer, D. Höper, S. Kamminga, M.C.W. Feltkamp, J. Rodríguez-Díaz, E. Keyaerts, X.C. Nielsen, E. Puchhammer-Stöckl, A.C. M. Kroes, J. Buesa, J. Breuer, E.C.J. Claas, de Vries JJC, Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: wet lab procedure, J. Clin. Virol. 134 (2021), 104691.
[17] T. Bharucha, C. Oeser, F. Balloux, J.R. Brown, E.C. Carbo, A. Charlett, C.Y. Chiu, E.C.J. Claas, M.C. de Goffau, de Vries JJC, M. Eloit, S. Hopkins, J.F. Huggett, D. MacCannell, S. Morfopoulou, A. Nath, D.M. O'Sullivan, L.B. Reoma, L.P. Shaw, I. Sidorov, P.J. Simner, L. Van Tan, E.C. Thomson, L. van Dorp, M.R. Wilson, J. Breuer, N. Field, STROBE-metagenomics: a STROBE extension statement to guide the reporting of metagenomics studies, Lancet Infect. Dis. 20 (10) (2020) e251–e260.