Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development of an Agnostic Clinical Metagenomic Workflow for Respiratory RNA Viruses

Kate E. Dingle

kate.dingle@ndm.ox.ac.uk

University of Oxford

Katie M.V. Hopkins
University of Oxford

Nicholas D. Sanderson
University of Oxford

Ali Vaughan
University of Oxford

Daisy Anderton
University of Oxford

Matthew Colpus
University of Oxford

Melody Parker
University of Oxford

Syeda Anam Fatima
Oxford University Hospitals NHS Foundation Trust

Jessica Gentry
Oxford University Hospitals NHS Foundation Trust

Laura Dunn
Oxford University Hospitals NHS Foundation Trust

Nicole Stoesser
University of Oxford

Ann Sarah Walker
University of Oxford

Philip Bejon
University of Oxford

David Eyre
University of Oxford

Bernadette C. Young
University of Oxford

Method Article

Keywords: Clinical metagenomics, respiratory virus diagnostics, metagenomic sequencing, molecular diagnostics, agnostic virus detection, sequence-agnostic genomic sequencing, pandemic preparedness

Posted Date: August 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10471119/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Development of an Agnostic Clinical Metagenomic Workflow for Respiratory RNA Viruses

Kate E. Dingle[1,2], Katie M.V. Hopkins[1,2], Nicholas D. Sanderson[1,2], Ali Vaughan[1,2], Daisy Anderton[2], Matthew Colpus[2], Melody Parker[1,2], Syeda Anam Fatima[3], Jessica Gentry[3], Laura Dunn[3], Nicole Stoesser[1,2,3,4], Ann Sarah Walker[1,2,4], Philip Bejon*[1,2,3,4], David Eyre*[1,2,4], Bernadette C. Young*[1,2,3]

* contribution considered equal

Corresponding Author: Kate E. Dingle (email: kate.dingle@ndm.ox.ac.uk)

Affiliations

1 NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

2 Modernising Medical Microbiology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK

3 Oxford University Hospitals NHS Foundation Trust, Oxford, UK

4 NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford, Oxford, UK

Keywords: Clinical metagenomics, respiratory virus diagnostics, metagenomic sequencing, molecular diagnostics, agnostic virus detection, sequence-agnostic genomic sequencing, pandemic preparedness.

ABSTRACT

Background: Clinical metagenomics has the potential to provide an agnostic approach to respiratory virus diagnostics, but there are substantial methodological challenges. Viral loads vary by orders of magnitude among respiratory samples and nucleic acid extracts are dominated by non-target sequences, particularly for samples from the upper respiratory tract. This necessitates upstream virus enrichment and non-specific amplification approaches adaptable to high-throughput assays. We aimed to develop such a workflow for respiratory RNA viruses.

Methods: We used nasopharyngeal swabs in viral transport medium remaining after routine microbiological testing of clinical samples at Oxford University Hospital NHS Foundation Trust (UK) was completed. Pre-extraction virus enrichment and non-target nucleic acid depletion methods evaluated included differential centrifugation, bead-beating, nuclease treatment and virus precipitation. RNA was extracted, then reverse transcribed using random N_9 priming to include the oligonucleotide (5'-GATGATAGTAGGGCTTCGTCAC-N_9-3') used in subsequent amplification. Two amplification approaches were designed and compared side-by-side: isothermal multiple displacement amplification via Equiphi29 DNA polymerase using circularised template versus PCR-based sequence independent single primer amplification. Sequencing used the Oxford Nanopore Technologies platform.

Results: Our final optimised workflow combined multiple approaches to virus enrichment; sample clarification (3,200g centrifugation, 30 minutes), nuclease treatment, virus precipitation and centrifugation (13,200g, 30 minutes). RNA was extracted from the resultant pellet. Sequence independent single primer amplification was adopted in the final workflow as it consistently out-performed multiple displacement amplification. During development work, we detected a wide range of respiratory viruses, including coronaviruses (SARS-Cov-2, OC43, NC63, 229E), influenza A and B, respiratory syncytial virus, parainfluenza 2, 3 and 4, human metapneumovirus, human rhino-

and enteroviruses, together with an extraction control (MS2) and controls for reverse transcription and amplification controls (Zika, murine respirovirus and orthoreovirus RNA).

Conclusions: We have developed an agnostic clinical metagenomics workflow for RNA respiratory viruses, with potential for use in molecular diagnostics. Subsequent work will determine sensitivity and specificity relative to performance current standard-of-care diagnostics such as commercial respiratory PCR panels.

BACKGROUND

Sequencing of viral pathogens has become a key tool for global surveillance and public health protection [1,2], facilitating the identification and investigation of outbreak, epidemic and pandemic situations. Recent examples include the H1N1 influenza pandemic (2009), the Western Africa Ebola epidemic (2014-2016), Zika virus in South America (2015-2016), the SARS-Cov-2 pandemic (2020-2023), the emergence of mpox transmissible among humans, and most recently a cruise-ship related cluster of hantavirus cases [3-8].

Clinical metagenomics (CMg) extends the public health benefits of sequencing by determining whole genomes directly from clinical samples such as blood, CSF or respiratory samples, without culture or prior knowledge of infecting pathogen sequences. Pathogen reads are identified and genomes assembled bioinformatically.

Notable applications of CMg to respiratory viruses include determination of the SARS-Cov-2 genome early in the pandemic [9] and more broadly the detection of viruses not routinely captured by routine diagnostics [10]. Proof of principle uses in clinical settings include ongoing pandemic SARS-Cov-2 identification in patients with pneumonia [11], divergent human rhinovirus strain detection in hospitalised neonates [12], and analysis of acute respiratory samples [13-16]. If widely adopted,

routine diagnostic CMg would support the detection of both routine infections and emerging viral threats or novel strains [17-21].

The global burden of respiratory virus infections remains high [22-23], with current clinical laboratory diagnostics focused on syndromic PCR panels, such as the BIOFIRE FilmArray Respiratory 2.1 panel (Biomerieux) or Alinity Resp-4-plex assay (Abbott). In addition to only detecting a finite number of targets, such assays may also miss emergent or rapidly evolving strains due to primer or probe mismatches. Tests for RNA viruses with high mutation rates ~0.1–1 mutation per genome per replication [24,25] are particularly vulnerable [26] the pandemic SARS-CoV-2 S-gene failure in the ThermoFisher TaqPath assay being a classic example [27].

The lack of routine CMg protocols delivering high sensitivity for all samples, especially those with low viral titres, reflects multiple technical and regulatory challenges [28]. Respiratory samples vary by orders of magnitude in relative and absolute amounts of viral and other (human and bacterial) nucleic acids, also depending on whether samples come from the lower or upper respiratory tract. Respiratory viruses themselves vary in size, genome organisation and envelope presence/absence. CMg workflows for respiratory viruses are being developed and are in the early stages of clinical evaluation [15, 29-32].

We aimed to develop a sensitive, catch-all workflow prioritising detecting RNA respiratory viruses from upper respiratory tract samples, most commonly sampled. This choice reflected their role in six of the eight World Health Organisation Public Health Emergencies of International Concern since 2005, and previously reported sensitivity challenges for RNA viruses in respiratory CMg detection [30]. We chose an iterative approach, evaluating different procedures to enrich viral nucleic acids relative to background, combined with parallel testing of two novel agnostic cDNA amplification strategies.

METHODS

Clinical Samples

Respiratory RNA virus-positive and negative clinical samples comprised nose and/or throat swabs in viral transport medium (Sigma-Virocult (MWE)). Samples were submitted for routine clinical testing by PCR at the Microbiology Laboratory, Oxford University Hospital NHS Foundation Trust (UK). Respiratory virus positive samples were identified via the Alinity Resp-4-plex assay (Abbott, detecting influenza A, B, RSV and SARS-CoV-2), Xpert Xpress SARS-CoV-2/Flu/RSV plus (Cepheid) or BIOFIRE FilmArray Respiratory 2.1 panel (Biomerieux, 17 viral targets) and were collected retrospectively. Negative samples were identified by BIOFIRE FilmArray Respiratory 2.1 panel (Biomerieux) as it has the broadest target range of these commercial panels. Samples were obtained post-discard, stored at 4°C and used for assay development after routine testing was complete, therefore quality (volume, age) varied.

Sample Fractionation and Virus Enrichment

Four experimental approaches to virus enrichment were evaluated in different combinations; (i) centrifugation (to remove intact human cells), (ii) nuclease digestion of free nucleic acid within VTM (to remove free human and bacteria DNA), (iii) precipitation of intact virions and (iv) bead beating (to release human DNA prior to nuclease digestion). Details are summarised in Additional File 1.

Control Materials

Commercial preparations of internal process controls (Table 1) were used. The extraction control comprised intact bacteriophage MS2 particles (a small RNA virus that infects bacteria) added after sample fractionation, and immediately before RNA extraction (Additional File 1). The reverse transcription/amplification controls, genomic RNA from 3 RNA viruses (Zika, murine respirovirus and orthoreovirus), were added to RNA extracts just before reverse transcription (Additional File 1).

Table 1: Internal Process Controls

Process	Internal Control	Quantitative or Qualitative	RNA Genome	Stock Concentration	Supplier	Product Code
RNA Extraction	NATtrol™ MS2 Bacteriophage	Quantitative	3.6 kb ss +ve	1x10^7 particles/ml	Zeptometrix, Buffalo NY, USA	NATMS2-STQ
	MS2 Bacteriophage	Qualitative	3.6 kb ss +ve	As supplied	Applied Biosystems, Thermo-Fisher	100093312
Reverse transcription and amplification	Zika virus genomic RNA Strain: MR 766	Quantitative	10.8 kb ss +ve	≥ 1x10^5 genome copies/µl	ATCC, VA, USA	ATCC-VR-1838
	Murine respirovirus (Sendai) genomic RNA	Quantitative	15.4 kb ss -ve	≥ 1x10^5 genome copies/µl	ATCC, VA, USA	ATCC-VR-907
	Orthoreovirus genomic RNA (Reovirus Serotype 3; Strain: Dearing)	Quantitative	23.6 kb ds segmented (n=10) 1.15 - 3.92 kb	≥ 1x10^5 genome copies/µl	ATCC, VA, USA	ATCC-VR-824

ss: single stranded, ds: double stranded, +ve: positive sense RNA, -ve: negative sense RNA

RNA Extraction

Three commercial (automatable) RNA extraction kits were evaluated (Additional File 1), according to manufacturers' protocols for manual use. Extraction was done initially using either the MagMAX viral RNA or the MagMAX Viral/Pathogen Ultra Nucleic Acid Isolation Kit (ThermoFisher Scientific, San Diego, USA), to extract 340 or 390 µl viral transport medium (VTM), (plus 10 µl MS2 extraction control). Then the 50 µl eluate in DEPC treated water (Ambion) was concentrated to 9 µl (RNA clean and concentrate kit, Zymo Research, Irving, California, USA) prior to reverse transcription. Later experiments used a different approach, the Direct-zol RNA Microprep kit (Zymo Research) to extract fractionated pelleted material derived from 1ml VTM, resuspended in 20 µl PBS (plus 10 µl MS2 extraction control). The method used 9 volumes (270 µl) Tri reagent (Zymo Research), then RNA was eluted from the DNAse-treated column in 10µl DEPC treated water.

First Strand cDNA Synthesis

Reverse transcription was performed using the Maxima H Minus First Strand cDNA Synthesis Kit, with ds DNase (ThermoFisher) in a 20 µl reaction volume, containing 8 µl RNA extract, and 1 µl each reverse transcription/amplification control, Zika, murine respirovirus and orthoreovirus (Table 1). cDNA was primed randomly via the 3'-terminal N_9 of a custom oligonucleotide 5'/phos/GATGATAGTAGGGCTTCGTCAC-**N_9**-3' (Integrated DNA Technologies, Leuven, Belgium) at 5 µM final concentration, with incubation according to the manufacturer's instructions.
Single stranded cDNA underwent size selection and purification using SPRI (Solid-Phase Reversible Immobilization) beads; 0.9 volumes AMPure XP magnetic beads (Beckman Coulter, Indianapolis, Indiana, USA) were used and two 70% ethanol washes performed before elution of cDNA >~100 nucleotides in DEPC treated water (Ambion). cDNA was then taken forward to sequence independent single primer amplification (SISPA) PCR or multiple displacement amplification (MDA) using EquiPhi29 DNA Polymerase (ThermoFisher).

cDNA Amplification: SISPA PCR

Prior to SISPA, single stranded cDNA was made double stranded, using the same custom oligonucleotide as the first strand, (5'-GATGATAGTAGGGCTTCGTCAC-**N_9**-3', 2.5 µM final concentration) to incorporate the 'tail' sequence on the opposing strand. Sequenase™ Version 2.0 DNA Polymerase (ThermoFisher) was used with reaction conditions adapted from previous work [33]. Following a SPRI bead purification, ds DNA >~100 bp was eluted in 22 µl DEPC treated water (Ambion) and the 20 µl amplified by PCR using the oligonucleotide primer SISPA-1: 5'-GATGATAGTAGGGCTTCGTCAC-3'.

Each PCR (50 µl) comprised 20 µl ds cDNA, 5 µl SISPA 1 oligonucleotide (20 µM) and 25 µl NEBNext® Ultra™ II Q5® Master Mix (New England Biolabs, Ipswich, Massachusetts, USA). Amplification conditions were denaturation at 98 °C for 30 seconds (1 cycle), followed by 30 cycles of denaturation at 98 °C for 10 seconds, annealing 64 °C for 30 seconds, and extension 72 ˚C for 1 min 30sec (30 cycles), followed by a final extension of 72 ˚C for 2 minutes. A final SPRI bead purification was followed by elution in 15 µl DEPC treated water. The amplification products were quantified by Qubit fluorometer using the dsDNA 1X BR Assay (ThermoFisher).

cDNA Amplification: MDA

Single stranded cDNA from the first strand synthesis reaction was treated with RNAse H (New England Biolabs, Ipswich, Massachusetts, USA) then circularised by intramolecular ligation using CircLigase™ ssDNA Ligase (Biosearch Technologies, Teddington, UK). The 5' terminal G of the cDNA synthesis primer 'tail' was chosen to meet the base preference of CircLigase™ G > A >>> T >>> C, and contained a 5' terminal monophosphate to facilitate ligation. After sequencing, the 'tail' sequence 5'-GATGATAGTAGGGCTTCGTCAC-3' was identified bioinformatically and used to de-concatenate sequences derived from the circularised cDNA (Figure 1). The manufacturer's recommended 10-15% PEG 8000 (polyethylene glycol) was included during ligation to improve circularisation of longer fragments. The 20 µl reaction comprised 11 µl ssDNA template, 2 µl 10x reaction buffer, 1 µl 50mM $MnCl_2$, 1 µl 1mM ATP, 4 µl 50% PEG 8000 (buffer pack B0216L, New England Biolabs) (10% final), 1 µl

Circligase™. Incubation was at 60 °C for 1 hour, followed by inactivation at 80 °C for 10 minutes. Remaining linear single stranded DNA and linear double stranded DNA were removed by incubation with 20 units exonuclease I and 100 units exonuclease III (New England Biolabs), with incubation at 37˚C for 45 minutes. After SPRI beads purification, elution was in 15 µl DEPC treated water. MDA was primed using exonuclease resistant random hexamers and performed using thermostable Equiphi29 DNA polymerase (ThermoFisher) according to the manufacturer's instructions. After SPRI bead purification, 'debranching' was performed using T7 endonuclease (17 µl DNA, 2µl NEB buffer 2, 1 µl T7 endonuclease 1 (New England Biolabs, Ipswich, MA, USA), followed by a final SPRI purification and quantitation by Qubit fluorometer using the dsDNA 1X BR Assay (ThermoFisher).

Sequencing library preparation and sequencing

Libraries were prepared according to the Rapid Sequencing V14-Amplicon Sequencing protocol from Oxford Nanopore Technologies (ONT, Oxford, UK). Samples were sequenced using the Rapid Barcoding Kit V14 (RBK114.96; ONT) on R10.4.1 Flow Cells (ONT). Up to 26 samples were sequenced per flow cell for 72 hours with SUP base calling enabled.

Bioinformatics Methods

Amplified reads were detected from sequencing reads using an in-house pipeline then taxonomically classified and mapped to reference genomes. Reads were identified by the presence of the reverse transcription primer sequence using blastn v2.16.0+ [34] with custom settings (-task blastn-short -word_size 7 -gapopen 2 -max_target_seqs 10000000). For reads from SISPA PCR, primer sequences were trimmed, for reads from MDA, reads were split at every location where primer was identified and each fragment treated as an independent read. Kraken2 v2.17.1 [35] (default settings) with the k2_pluspf_20240605 index was used for taxonomic classification counts of amplified reads for each virus target. Trimmed reads were aligned to curated reference genomes using minimap2 [36] for coverage breadth and depths. All scripts are available in GitHub repository https://github.com/oxfordmmm/AD_analysis.

Statistical Tests

Read counts are presented as medians with interquartile ranges (IQRs) and ranges. Comparisons between two related groups, including different amplification methods, or treatments applied to the same clinical sample, were performed using two-sided Wilcoxon signed-rank tests. Comparisons between two independent samples were performed using two-sided Mann–Whitney U tests. Where three treatment conditions were compared within the same clinical samples, an overall Friedman test for repeated measures was used. A p-value < 0.05 was considered statistically significant.

Nucleotide Sequence Data

Data were uploaded to European Nucleotide Archive (ENA) under the accession PRJEB102181. Accession numbers are provided (Additional File 2).

RESULTS

Over a series of experiments, we identified best-performing combinations of virus enrichment, RNA extraction and non-specific amplification (Figure 1) strategies (Additional File 1).

Control Performance

Following initial spiking experiments using two negative clinical sample 'backgrounds' (BG1 and BG2), 10^5 or 10^4 genomes per reaction was selected for the reverse transcription/amplification controls (Zika, murine respirovirus or orthoreovirus genomic RNA) (Figure 2). A higher control concentration (10^6) was not used, to avoid competition from excessive non-target reads and for reasons of cost. The extraction control, MS2 was not detected (Additional File 3, Experiment 2). To investigate, MS2 RNA was extracted directly from control dilutions (without spiking) and amplified using a commercial qPCR assay (Applied Biosystems - ThermoFisher). MS2 was detected (Table 2), confirming successful extraction but at an insufficient number of copies to be detected by sequencing under the spiking conditions tested using BG1 and BG2. This sensitivity issue was

Table 2. Confirmation of MS2 Extraction

Sample	MS2 Copies	Ct Value
1	10^6	17
2	10^5	21
3	10^4	25
4	10^3	36
5	Negative 1	N/A
6	Negative 2	N/A

Quantitative MS2 control (Zeptometrix) was extracted using the MagMAX™ Viral RNA Isolation Kit (Thermo-Scientific) and then detected using a commercial PCR-based assay, which is designed to detect MS2 (Applied Biosystems - Thermo-Fisher SARS-Cov-2/MS2 PCR assay). Data obtained for the MS2 control are shown.

addressed in subsequent experiments by approaches aiming to (i) deplete non-target background DNA (ii) improve lysis of non-enveloped viruses and (iii) evaluate an alternative MS2 preparation.

Non-specific Amplification Strategies: SISPA out-performed MDA

To avoid running all experiments in duplicate, prior to evaluating the optimal background depletion and RNA extraction processes, we first identified the optimal non-specific amplification strategy. SISPA out-performed MDA for the control virus genomes (Figure 2). The performance of SISPA and MDA was also compared using 37 respiratory virus positive samples in which ≥2 virus reads were detected by SISPA (Figure 3). In 32 out of 37 instances, SISPA was equivalent to or superior to MDA, (both methods generated 3 reads from one sample), yielding a median (IQR) of 146 (5.5–4901.5) reads and 10 (1–254.5), respectively (Wilcoxon signed-rank test, $p < 0.0001$).

Overall, fewer reads were obtained for non-enveloped human rhino- and enterovirus (HRE; labelled as “picornavirus”/*Picornaviridae*” in Figure 3) relative to enveloped viruses, e.g., using SISPA, 4 (2–15.5) reads vs. 739 (31–7928.5) reads, respectively (Mann–Whitney U test, $p < 0.001$) (Figure 3). SISPA was therefore used in preference to MDA in optimisation of virus enrichment and host depletion.

Centrifugation: Pellet Fraction Best-Performing

Clinical samples (n=12) underwent fractionation by centrifugation (5,000 g, 30 minutes), as described [37] for the removal of cellular debris from virus-containing cell culture fluid. Sequencing reads (n) obtained for the per sample fractions were compared (unfractionated 'neat' VTM, pellet and supernatant; Figure 4, Experiments 5 and 6). In 6/9 samples in which virus was detected (7/10 observations due to one dual infection), the pellet fraction contained more reads than the corresponding supernatant fraction, median (IQR; range) read counts of 239 (5–5053; 0–61,075) and 2 (1–16,380; 0–28,343), respectively (Wilcoxon signed-rank test, p = 0.33). The pellet fraction also performed better than the neat VTM in 6/9 samples (Figure 4), (7/10 observations as before), with median (IQR; range) read counts of 239 (5–5053; 0–61,075) and 80 (2–2491; 0–54,786), respectively; (Wilcoxon signed-rank test, p = 0.15).

Even after fractionation, detection of the extraction control MS2 (10^5 per extraction) and HRE remained poor; notably, both are non-enveloped and contain the smallest genomes of all viruses targeted. We therefore used the pellet fraction only in the immediately following experiments, before subsequently revising this strategy in the final workflow (described below).

Improved Detection of Non-Enveloped Viruses

Because of the challenges of obtaining reads from non-enveloped viruses, to improve disruption of tough non-enveloped viral capsids, a phenol and guanidine isothiocyanate-based approach to RNA extraction [38] was evaluated using the Direct-zol RNA Microprep kit (Zymo Research). This kit also included an on-column DNAse 1 treatment and reduced the RNA elution volume (10 µl). DNAse1 aimed to reduce background DNA, while the small elution volume allowed virtually all RNA derived from the initial 1ml VTM to be reverse transcribed. All these changes were introduced with the aim of improving workflow sensitivity.

A further 16 clinical samples, which were PCR positive for 22 respiratory viruses (including 8 HRE) were obtained and processed to a pellet as above (centrifugation 5,000 g, 30 minutes). Following the

phenol-based Direct-zol extraction, at least one respiratory virus was detected in 12/16 samples, with 15/22 respiratory viruses detected overall (Figure 5, Additional File 3, Experiment 7). The nine HRE positive samples yielded 1, 357, 1, 2, 0, 4, 0, 0, 2 reads (≥1 HRE read from 6/9 pellets). This compared to 5, 0, 3, 25, 0, 0, and 0 reads (≥1 HRE read from 3/7 pellets) for HRE positive samples in previous experiments using MagMAX extraction kits (Figure 4). The difference in HRE reads between the Direct-zol and MagMAX approaches was not statistically significant, with median (IQR; range) read counts of 1 (0–3; 0–357) and 0 (0–5; 0–25), respectively (Mann–Whitney U test, p = 0.93). However, Direct-zol was chosen for subsequent work since (a) one sample yielded substantially more HRE reads (357), confirming that higher-level read counts are achievable using this kit, and (b) HRE detection was slightly better by Direct-zol (although not statistically significant), with ≥1 HRE read detected in 6/9 samples, versus 3/7 samples by MagMAX. The MS2 (10^5) extraction control was detected in 4/16 samples (1, 1, 1, 3 reads).

Yield Enhanced by Intact Virus Precipitation

Returning next to try to further improve the enrichment of virus RNA prior to extraction, a combination of differential centrifugation and precipitation were assessed, with the phenol-based Direct-zol extraction. Three pellets were generated from each of 10 clinical samples. Firstly, 1 ml VTM was clarified at reduced speed (3,200g, 30 minutes - **pellet 1**) to ensure the majority of viral RNA remained in the supernatant. The supernatant from this underwent a second, high-speed clarification (13,200g, 30 minutes **- pellet 2**). Then the supernatant from this underwent PEG precipitation using intact virus precipitation reagent (Invitrogen, Paisley, UK), with a final high-speed centrifugation step yielding a third pellet (13,200g, 30 minutes - **pellet 3**). The respiratory virus reads (n) obtained for each pellet were compared across samples (Figure 6, Additional File 3, Experiment 8).

Respiratory virus reads were detected in 6/10 samples, and across all three pellets in 5 of these (Figure 6). In the 6 samples, pellets 2 and 3 always out-performed pellet 1. In 4/6 samples pellets 2

and 3 both yielded well over 10-fold higher numbers of reads than pellet 1. For the same six samples, pellets 1, 2 and 3 yielded median (range) read counts of 42 (0–133), 706 (0–4383), and 681 (4–3189), respectively (overall P-value for differences, p = 0.048 (Friedman test)). This suggested that the large pellet 1 generated by slower initial centrifugation functioned as intended, clarifying the VTM but with limited viral material co-pelleting with the cellular material.

Detection of the MS2 control improved relative to previous experiments with reads detected in at least one pellet for 8/10 samples (Figure 6), compared to only 3/16 samples previously (Figure 5). Here, MS2 was detected in 3/3 pellets from two clinical samples, 2/3 pellets from four samples, 1/3 pellets from two samples and undetected in two samples. This may reflect improved MS2 extraction and, or better depletion of non-target nucleic acid. However, HRE detection remained challenging, with only one of the three samples tested yielding any reads (Figure 6).

Nuclease Treatment of Clarified VTM

Next, we examined nuclease treatment of clarified VTM using two different salt-active nucleases. In the first of two experiments (Figure 7A, Additional File 3, Experiment 10), 3x1ml clarified VTM aliquots were prepared per clinical sample (n=8). They underwent either: (i) no nuclease treatment (control), (ii) M-SAN HQ (Arcticzymes) incubation at 37°C for 30 minutes, followed by PEG precipitation (Intact virus precipitation reagent, Invitrogen) or (iii) SAN HQ (Arcticzymes) incubation during the PEG precipitation. Then they underwent high-speed centrifugation, (13,200g, 30 minutes) generating a pellet for extraction.

We found that SAN treatment was unsuccessful (no virus reads detected; Figure 7A) and it was not pursued. However, M-SAN treatment was investigated further. An additional 10 clinical samples were treated and compared to ‘no nuclease’ control, bringing the total number of samples assessed in this way to 18 (Figure 7A & B, Additional File 3, Experiment 10 & 10A).

Respiratory virus reads were detected in 14/18 samples (both with and without M-SAN treatment). Of the fourteen, 5 were only weakly positive, having ≤15 reads (again, by both methods). The

remaining 9 samples were used to assess the effectiveness of M-SAN treatment. M-SAN pre-treatment yielded higher read counts in 6/9 samples (Figure 7). Median (IQR; range) read counts were 87 (44–1711; 21–5416) with M-SAN, versus 66 (25.5–1583.5; 4–3805) without. Although, this difference was not statistically significant (Wilcoxon signed-rank test, p = 0.30), due to the potential benefit here, and a report of adoption by others in place of HL-SAN [39], M-SAN was retained in our final workflow.

With respect to the MS2 extraction control, an alternative qualitative preparation (ThermoFisher) was tested. MS2 reads were detected in all but one of the 18 clinical samples, and in 32/44 fractions overall, with the number of reads exceeding 10 in 19 of these (Figures 7A, 7B). This alternative MS2 stock was therefore adopted going forwards. The RT/amplification controls performed robustly at 10^4 genomes per extraction (Figures 7A, B).

Mechanical Lysis of Human cells

Within a final set of experiments, for each of 10 clinical samples, a mechanical lysis bead beating treatment before M-SAN incubation was compared to no bead beating pre-treatment. Respiratory virus reads were detected in 7/10 samples (5 by both methods, 1 by each alone) (Figure 8, Additional File 3, Experiment 11). An increased number of reads post bead beating/M-SAN treatment were detected in only 3/7 samples but, in 2/3 the numbers of reads were <10. For the seven samples, bead beating then M-SAN yielded median (IQR; range) read counts of 1 (1–47; 0–802), and M-SAN alone read counts of 3 (1–29; 0–1253), with no significant difference between treatments (Wilcoxon signed-rank test, p = 0.58). Bead beating prior to nuclease treatment was not pursued further. The qualitative MS2 stock (ThermoFisher) performed consistently (Figure 8) as before (Figure 7A, B).

Final workflow and Genome Coverage

Our final workflow comprised nuclease treatment of clarified VTM, intact virus precipitation, then high-speed centrifugation, followed by extraction of the resultant pellet (Figure 9). Double stranded cDNA underwent non-specific amplification by SISPA before sequencing using the ONT platform. For

all samples in the final workflow, (Figure 8, Additional File 3, Experiment 11 – samples without beat beating) the median (IQR)[min-max] coverage breadth for MS2, Zika, murine respirovirus, and orthoreovirus was 52% (21%)[0-87%], 72% (13%)[2-91%], 64% (29%)[0-92%], and 34% (22%)[1-54%], respectively. Reads mapped across the length of the genomes for the extraction control, the reverse transcription/amplification controls and the infecting respiratory virus, SARS-CoV-2 are shown for samples 6 and 10, (Experiment 11, Additional File 3) as there were sufficient reads for coverage plots (Figure 10).

DISCUSSION

During CMg workflow development, a total of 12 different respiratory viruses; coronaviruses (SARS-Cov-2, OC43, NC63, 229E), influenza A and B, RSV, parainfluenza 2, 3 and 4, human metapneumovirus and HRE were successfully detected (Figures 4-8). Sequencing reads were used to measure and compare the performance of different workflow iterations, and data are shown for a range of optimisation experiments (Figures 2-8, Additional File 1). Sequence data for 66 clinical samples are shown in Figures 4-8. Samples were split into 146 different fractions sequenced through multiplexing at 13 – 26 fractions per flow cell to reduce cost. Data for two experiments conducted for operator training are shown in Additional File 3 only, as these yielded few informative data points.

Since throat/nasal swabs contain large amounts of non-target nucleic acid, obtaining useful numbers of viral reads depends on pre-extraction virus enrichment or non-target nucleic acid depletion. Chemical and enzymatic methods have been used previously to achieve human DNA depletion, for example saponin/nuclease treatment, MolYsis, HostZERO (Zymo Research) and DNA Microbiome (QIAamp), [40-42]. However, these are unsuitable for RNA viruses. They are not optimised for the recovery of RNA, which is much more easily hydrolysed than DNA and subject to degradation by cellular RNAses. Instead, post-extraction human rRNA depletion by RNAse H and specific probes has

371 been used successfully [15, 43-45]. But the cost of commercial RNAse H-based kits is high, adding at
372 least £40-£50/sample and this was not evaluated here.

373 Instead, to maintain the advantages of CMg as a potentially agnostic detection method, we
374 identified approaches to pre-extraction virus enrichment with the potential to work equally
375 efficiently for all RNA virus taxa. We evaluated combinations of differential centrifugation, bead
376 beating, endonuclease treatment and virus precipitation (Figures 4-8). Low speed centrifugation is
377 often used to sediment large human cells and debris, with the supernatant taken forward for further
378 processing. We discarded an obvious pellet from most samples, while removing relatively little
379 target virus RNA (Figure 6). Filtration is another option for large debris removal, but centrifugation
380 was preferred due to simplicity and reports of similar performance for both [33]. Free nucleic acid
381 was removed from our clarified supernatant using the salt tolerant endonuclease M-SAN, which out-
382 performed SAN and bead beating under the conditions tested (Figure 7, 8). Following M-SAN
383 treatment, intact virions were precipitated using Intact Virus Precipitation Reagent (Invitrogen) and
384 collected by high-speed centrifugation. This yielded a pellet from which RNA was extracted. Our
385 results for nuclease-treatment using M-SAN were variable, certain samples benefiting, but others
386 not (Figure 7) despite an incubation time of 30 minutes and optimised Mg^{2+} concentrations. Given
387 the cost (~£35 per sample), it may be worthwhile exploring alternatives such as human rRNA
388 depletion kits, the cost of which may be justified if consistent benefits were achieved for every
389 sample.

390 An advantage of PEG precipitation was it allowed almost 1ml starting volume of VTM, around twice
391 that of other respiratory CMg protocols [15,29], while generating a compact pellet for RNA
392 extraction. Virus precipitation reagents typically contain buffered polyethylene glycol and NaCl.
393 Precipitation occurs by physical chemistry; PEG hydration reduces available solvent and NaCl helps
394 neutralise surface charges causing virus aggregates to form, which can be pelleted by centrifugation
395 [46]. This process can be automated using magnetic beads and KingFisher™ instruments and

396 although optimised for SARS-CoV-2, the manufacturer's datasheet indicates success with influenza
397 A, Ebola, Zika and RSV. Since CMg for non-enveloped HRE proved challenging, we chose an
398 extraction kit combining the advantages of phenol and guanidine isothiocyanate [38] for capsid lysis
399 with silica column-based RNA purification, aiming to improve sensitivity for HRE without
400 compromising enveloped viruses.

401 We found that SISPA out-performed MDA for cDNA amplification (Figure 3). In the 66 clinical
402 samples represented in Figure 4-8, genome coverage (depth and breadth) allowed complete
403 genomes to be determined for higher titre samples. Ct values were recorded when available for
404 samples positive by Alinity Resp-4-plex assay, or Xpert Xpress CoV-2/Flu/RSV plus (Additional File 3),
405 (but not BIOFIRE FilmArray Respiratory 2.1 panel positives as Ct is not provided). Coverage declined
406 towards the 5' and 3' termini of genomes/genome segments (Figure 9, 10), reflecting the nature of
407 RT-PCR.

408 The ONT platform was used, together with the Rapid sequencing V14 Amplicon sequencing kit (SQK-
409 RBK114). Advantages for CMg include ease of setting up sequencing runs, and real-time data
410 analysis, together with longer read lengths. Costs and hands on time per sample (estimated as
411 £133.66, and 70.3 minutes [47]) were mitigated by multiplexing up to 26 samples per run.

412 Our workflow was designed to maximise sensitivity for RNA viruses, which has been previously
413 highlighted as challenging [29]. The detection of non-enveloped HRE remained more challenging
414 than enveloped viruses, even after modifications designed to improve this. Difficulties of identifying
415 HRE may reflect long term shedding of rhinovirus RNA post infection at low, sub clinical levels [48],
416 or may be due to differences in efficiency of PEG precipitation for different RNA viruses. PEG
417 precipitation has been proposed as a generalisable solution for both enveloped and non-enveloped
418 viruses, [49-51] but we did not find this to be the case in our workflow, and results elsewhere have
419 been variable [52-53].

Our study has several limitations. Firstly, small sample numbers, due to seasonal availability and high manual processing time, limited the total number assessed and our power to differentiate methods. Secondly, sample age ranged from 1 to 33 days (Additional File 3) affecting RNA quality; viral load per sample was not quantified at the time of extraction however, therefore the impact of starting virus titre cannot be fully assessed. Thirdly, a formal estimation of the fold-enrichment provided by each pre-extraction step would have been informative, but would have required extraction of supernatant (as well as pelleted material) and consequently varied extraction methods.

Overall, our experience illustrates the difficulty of delivering a single CMg protocol with maximal sensitivity for all viruses, since modifications targeting one virus may compromise another. cDNA amplification may also bias sensitivity, and it has been proposed that SISPA selectively amplifies longer cDNA molecules [54]. Our control data support this, with single stranded 15.4kb murine respirovirus and 10.8kb Zika virus yielding significantly more reads than 10-segmented orthoreovirus (1.15 - 3.92 kb segments: 23.6kb total). This amplification-related effect may also have impacted our MS2 extraction control (3.6kb RNA genome) and HRE (~7.5kb RNA).

Despite improvements, CMg for respiratory viruses remains technically challenging and it cannot yet equal the sensitivity of PCR. Workflow sensitivity could potentially be increased further by host rRNA depletion, or increasing sequencing depth [15], for example by reducing the numbers of samples multiplexed per flow cell, but this would increase costs. Finally, lower sensitivity for highly genetically diverse non-enveloped RNA viruses may also reflect bioinformatic challenges if all variants are not detected with equal efficiency.

CONCLUSIONS

A CMg workflow for RNA respiratory viruses was developed. Enrichment strategies including centrifugation, nuclease treatment and virus precipitation were combined to maximise sensitivity, together with SISPA and sequencing using the ONT platform. In total, twelve clinically important

444 respiratory RNA viruses were detected, with whole genome sequences determined for high titre
445 samples. Subsequent work [47] will determine workflow sensitivity and specificity prospectively
446 relative to commercial respiratory PCR panels, to assess its potential for routine use.

447 **FIGURE LEGENDS**

448 **Figure 1: Overview of cDNA Amplification: SISPA and MDA**

449 **Figure 2: Performance of Control Materials**
450 Reads classified per sample, using Kraken 2, for reverse transcription and amplification controls Zika,
451 murine respiro and orthoreovirus; reads obtained following MDA versus SISPA (Experiment 2,
452 Additional File 3).

453 **Figure 3: Performance of MDA versus SISPA**
454 Reads (n) classified using Kraken 2, for 'target virus' for 37 respiratory virus positive clinical samples.
455 Each vertical line represents one sample, from which cDNA was synthesised; half was amplified by
456 SISPA and half by MDA then sequenced on the ONT platform. Samples yielding ≥ 2 reads following
457 SISPA are shown, ranked by respiratory virus reads (n) (Experiments 5-8, Additional File 3).

458 **Figure 4: Sample Clarification by Low Speed Centrifugation**
459 Reads (n) classified for 12 respiratory virus-positive clinical samples, using Kraken 2 (Experiments 5
460 and 6, Additional File 3). The viruses detected previously in each sample by respiratory virus PCR
461 panel are indicated by red and blue boxes above the plot; five samples contained dual infections.
462 Each sample underwent three pre-extraction treatments, indicated in the blue box below the plot,
463 then extraction used the kit indicated by the grey bar above. Respiratory virus reads (n) obtained are
464 indicated by bright red and blue dots. MS2 extraction control was used at 10^5 genomes per
465 extraction and MS2 reads are indicated by blue/grey dots. The reverse transcription/amplification
466 control amount (Zika, murine respiro and orthoreoviruses) is indicated by the yellow box/line; 10^5
467 genomes per reaction, and their reads (n) by grey dots.

Figure 5: Improving detection of non-enveloped HRE and MS2

Reads (n) classified using Kraken 2 for 16 clinical samples, containing 22 respiratory viruses (6 dual infections) (Experiment 7, Additional File 3). Virus(es) previously detected in each sample by respiratory virus PCR panel are indicated above the plot, and the clinical samples are numbered below the plot. 1ml VTM per sample underwent low speed centrifugation (5000g, 30 minutes). The supernatant was discarded and the pellet (together with MS2 extraction control at 10^5 per sample) was extracted using Direct-zol RNA Microprep (Zymo). Respiratory virus reads (n) are indicated by bright red and blue dots, MS2 reads by blue/grey dots and reverse transcription/amplification control reads (Zika, murine respiro and orthoreoviruses; 10^5 genomes per reaction - yellow box and arrow) are indicated by grey dots.

Figure 6: Sample Fractionation and Intact Virus Precipitation

Reads (n) classified using Kraken 2, for 10 virus-positive clinical samples (Experiment 8, Additional File 3). The virus previously detected in each sample by respiratory virus PCR panel is indicated above the plot, and they are numbered below the plot. 1ml VTM from each sample was fractionated into three pellets, as indicated by 'sample treatment' (blue box). The supernatant from pellet 1 was used to generate pellet 2 and the supernatant from pellet 2 was used to generate pellet 3; the final supernatant was discarded, then pellets together with MS2 control at 10^5, were extracted using the Direct-zol RNA Microprep kit (Zymo). Respiratory virus reads (n) are indicated by red dots, MS2 reads by blue/grey dots, and the reverse transcription/amplification control reads (10^5 genomes per reaction Zika, murine respiro and orthoreoviruses, yellow line and box) by grey dots.

Figure 7: Nuclease treatment of Clarified Supernatant

(A) Reads (n) classified by Kraken 2 for 8 virus-positive clinical samples (Experiment 10, Additional File 3). Treatments 1, 2 and 3 are as indicated (blue box), and each generated a pellet, which was extracted with the Direct-zol RNA Microprep kit (Zymo). Virus previously detected in each sample by respiratory PCR panel is indicated above the plot. Clinical samples are numbered below

the plot. The concentration of reverse transcription/amplification controls used (Zika, murine respiro and orthoreoviruses) is indicated by the yellow box/line (10^4 genomes per reaction) and their read numbers by grey circles. MS2 (qualitative stock, Thermo Fisher) reads are indicated by blue/grey dots. Respiratory virus reads (n) detected are indicated by red dots.

(B) Reads (n) classified by Kraken 2 for 10 virus-positive clinical samples (Experiment 10A, Additional File 3); treatments 1 and 2 are as indicated (blue box), these are as in (A). All other details are as (A).

Figure 8: Human cell disruption by bead beating before nuclease treatment

Reads classified by Kraken 2 for 10 virus-positive clinical samples (Experiment 11, Additional File 3). The virus previously detected in each sample by respiratory PCR panel is indicated above the plot, and they are numbered below the plot. Two separate 1ml VTM aliquots underwent either of two treatments as indicated (blue box), prior to extraction of the pellet (together with MS2) with the Direct-zol RNA Microprep kit (Zymo). Respiratory virus reads (n) are indicated by red dots, MS2 reads by blue/grey dots, and the reverse transcription/amplification control reads (10^4 genomes per reaction Zika, murine respiro and orthoreoviruses, yellow line and box) by grey dots.

Figure 9: Final Workflow

Figure 10: Amplified read genome coverage for two samples positive for SARS-CoV-2

Read genome coverage plots for two samples (6 and 10, Experiment 11, Additional File 3) with sufficient reads mapping for the target pathogen (SAR-CoV-2) to display coverage traces. Also included are coverage plots for the extraction control MS2 and the reverse transcription/amplification controls (Zika, murine respiro and orthoreoviruses). Red and blue lines display results following treatments 1 and 2, respectively (treatment 1 being incubation of clarified supernatant with M-SAN followed by virus precipitation and high-speed centrifugation and treatment 2 as 1 but with prior bead beating). Coverage has been truncated to 20x read depth.

REFERENCES

1. Gauthier NPG, Chorlton SD, Krajden M, Manges AR. Agnostic sequencing for detection of viral pathogens. Clin Microbiol Rev. 2023;36(1):e00119-22. doi:10.1128/cmr.00119-22.

2. Caini S, Meijer A, Nunes MC, Henaff L, Zounon M, Boudewijns B, Del Riccio M, Paget J. Probable extinction of influenza B/Yamagata and its public health implications: a systematic literature review and assessment of global surveillance databases. Lancet Microbe. 2024;5(8):100851. doi:10.1016/S2666-5247(24)00066-1.

3. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature. 2015;524(7563):97-101. doi:10.1038/nature14594.

4. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC, Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12(6):1261-1276. doi:10.1038/nprot.2017.066.

5. Chiara M, D'Erchia AM, Gissi C, Manzari C, Parisi A, Resta N, Zambelli F, Picardi E, Pavesi G, Horner DS, Pesole G. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. Brief Bioinform. 2021;22(2):616-630. doi:10.1093/bib/bbaa297.

6. Kumar S, Henrickson KJ. Update on influenza diagnostics: lessons from the novel H1N1 influenza A pandemic. Clin Microbiol Rev. 2012;25(2):344-361. doi:10.1128/CMR.05016-11.

7. Srivastava S, Sharma D, Sridhar SB, Kumar S, Koteswara Rao GSN, Budha RR, Babu MR, Sahu R, Sah S, Mehta R, Giraldo-Corrales NA, Feehan J, Apostolopoulos V, Rodriguez-Morales AJ.

Comparative analysis of Mpox clades: epidemiology, transmission dynamics, and detection strategies. BMC Infect Dis. 2025;25(1):1290. doi:10.1186/s12879-025-11784-8.

8. Palacios G, Hoxha A, Oude Munnink B, Rambaut A, Andersen KG, Bedford T, et al. Preliminary analysis of Orthohantavirus andesense virus sequences from a cruise-ship related cluster, May 2026. Virological.org. 2026. Available from: https://virological.org/t/preliminary-analysis-of-orthohantavirus-andesense-virus-sequences-from-a-cruise-ship-related-cluster-may-2026/1029. Accessed 11 Jun 2026.

9. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–269. doi:10.1038/s41586-020-2008-3.

10. Lu L, Robertson G, Ashworth J, Pham Hong A, Shi T, Ivens A, Thwaites G, Baker S, Woolhouse M, VIZIONS Consortium. Epidemiology and phylogenetic analysis of viral respiratory infections in Vietnam. Front Microbiol. 2020;11:833. doi:10.3389/fmicb.2020.00833.

11. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W; China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727–733. doi:10.1056/NEJMoa2001017.

12. Li H, Xiao Y, Lü J, Zhang W, Lu H. Viral metagenomic analysis reveals a human rhinovirus from hospitalized neonates. Microbiol Resour Announc. 2021;10:e00106-21. doi:10.1128/MRA.00106-21.

13. Sandybayev N, Beloussov V, Strochkov V, Solomadin M, Granica J, Yegorov S. Metagenomic profiling of nasopharyngeal samples from adults with acute respiratory infection. R Soc Open Sci. 2024;11:240108. doi:10.1098/rsos.240108.

14. Babiker A, Bradley HL, Stittleburg VD, Ingersoll JM, Key A, Kraft CS, Waggoner JJ, Piantadosi A. Metagenomic sequencing to detect respiratory viruses in persons under investigation for COVID-19. J Clin Microbiol. 2020;59:e02142-20. doi:10.1128/JCM.02142-20.

15. Tan JK, Servellita V, Stryke D, Kelly E, Streithorst J, Sumimoto N, Foresythe A, Huh HJ, Nguyen J, Oseguera M, Brazer N, Tang J, Ingebrigtsen D, Fung B, Reyes H, Hillberg M, Chen A, Guevara H, Yagi S, Morales C, Wadford DA, Mourani PM, Langelier CR, de Lorenzi-Tognon M, Benoit P, Chiu CY. Laboratory validation of a clinical metagenomic next-generation sequencing assay for respiratory virus detection and discovery. Nat Commun. 2024;15(1):9016. doi:10.1038/s41467-024-51470-y.

16. Chen XH, Zhou SJ, Liu YY, Cao H, Zheng YR, Chen Q. Application value of metagenomics next-generation sequencing in the diagnosis of respiratory virus infection after congenital heart surgery. *Transl Pediatr*. 2024;13(2):260–270. doi:10.21037/tp-23-341.

17. Dos Santos MM, Salles TS, de Campos GM, Brcko IC, Lima ARJ, Sampaio SC, Elias MC, Giovanetti M, Slavov SN. Acute respiratory infections (ARIs): current etiological perspectives and advances in viral metagenomics-a review. Viruses. 2025;17(12):1554. doi:10.3390/v17121554.

18. Makoni M. Launch of genomic surveillance system for respiratory viruses. Lancet Microbe. 2023;4(4):e214. doi:10.1016/S2666-5247(23)00073-3.

19. Downie DL, Rao P, David-Ferdon C, Courtney S, Lee JS, Quiner C, MacDonald PDM, Barnes K, Fisher S, Andreadis JL, Chaitram J, Mauldin MR, Salerno RM, Schiffer J, Gundlapalli AV. Surveillance for emerging and reemerging pathogens using pathogen agnostic metagenomic sequencing in the United States: a critical role for federal government agencies. Health Secur. 2024;22(2):85–92. doi:10.1089/hs.2023.0099.

20. Ko KKK, Chng KR, Nagarajan N. Metagenomics-enabled microbial surveillance. Nat Microbiol. 2022;7(4):486–496. doi:10.1038/s41564-022-01089-w.

586 21. Torres Montaguth EO, Buddle S, Morfopoulou S, Breuer J. Clinical metagenomics for
587 diagnosis and surveillance of viral pathogens. Nat Rev Microbiol. 2026;24(1):61–75.
588 doi:10.1038/s41579-025-01223-5.

589 22. Hanage WP, Schaffner W. Burden of acute respiratory infections caused by influenza virus,
590 respiratory syncytial virus, and SARS-CoV-2 with consideration of older adults: a narrative review.
591 Infect Dis Ther. 2025;14(1):5-37. doi:10.1007/s40121-024-01080-4.

592 23. Jin X, Ren J, Li R, Gao Y, Zhang H, Li J, Zhang J, Wang X, Wang G. Global burden of upper
593 respiratory infections in 204 countries and territories, from 1990 to 2019. EClinicalMedicine.
594 2021;37:100986. doi:10.1016/j.eclinm.2021.100986.

595 24. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and
596 determinants. Nat Rev Genet. 2008;9(4):267-276. doi:10.1038/nrg2323.

597 25. Duffy S. Why are RNA virus mutation rates so damn high? PLoS Biol. 2018;16(8):e3000003.
598 doi:10.1371/journal.pbio.3000003.

599 26. Klungthong C, Chinnawirotpisan P, Hussem K, Phonpakobsin T, Manasatienkij W,
600 Ajariyakhajorn C, Rungrojcharoenkit K, Gibbons RV, Jarman RG. The impact of primer and probe-
601 template mismatches on the sensitivity of pandemic influenza A/H1N1/2009 virus detection by real-
602 time RT-PCR. J Clin Virol. 2010;48(2):91–95. doi:10.1016/j.jcv.2010.03.012.

603 27. Kidd M, Richter A, Best A, Cumley N, Mirza J, Percival B, Mayhew M, Megram O, Ashford F,
604 White T, Moles-Garcia E, Crawford L, Bosworth A, Atabani SF, Plant T, McNally A. S-variant SARS-
605 CoV-2 lineage B1.1.7 is associated with significantly higher viral load in samples tested by TaqPath
606 polymerase chain reaction. J Infect Dis. 2021;223(10):1666–1670. doi:10.1093/infdis/jiab082.

607 28. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20(6):341-355.
608 doi:10.1038/s41576-019-0113-7.

29. Alcolea-Medina A, Alder C, Snell LB, Charalampous T, Aydin A, Nebbia G, Williams T, Goldenberg S, Douthwaite S, Batra R, Cliff PR, Mischo H, Neil S, Wilks M, Edgeworth JD. Unified metagenomic method for rapid detection of microorganisms in clinical samples. Commun Med (Lond). 2024;4(1):135. doi:10.1038/s43856-024-00554-3.

30. Alcolea-Medina A, Snell LB, Humayun G, Al-Yaakoubi N, Ward D, Alder C, Patel V, Vivian F, Meadows CIS, Wyncoll D, Paul R, Barratt N, Batra R, Edgeworth J, Nebbia G, Whitehorn J. Rapid pan-microbial metagenomics for pathogen detection and personalised therapy in the intensive care unit: a single-centre prospective observational study. Lancet Microbe. 2025;6(10):101174. doi:10.1016/j.lanmic.2025.101174.

31. Gauthier NPG, Chan W, Locher K, Smailus D, Coope R, Charles M, Jassem A, Kopetzky J, Chorlton SD, Manges AR. Validation of an automated, end-to-end metagenomic sequencing assay for agnostic detection of respiratory viruses. J Infect Dis. 2024;230(6):e1245–e1253. doi: 10.1093/infdis/jiae226.

32. Su YS, Tsai WH, Wu HC, Chiu YT, Jiang NR, Lee CY, Cheng SH, Huang CT, Chi CY, Lin EJ, Kuo YP, Tsai WT, Tien CF, Liao YC, Lee KL, Chen FJ, Yu GY. Multiplex metagenomic sequencing for rapid viral pathogen identification and surveillance in clinical specimens. BMC Infect Dis. 2025;25(1):1531. doi:10.1186/s12879-025-11952-w.

33. Lewandowski K, Xu Y, Pullan ST, Lumley SF, Foster D, Sanderson N, Vaughan A, Morgan M, Bright N, Kavanagh J, Vipond R, Carroll M, Marriott AC, Gooch KE, Andersson M, Jeffery K, Peto TEA, Crook DW, Walker AS, Matthews PC. Metagenomic nanopore sequencing of influenza virus direct from clinical respiratory samples. J Clin Microbiol. 2020;58(1):e00963-19. doi:10.1128/JCM.00963-19.

34. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi:10.1186/1471-2105-10-421.

633 35. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol.
634 2019;20:257. doi:10.1186/s13059-019-1891-0.

635 36. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics.
636 2018;34(18):3094–3100. doi:10.1093/bioinformatics/bty191.

637 37. Payne S. Methods to study viruses. In: Viruses: from understanding to investigation. London:
638 Academic Press; 2017. p. 37–52. doi:10.1016/B978-0-12-803109-4.00004-0.

639 38. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium
640 thiocyanate-phenol-chloroform extraction. Anal Biochem. 1987;162(1):156–159.
641 doi:10.1006/abio.1987.9999.

642 39. Doyle RM, Patel V, Humayun G, Chang R, Gumbanmarasigan M, Vorslava V, Leka B, Girgis S,
643 Ward D, Truesdale L, Brown JR, Hammond R, Kopec A, Rock SA, Valappil M, Laws K, Parmar S, May S,
644 Patel P, Ibrahim M, Green S, Saeed K, Norton N, Grant P, Nebbia G, Batra R, Edgeworth J, Snell LB,
645 Alcolea-Medina A. Optimisation of whole cell human depletion provides increased sensitivity and
646 microbial genome coverage from respiratory metagenomic assays. medRxiv [Preprint]. 2025.
647 doi:10.1101/2025.10.10.25337716.

648 40. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, Rae D, Grundy S, Turner
649 DJ, Wain J, Leggett RM, Livermore DM, O'Grady J. Nanopore metagenomics enables rapid clinical
650 diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783-792.
651 doi:10.1038/s41587-019-0156-5.

652 41. Heravi FS, Zakrzewski M, Vickery K, Hu H. Host DNA depletion efficiency of microbiome DNA
653 enrichment methods in infected tissue samples. J Microbiol Methods. 2020;170:105856.
654 doi:10.1016/j.mimet.2020.105856.

655 42. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker TM, Quan TP,
656 Wyllie DH, Del Ojo Elias C, Wilcox M, Walker AS, Peto TE, Crook DW. Mycobacterial DNA extraction

for whole-genome sequencing from early positive liquid (MGIT) cultures. J Clin Microbiol. 2015;53(4):1137-1143. doi:10.1128/JCM.03073-14.

43. Matranga CB, Andersen KG, Winnicki S, Busby M, Gladden AD, Tewhey R, Stremlau M, Berlin A, Gire SK, England E, Moses LM, Mikkelsen TS, Odia I, Ehiane PE, Folarin O, Goba A, Kahn SH, Grant DS, Honko A, Hensley L, Happi C, Garry RF, Malboeuf CM, Birren BW, Gnirke A, Levin JZ, Sabeti PC. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol. 2014;15(11):519. doi.org/10.1186/s13059-014-0519-7.

44. Rajagopala SV, Bakhoum NG, Pakala SB, Shilts MH, Rosas-Salazar CR, Mai A, Boone HH, McHenry R, Yooseph S, Halasa N, Das SR. Metatranscriptomics to characterize respiratory virome, microbiome, and host response directly from clinical samples. Cell Rep Methods. 2021;1(6):100091. doi:10.1016/j.crmeth.2021.100091.

45. Gámbaro F, Pérez AB, Prot M, Agüera E, Baidaliuk A, Sánchez-Seco MP, Martínez-Martínez L, Vázquez A, Fernandez-Garcia MD, Simon-Loriere E. Untargeted metagenomic sequencing identifies Toscana virus in patients with idiopathic meningitis, southern Spain, 2015 to 2019. Euro Surveill. 2023;28(45):2200913. doi: 10.2807/1560-7917.ES.2023.28.45.2200913.

46. Atha DH, Ingham KC. Mechanism of precipitation of proteins by polyethylene glycols: analysis in terms of excluded volume. J Biol Chem. 1981;256(23):12108–12117. doi:10.1016/S0021-9258(18)43240-1.

47. Sanderson ND, Dingle KE, Hopkins KMV, Vaughan A, Colpus M, Parker M, Dietz EV, Gentry J, Justice A, Oakley S, Barrett L, Quan TP, Stoesser N, Eyre DW, Bejon P, Walker AS, Young BC. Clinical validation of a novel metagenomic nanopore sequencing method for detecting viral respiratory pathogens: diagnostic accuracy study. medRxiv [Preprint]. 2026. doi:10.64898/2026.02.06.26345651.

48. Wright PF, Deatly AM, Karron RA, Belshe RB, Shi JR, Gruber WC, Zhu Y, Randolph VB. Comparison of results of detection of rhinovirus by PCR and viral culture in human nasal wash specimens from subjects with and without clinical symptoms of respiratory illness. J Clin Microbiol. 2007;45(7):2126-2129. doi:10.1128/JCM.02553-06.

49. Kevill JL, Pellett C, Farkas K, Brown MR, Bassano I, Denise H, McDonald JE, Malham SK, Porter J, Warren J, Evens NP, Paterson S, Singer AC, Jones DL. A comparison of precipitation and filtration-based SARS-CoV-2 recovery methods and the influence of temperature, turbidity, and surfactant load in urban wastewater. Sci Total Environ. 2022;808:151916. doi:10.1016/j.scitotenv.2021.151916.

50. Diamanti C, Nousis L, Bozidis P, Koureas M, Kyritsi M, Markozannes G, Simantiris N, Panteli E, Koutsolioutsou A, Tsilidis K, Hadjichristodoulou C, Koutsotoli A, Christaki E, Alivertis D, Bartzokas A, Gartzonika K, Dovas C, Ntzani E. Wastewater surveillance of SARS-CoV-2: a comparison of two concentration methods. Viruses. 2024;16(9):1398. doi:10.3390/v16091398.

51. Baral R, Nwaubani DA, Solomon T, Sherchan SP. Assessing virus concentration methods for norovirus and SARS-CoV-2 detection in wastewater environments. Environments. 2026;13(2):86. doi:10.3390/environments13020086.

52. Torii S, Oishi W, Zhu Y, Thakali O, Malla B, Yu Z, Zhao B, Arakawa C, Kitajima M, Hata A, Ihara M, Kyuwa S, Sano D, Haramoto E, Katayama H. Comparison of five polyethylene glycol precipitation procedures for the RT-qPCR based recovery of murine hepatitis virus, bacteriophage phi6, and pepper mild mottle virus as a surrogate for SARS-CoV-2 from wastewater. Sci Total Environ. 2022;807(Pt 2):150722. doi:10.1016/j.scitotenv.2021.150722.

53. Thornton M, Eder G, Amman F, Pantielieieva A, Vierheilig J, Bergthaler A. Comparative wastewater virome analysis with different enrichment methods. Water Res. 2025;285:123985. doi:10.1016/j.watres.2025.123985.

54. Stepniak D, Constantinides B, Weaver M, Treagus S, Wilkinson SAJ, Quarton S, Behruznia M, Cumley N, Tyson J, McNally A, Loman NJ, Pullan S, Quick J. Library preparation strategy critically impacts RNA virus sensitivity in clinical metagenomics. medRxiv [Preprint]. 2026. doi:10.64898/2026.05.18.26353500.

ADDITIONAL FILES

File name: Additional File 1

File format: .xls

Title of data: Experimental Methods

Description of data: Summary of experimental methods used in workflow development

File name: Additional File 2

File format: .xls

Title of data: ENA Accession Numbers

Description of data: ENA Accession Numbers

File name: Additional File 3

File format: .xls

Title of data: Sequencing Reads

Description of data: Numbers of sequencing reads obtained per experiment, (one tab per experiment) together with metadata.

ABBREVIATIONS

ATCC: American Type Culture Collection

CMg: Clinical metagenomics

ENA: European Nucleotide Archive

HRE: human rhino- and enterovirus

MDA: Multiple displacement amplification

M-SAN (HQ): Medium salt active nuclease (high quality)

PEG: Polyethylene glycol

ONT: Oxford Nanopore Technologies

SAN (HQ): Salt active nuclease (high quality)

SISPA: Sequence independent single primer amplification

VTM: Viral transport medium

DECLARATIONS

RESEARCH ETHICS STATEMENT

The use of surplus material from respiratory samples processed by the Oxford University Hospitals NHS Foundation Trust Clinical Microbiology Laboratory, and the use of de-identified patient information from these samples, without patient-level consent, was approved by the Queens Square NHS Research Ethics Committee (REC), London, reference 17/LO/1420, as part of work on the development of rapid diagnostics. This REC decision (reference 17/LO/1420) waived the need to obtain consent to participate. Anonymised respiratory samples were used only post-discard, once routine clinical testing was completed. Research was performed in accordance with the Declaration of Helsinki.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

This project was funded by the NIHR Biomedical Research Centre, Oxford, and supported by the NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University in partnership with the UK Health Security Agency (UKHSA) (NIHR207379). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or UKHSA. ASW is a NIHR Senior Investigator. This work was supported by funding from the University of Oxford's John Fell Fund.

AUTHOR CONTRIBUTIONS

PB, ASW, NSt, DE, BY conceived the study.

KD, KH, DE, BY contributed to experimental design.

KD, KH, DA performed laboratory experimental work.

KH, AV, SAF, JG, LD, collected and organised samples and/or metadata.

NSa, MC, MP designed and performed bioinformatic data analysis and interpretation.

KD, PB wrote the first draft of the manuscript.

All authors critically reviewed the manuscript and approved its final submission.

DATA AVAILABILITY

All data generated or analysed during this study are included in this published article or uploaded to European Nucleotide Archive (ENA) under the accession PRJEB102181. Accession numbers are provided (Additional File 2) and also sample data and read numbers (Additional File 3).

COMPETING INTERESTS

The authors declare no competing interests.

Figure 1

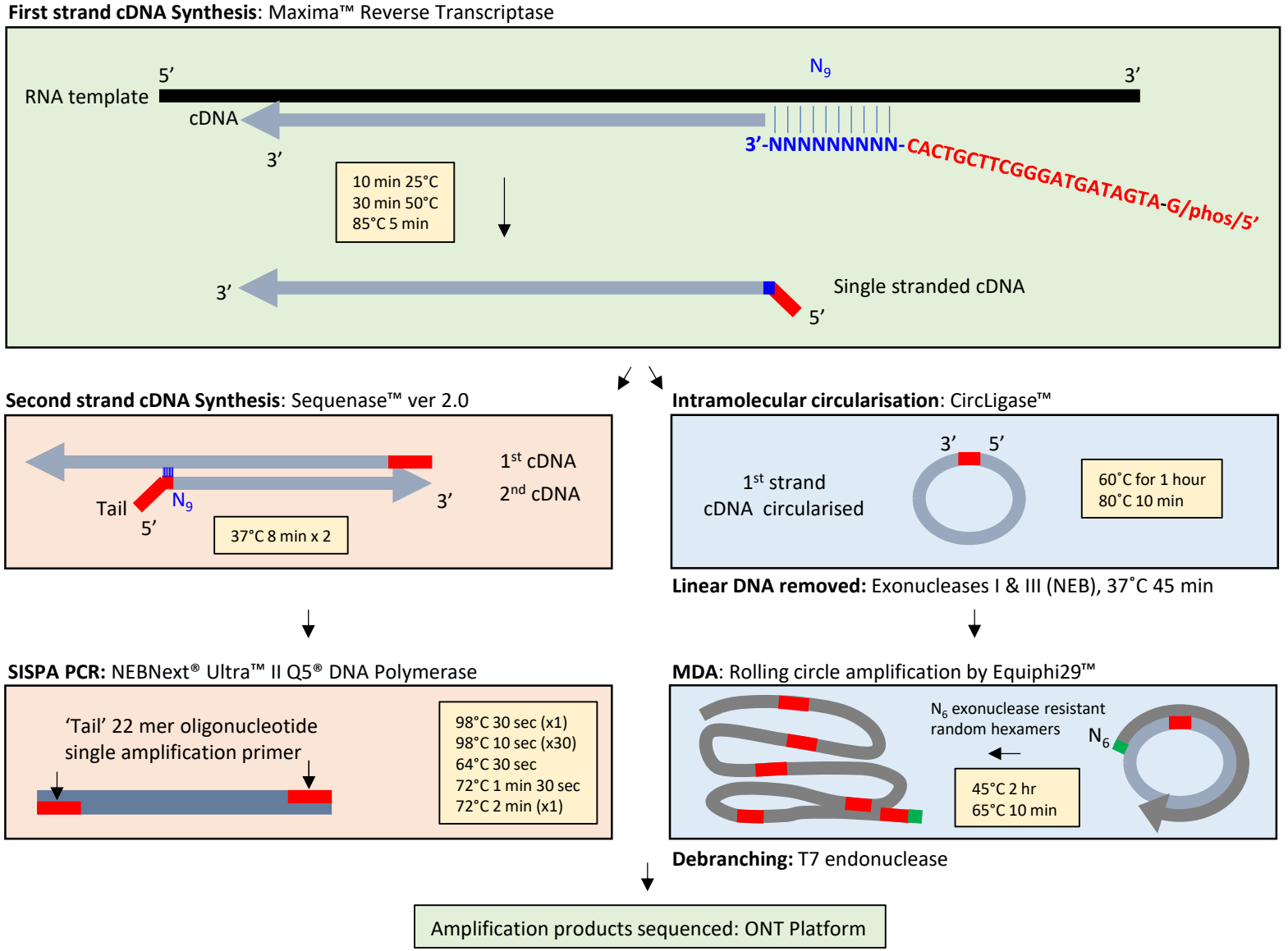

First strand cDNA Synthesis: Maxima™ Reverse Transcriptase
RNA template
5′
3′
N9
cDNA
3′
3′-NNNNNNNNN-CACTGCTTCGGGATGATAGTA-G/phos/5′
10 min 25°C
30 min 50°C
85°C 5 min
3′
5′
Single stranded cDNA
Second strand cDNA Synthesis: Sequenase™ ver 2.0
Tail
N9
5′
3′
37°C 8 min x 2
1st cDNA
2nd cDNA
Intramolecular circularisation: CircLigase™
1st strand
cDNA circularised
3′
5′
60˚C for 1 hour
80˚C 10 min
Linear DNA removed: Exonucleases I & III (NEB), 37˚C 45 min
SISPA PCR: NEBNext® Ultra™ II Q5® DNA Polymerase
'Tail' 22 mer oligonucleotide
single amplification primer
98°C 30 sec (x1)
98°C 10 sec (x30)
64°C 30 sec
72°C 1 min 30 sec
72°C 2 min (x1)
MDA: Rolling circle amplification by Equiphi29™
N6 exonuclease resistant
random hexamers
N6
45°C 2 hr
65°C 10 min
Debranching: T7 endonuclease
Amplification products sequenced: ONT Platform

Figure 2

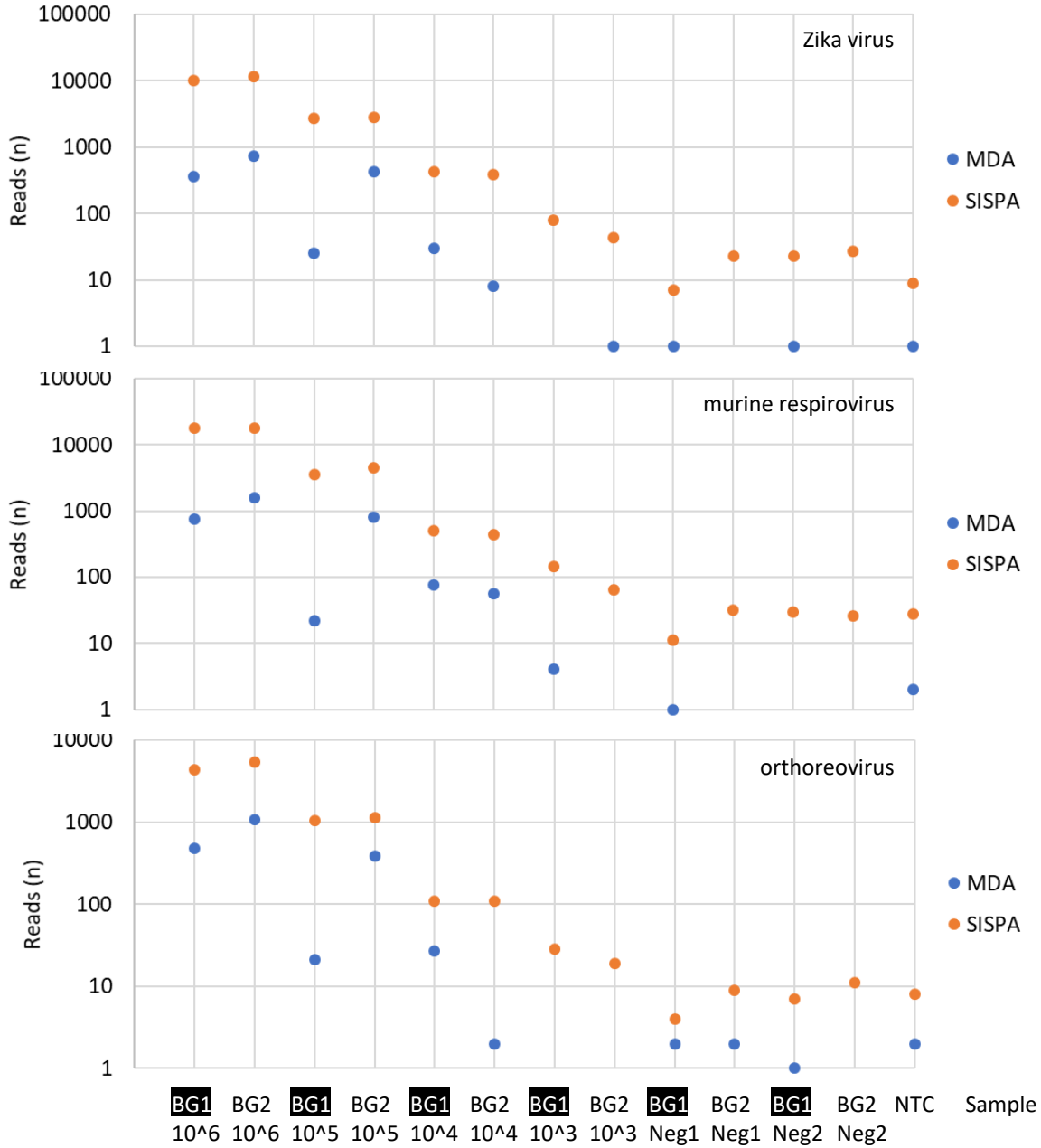

Zika virus
murine respirovirus
orthoreovirus
Reads (n)
100000
10000
1000
100
10
1
MDA
SISPA
BG1 10^6
BG2 10^6
BG1 10^5
BG2 10^5
BG1 10^4
BG2 10^4
BG1 10^3
BG2 10^3
BG1 Neg1
BG2 Neg1
BG1 Neg2
BG2 Neg2
NTC
Sample

Figure 3

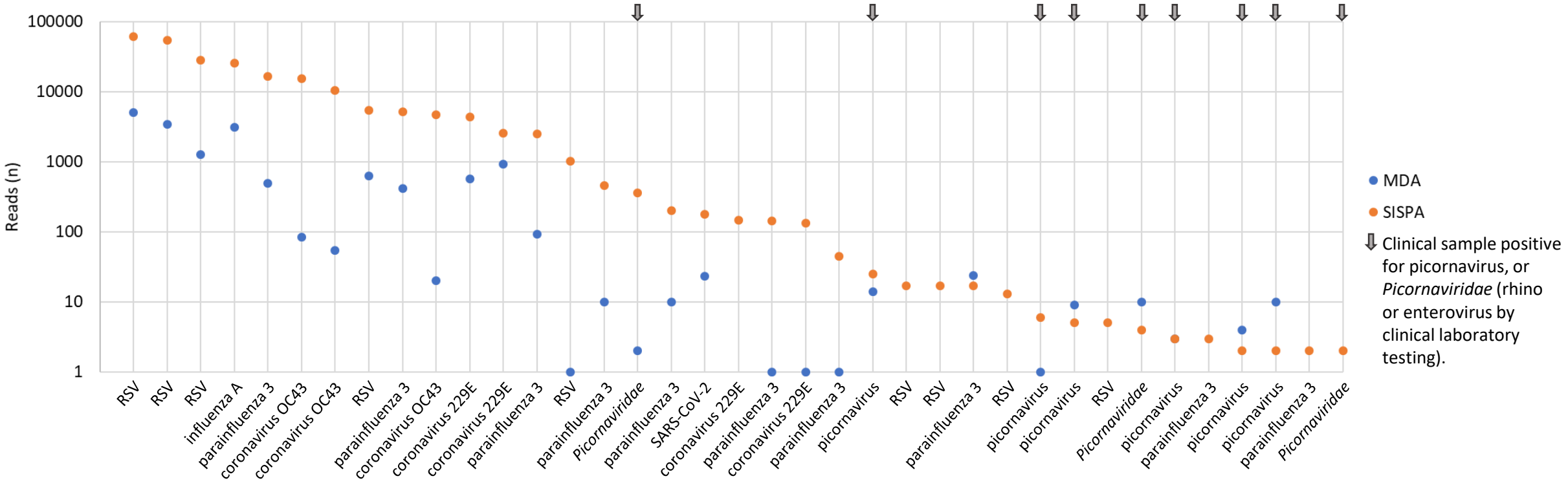
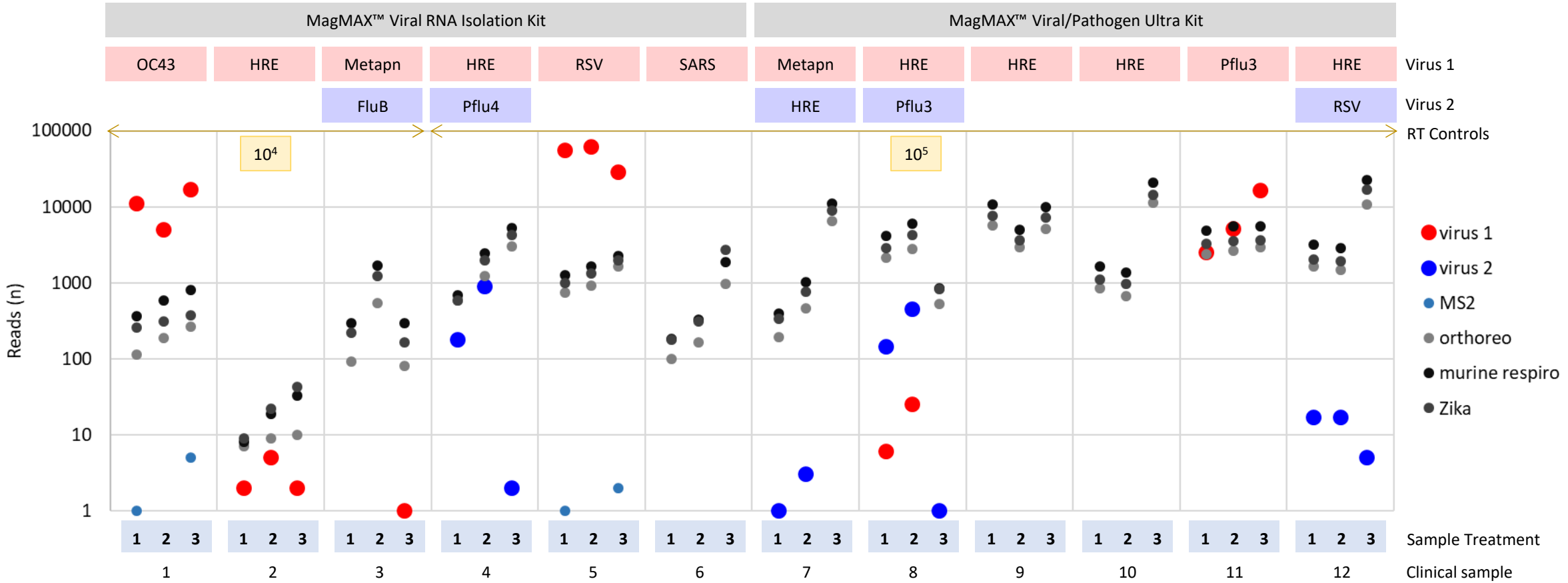

100000
10000
1000
100
10
1
Reads (n)
RSV
RSV
RSV
influenza A
parainfluenza 3
coronavirus OC43
coronavirus OC43
RSV
parainfluenza 3
coronavirus OC43
coronavirus 229E
coronavirus 229E
parainfluenza 3
RSV
parainfluenza 3
Picornaviridae
parainfluenza 3
SARS-CoV-2
coronavirus 229E
parainfluenza 3
coronavirus 229E
parainfluenza 3
picornavirus
RSV
RSV
parainfluenza 3
RSV
picornavirus
picornavirus
RSV
Picornaviridae
picornavirus
parainfluenza 3
picornavirus
picornavirus
parainfluenza 3
Picornaviridae
MDA
SISPA
Clinical sample positive for picornavirus, or Picornaviridae (rhino or enterovirus by clinical laboratory testing).

Figure 4

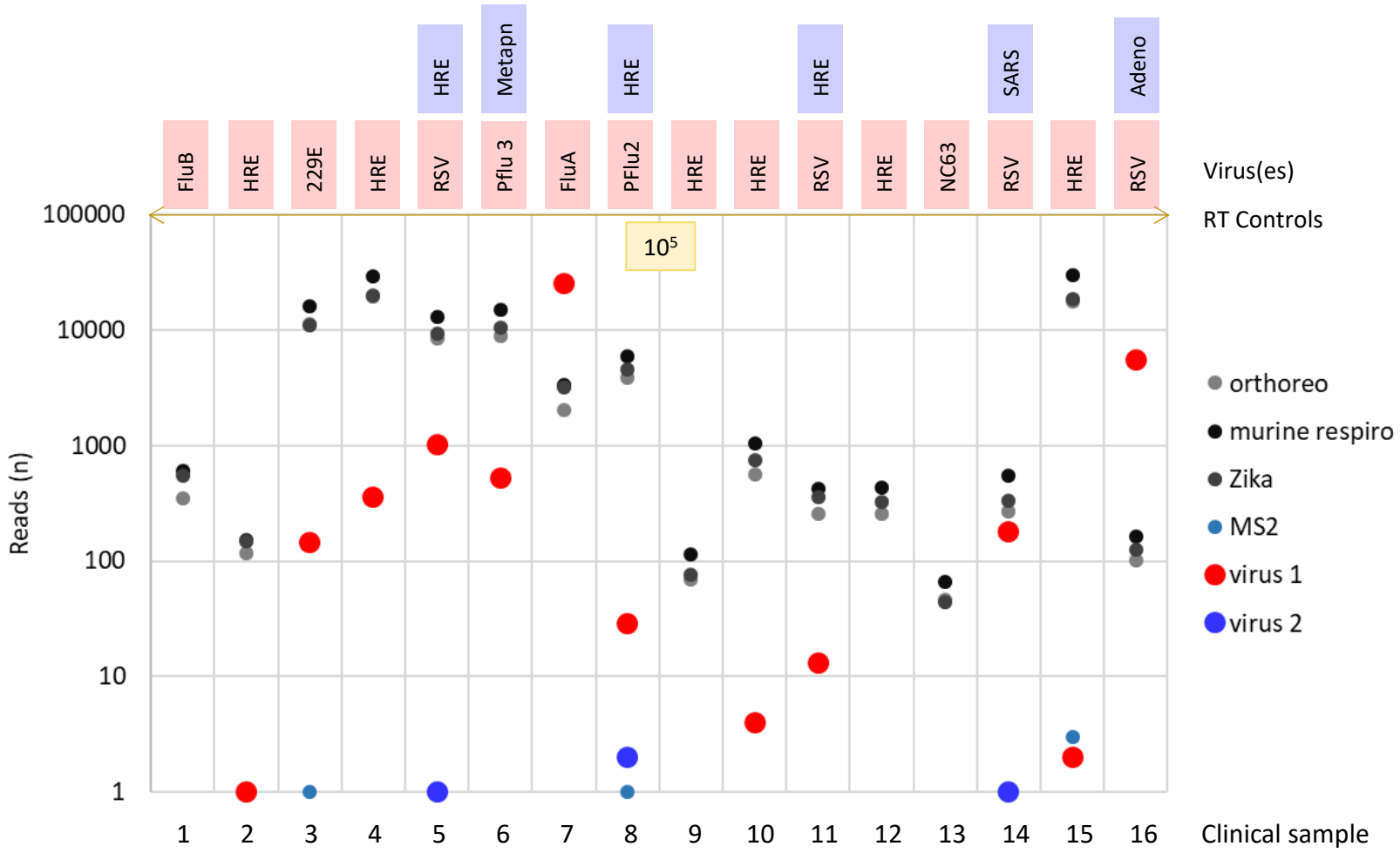


SAMPLE TREATMENT
1. Neat clinical sample
2. Centrifugation Pellet – 5,000g 30 min
3. Centrifugation Supernatant – 5,000g 30min

OC43: coronavirus OC43
SARS: SARS-CoV-2
FluA: influenza A
FluB: influenza B

Pflu3: parainfluenza virus 3
Pflu4: parainfluenza virus 4
HRE: human rhino and enteroviruses
Metapn: human metapneumovirus
RSV: respiratory syncytial virus

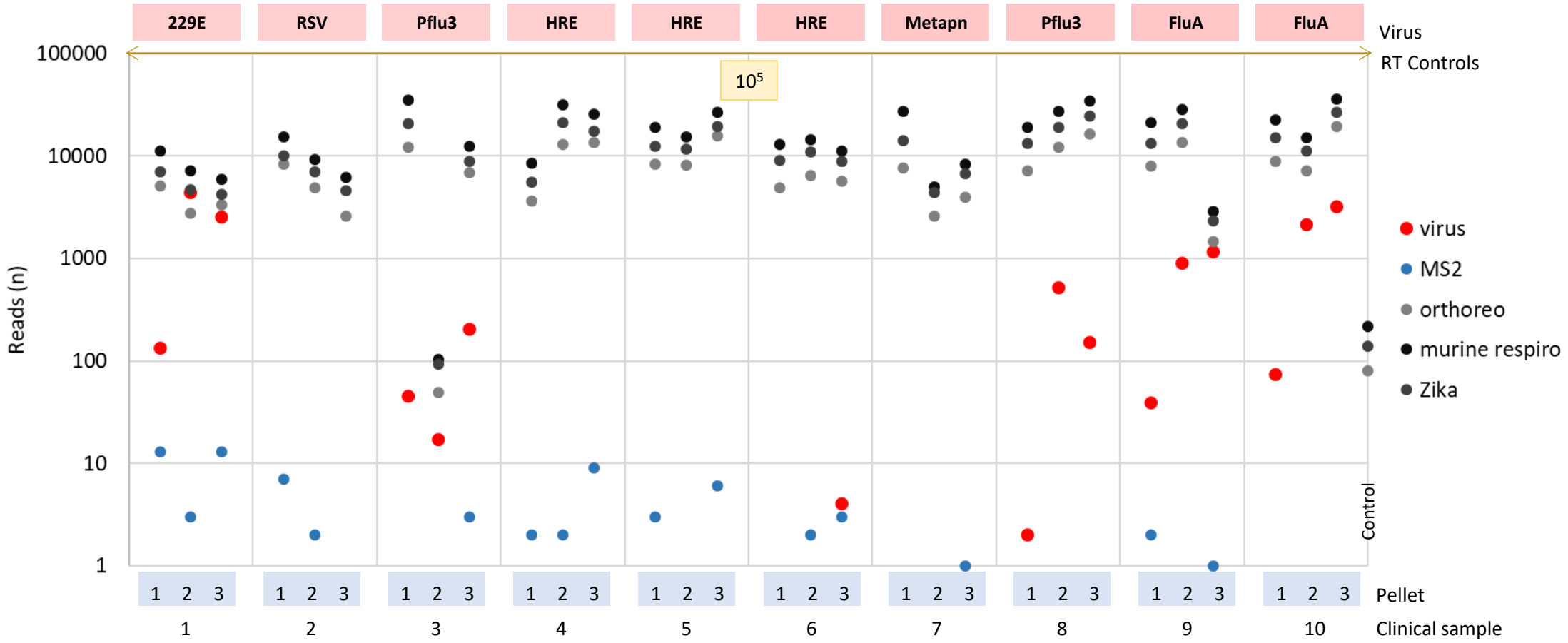
Figure 5



Viruses detected by Clinical Lab PCR
Pflu2: parainfluenza virus 2
Pflu3: parainfluenza virus 3
HRE: human rhino and enteroviruses
Metapn: human metapneumovirus
RSV: respiratory syncytial virus
FluA: influenza A
FluB: influenza B
OC43: coronavirus OC43
NC63: coronavirus NC63
SARS: SARS-CoV-2
Adeno: adenovirus

SAMPLE TREATMENT
Pellet – Centrifugation 5,000g, 30 min

Figure 6



SAMPLE TREATMENT
1ml VTM processed to 3 fractions
Pellet 1: Centrifugation 3,200g, 30 min.
Pellet 2: Centrifugation 13,200g, 30 min.
Pellet 3: Precipitation & Centrifugation 13,200g, 30 min.

FluA: influenza A
229E: coronavirus 229E
Pflu3: parainfluenza virus 3
HRE: human rhino and enteroviruses
Metapn: human metapneumovirus
RSV: respiratory syncytial virus

Figure 7

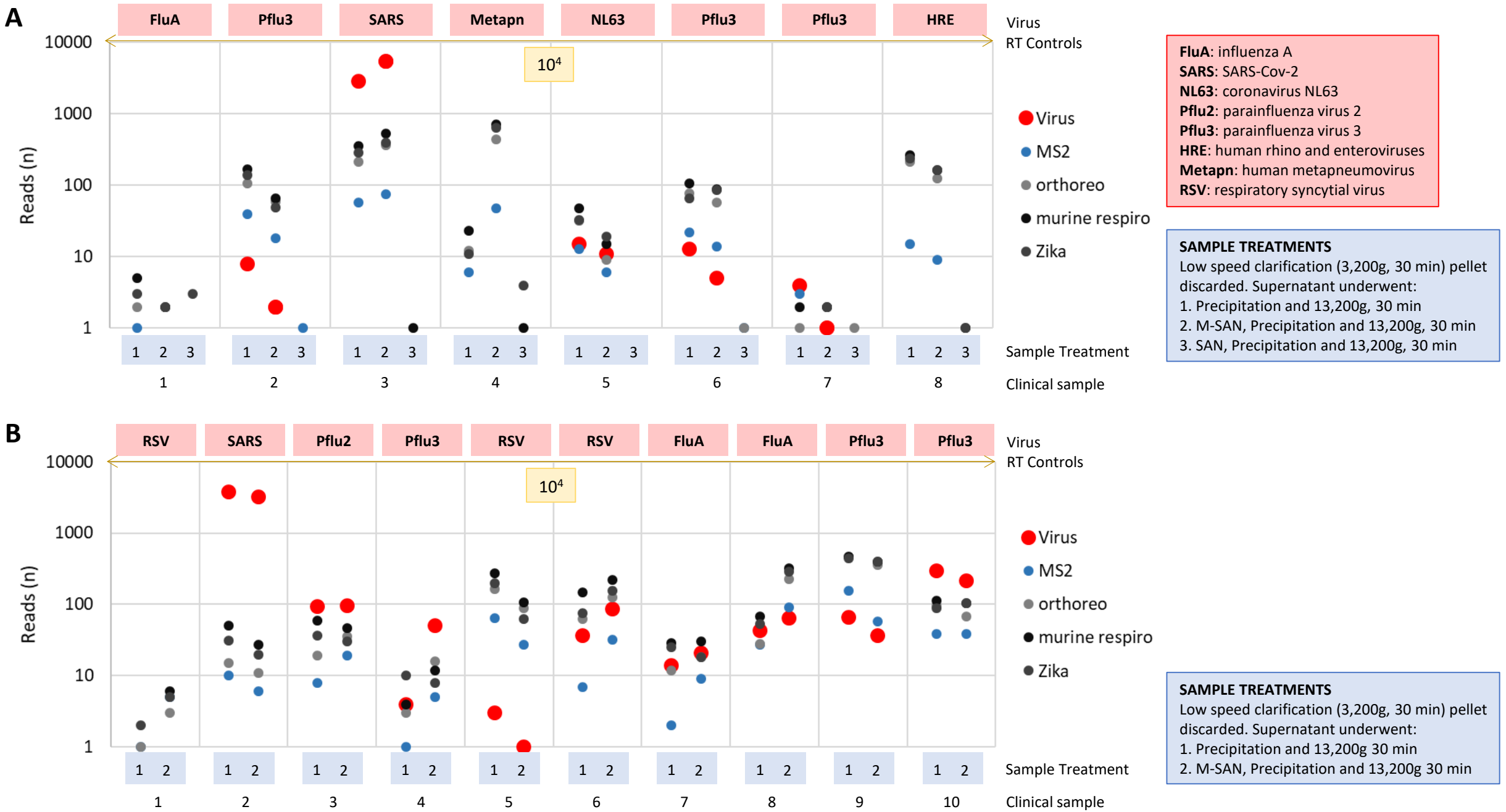
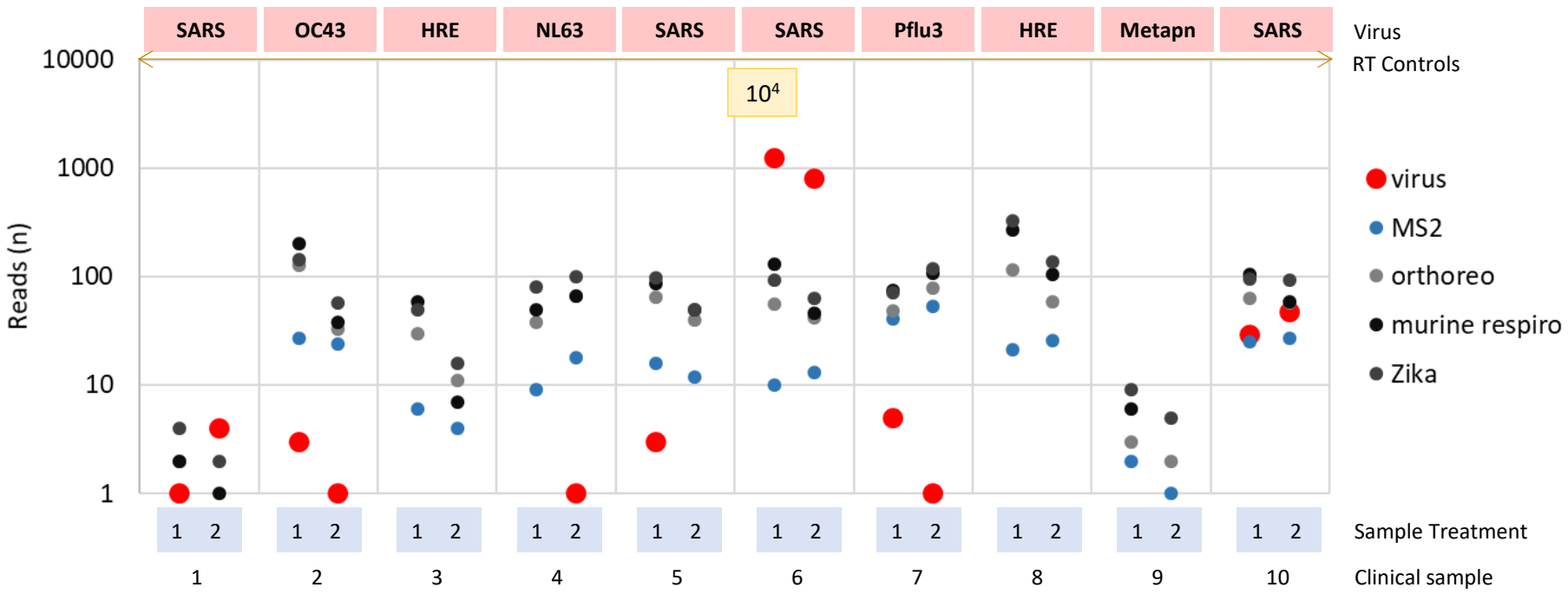

A
FluA
Pflu3
SARS
Metapn
NL63
Pflu3
Pflu3
HRE
Virus
RT Controls
10000
1000
100
10
1
Reads (n)
10⁴
Virus
MS2
orthoreo
murine respiro
Zika
1 2 3
Sample Treatment
1 2 3 4 5 6 7 8
Clinical sample
FluA: influenza A
SARS: SARS-Cov-2
NL63: coronavirus NL63
Pflu2: parainfluenza virus 2
Pflu3: parainfluenza virus 3
HRE: human rhino and enteroviruses
Metapn: human metapneumovirus
RSV: respiratory syncytial virus
SAMPLE TREATMENTS
Low speed clarification (3,200g, 30 min) pellet discarded. Supernatant underwent:
1. Precipitation and 13,200g, 30 min
2. M-SAN, Precipitation and 13,200g, 30 min
3. SAN, Precipitation and 13,200g, 30 min
B
RSV
SARS
Pflu2
Pflu3
RSV
RSV
FluA
FluA
Pflu3
Pflu3
Virus
RT Controls
10000
1000
100
10
1
Reads (n)
10⁴
Virus
MS2
orthoreo
murine respiro
Zika
1 2
Sample Treatment
1 2 3 4 5 6 7 8 9 10
Clinical sample
SAMPLE TREATMENTS
Low speed clarification (3,200g, 30 min) pellet discarded. Supernatant underwent:
1. Precipitation and 13,200g 30 min
2. M-SAN, Precipitation and 13,200g 30 min

Figure 8



SAMPLE TREATMENTS
Low speed clarification (3,200g, 30 min) pellet discarded.
Supernatant undergoes:
1. M-SAN, Precipitation and 13,200g 30 min
2. Bead beat, M-SAN, Precipitation and 13,200g 30 min

SARS: SARS-Cov-2
OC43: coronavirus OC43
NL63: coronavirus NL63
Pflu3: parainfluenza virus 3
HRE: human rhino and enteroviruses
Metapn: human metapneumovirus

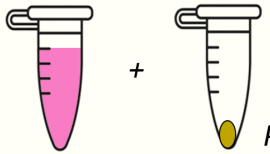
Figure 9

SAMPLE FRACTIONATION



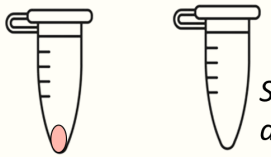
1ml VTM

CLARIFICATION: ↓ 3,200g 30 min



Pellet discarded

NUCLEASE: 37°C, 1 hour
PRECIPITATION: virus precipitation reagent
CENTRIFUGATION: ↓ 13,200g 30 min



Supernatant discarded

RNA EXTRACTION

+ PBS + MS2 extraction control
+270µl TRI Reagent



Direct-zol RNA Microprep kit (Zymo)

+ RNA RT/amplification controls 10^4 genomes:
Zika, murine respirovirus, orthoreovirus

cDNA SYNTHESIS

Random (N_9) primed, incorporating 'tail'
5'-GATGATAGTAGGGCTTCGTCAC-N_9-3'

ss cDNA
Maxima Reverse transcriptase™

ds cDNA
Sequenase™

AMPLIFICATION

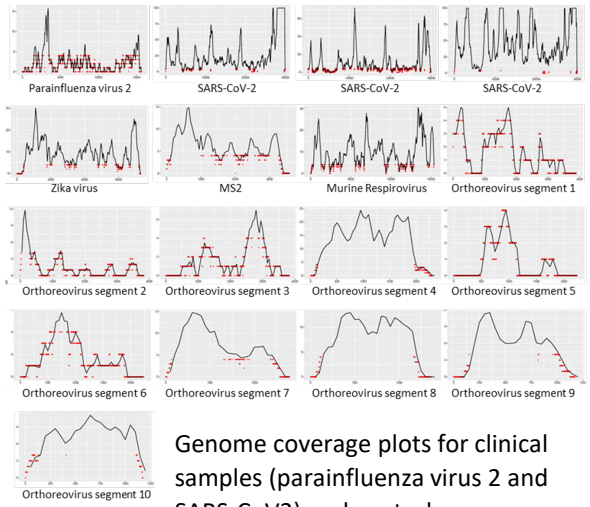
SISPA PCR
NEBNext® Ultra™ II Q5®
5'-GATGATAGTAGGGCTTCGTCAC-3'

SEQUENCING

Oxford Nanopore Technologies (ONT) platform
Rapid sequencing V14 Amplicon Sequencing (SQK-RBK114.96), R10.4.1 flow cell

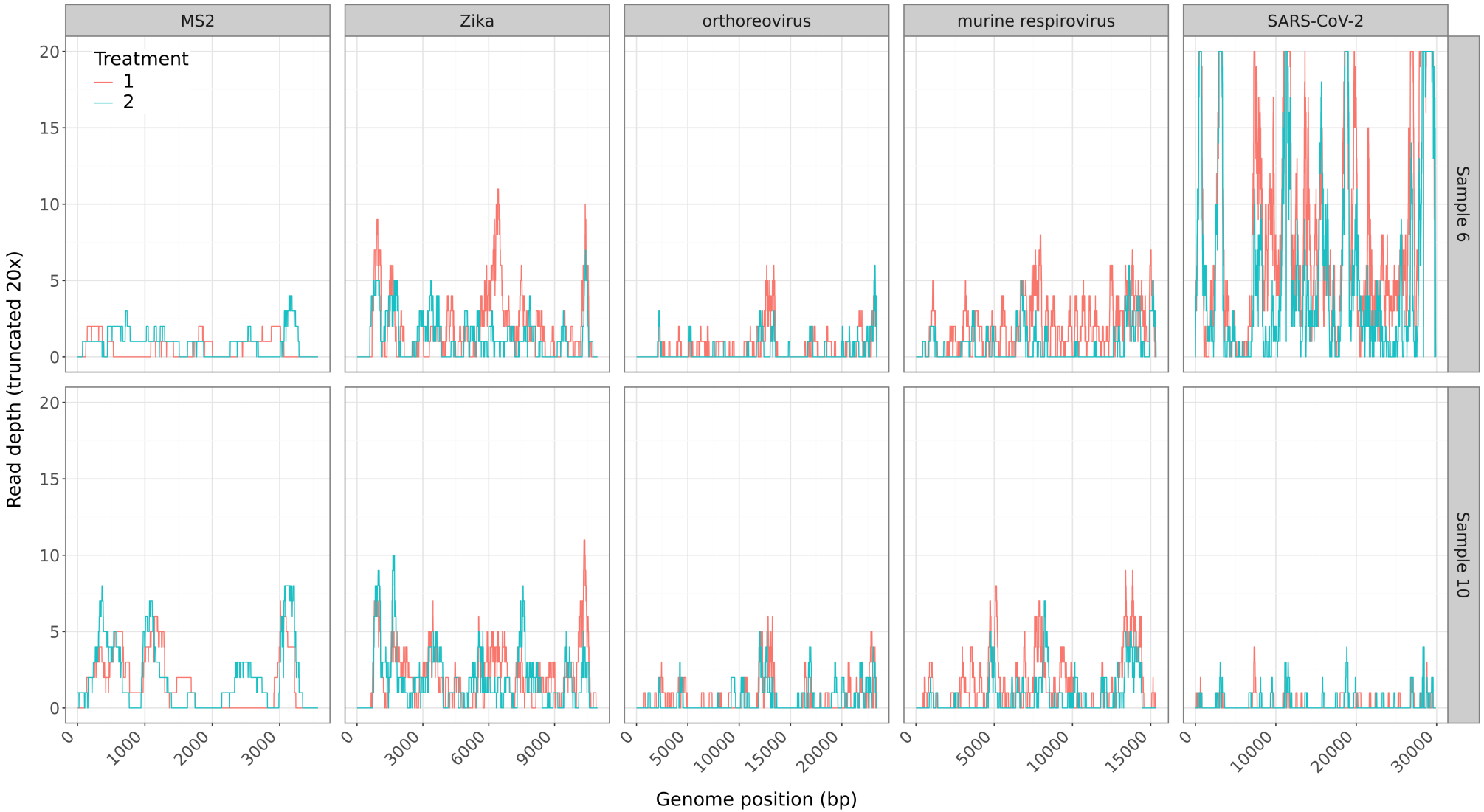
BIOINFORMATICS

Reads classified taxonomically with Kraken2 and mapped to panel of respiratory virus genomes using minimap2.



Genome coverage plots for clinical samples (parainfluenza virus 2 and SARS-CoV2) and controls.

Figure 10


MS2
Zika
orthoreovirus
murine respirovirus
SARS-CoV-2
Sample 6
Sample 10
Treatment
1
2
Read depth (truncated 20x)
Genome position (bp)
0
5
10
15
20
0
1000
2000
3000
0
3000
6000
9000
0
5000
10000
15000
20000
0
5000
10000
15000
0
10000
20000
30000

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1ExperimentalMethods.xlsx
- AdditionalFile2ENAAccessionNumbers.xlsx
- AdditionalFile3SequencingReads.xlsx