Comparison of the performance of two targeted metagenomic virus capture probe-based methods using reference control materials and clinical samples

Mourik, K.; Sidorov, I.; Carbo, E.C.; Meer, D. van der; Boot, A.; Kroes, A.C.M.; ... ; Vries, J.J.C. de

Journal of Clinical Microbiology

Virology | Full-Length Text

Comparison of the performance of two targeted metagenomic virus capture probe-based methods using reference control materials and clinical samples

Kees Mourik,[1] Igor Sidorov,[1] Ellen C. Carbo,[1] David van der Meer,[2] Arnoud Boot,[2] Aloysius C. M. Kroes,[1] Eric C. J. Claas,[1] Stefan A. Boers,[1] Jutte J. C. de Vries[1]

AUTHOR AFFILIATIONS See affiliation list on p. 10.

ABSTRACT Viral enrichment by probe hybridization has been reported to significantly increase the sensitivity of viral metagenomics. This study compares the analytical performance of two targeted metagenomic virus capture probe-based methods: (i) SeqCap EZ HyperCap by Roche (ViroCap) and (ii) Twist Comprehensive Viral Research Panel workflow, for diagnostic use. Sensitivity, specificity, and limit of detection were analyzed using 25 synthetic viral sequences spiked in increasing proportions of human background DNA, eight clinical samples, and American Type Culture Collection (ATCC) Virome Virus Mix. Sensitivity and specificity were 95% and higher for both methods using the synthetic and reference controls as gold standard. Combining thresholds for viral sequence read counts and genome coverage [respectively 500 reads per million (RPM) and 10% coverage] resulted in optimal prediction of true positive results. Limits of detection were approximately 50–500 copies/mL for both methods as determined by ddPCR. Increasing proportions of spike-in cell-free human background sequences up to 99.999% (50 ng/mL) did not negatively affect viral detection, suggesting effective capture of viral sequences. These data show analytical performances in ranges applicable to clinical samples, for both probe hybridization metagenomic approaches. This study supports further steps toward more widespread use of viral metagenomics for pathogen detection, in clinical and surveillance settings using low biomass samples.

IMPORTANCE Viral metagenomics has been gradually applied for broad-spectrum pathogen detection of infectious diseases, surveillance of emerging diseases, and pathogen discovery. Viral enrichment by probe hybridization methods has been reported to significantly increase the sensitivity of viral metagenomics. During the past years, a specific hybridization panel distributed by Roche has been adopted in a broad range of different clinical and zoonotic settings. Recently, Twist Bioscience has released a new hybridization panel targeting human and animal viruses. This is the first report comparing the performance of viral metagenomic hybridization panels.

KEYWORDS viral metagenomics, capture probes, targeted metagenomics, viral diagnostics

Viral metagenomics has been gradually applied for broad-spectrum pathogen detection of infectious diseases (1–5), surveillance of emerging diseases (3, 6–8), and pathogen discovery (9, 10). Though metagenomic approaches have been practiced for decades in the field of marine environments and the human microbiome, this approach is nowadays changing how physicians diagnose infectious diseases (2). Whereas amplicon-based metagenomics has been successfully adopted for bacterial diagnostics, no amplicon-based pan-viral approach is available, and therefore, viral

Editor Kirsten St. George, Wadsworth Center - NYSDOH, Albany, New York, USA

Address correspondence to Kees Mourik, k.mourik@lumc.nl, or Jutte J. C. de Vries, jjcdevries@lumc.nl.

D.V.D.M. and A.B. are employees of GenomeScan B.V. and provided the sequencing service. They were not involved in bioinformatic/statistical data analysis or interpretation of results.

See the funding table on p. 11.

Received 28 February 2024
Accepted 5 April 2024
Published 17 May 2024

metagenomics has not yet been widely deployed as a diagnostic tool in clinical laboratories (1). One of the main challenges for application in clinical settings is the low level of viral genomes in the presence of high levels of host material in patient samples. Several methods for depletion of host sequences and enrichment of viral sequences have been studied with varying success rates (11). For example, depletion of host cells before the extraction of nucleic acids (NAs) has been reported to be not advantageous in clinical samples as also intracellular viral particles or NA will be removed (11, 12). In contrast, viral enrichment by probe hybridization methods has been reported to significantly increase sensitivity in various sample types (1, 13–19), up to the level required for accurate detection of low-frequency virus variants (20).

Previously, the performance of a hybridization capture probe panel targeting vertebrate viruses in cerebrospinal fluids from patients with meningo-encephalitis has been analyzed (21). Viral target sequence read counts increased 100–10,000 fold compared to unenriched metagenomic sequencing, and sensitivity by enrichment was comparable with polymerase chain reaction (PCR) (21). Moreover, these earlier data showed that this hybridization panel of approximately 2 million capture probes designed in 2015 was suitable for the detection of novel coronaviruses by reactivity with other vertebrate beta coronavirus probes (10). During the past years, the SeqCap EZ Hypercap (Virocap) hybridization panel distributed by Roche has been adopted in a broad range of different clinical (15, 22–29) and zoonotic settings (22, 29, 30). Recently, Twist Bioscience has released a new Comprehensive Viral hybridization panel containing approximately 1 million capture probes targeting human and animal viruses. Reports comparing the performance of viral metagenomic hybridization panels are lacking.

Here, we compare the analytical performance of two targeted metagenomic virus capture probe-based methods: (i) SeqCap EZ HyperCap by Roche (ViroCap) and (ii) Twist Comprehensive Viral Research Panel workflow. Sensitivity, specificity, and limit of detection (LOD) were analyzed using reference control materials and clinical samples.

MATERIALS AND METHODS

(Reference) control materials and clinical samples

An overview of the validation panels and study design is shown in Fig. 1.

In order to mimic the complexity of clinical samples while reducing the number of additional viral sequences, enabling sensitivity and specificity analyses, a panel was prepared of synthetic ssRNA and ssDNA viral sequences (10^0–10^7 copies/mL) spiked in human cell-free DNA background sequences (cfDNA, Twist pan-cancer reference standard set, Twist Bioscience, San Francisco, USA). Synthetic viral sequences covering >99.9% of the viral genomes of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) B1.1429 Epsilon 2020 (ssRNA, template: EPI_ISL_672365), influenza A virus H1N1 2009 (Inf A, ssRNA, NC_026438), measles (ssRNA, NC_001498.1), enterovirus D68 (ssRNA, NC_038308.1), and bocavirus 1 (ssDNA, MG953830.1), in non-overlapping fragments of maximal 5 kB with 50 bp gaps for biosafety reasons according to the manufacturer's policy. Viral sequences were mixed with a range of proportions of human cfDNA; 90–99.99990% of weight (up to 5 ng per 100 uL), corresponding with proportions of 10–0.0001% (down to 0.00004 pg per 100 uL, see Table S1) of viral nucleotides based on the reported abundancy in different types of low biomass clinical samples (12, 31–33). Concentrations of viral sequences were determined by digital droplet (reverse transcriptase) PCR in triplicate (BioRad QX200). A total of 25 synthetic mixtures were constructed (Table S1).

Furthermore, a dilution of ATCC Virome Virus Mix (MSA-2008, ATCC, Manassas, USA) of cultivated adenovirus type F (ADV, dsDNA), cytomegalovirus (CMV, dsDNA), respiratory syncytial virus (RSV, ssRNA), influenza B virus (ssRNA), reovirus 3 (dsRNA), and zika virus (ssRNA) was included.

In addition, clinical EDTA plasma samples (n = 8) with dsDNA viruses, previously submitted to the Clinical Microbiological Laboratory for routine diagnostic testing and

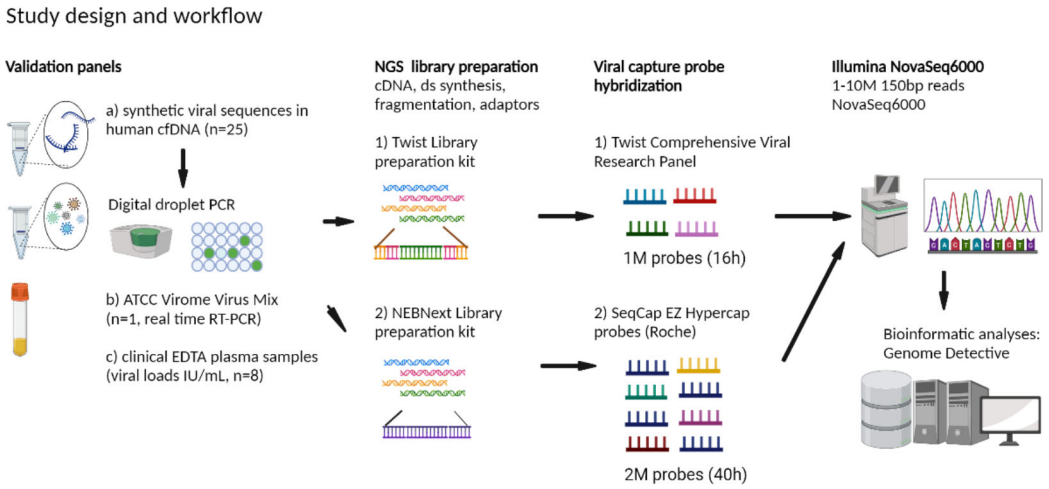

FIG 1 Workflows of the capture probe-based targeted metagenomic protocols compared in this study, Twist Comprehensive Viral Research, and the SeqCap EZ HyperCap (ViroCap, Roche) and, both in combination with identical bioinformatic analyses pipeline. Created using BioRender.

tested positive by qPCR (21, 34) for ADV, Epstein–Barr virus (EBV) and Hepatitis B virus (HBV), were included in the comparison. Viral loads (VLs) selected ranged from 500 to 50,000 International Units (IU)/mL and 10-fold and 100-fold dilutions in elution buffer were prepared (Table S1).

NA extraction

In total, 200 uL of clinical samples and diluted ATCC whole virus mixture were subjected to extraction of total NAs after spike-in of internal controls equine arteritis virus (EAV) and phocid herpes virus (PhHV), using the MagNAPure 96 DNA and Viral NA Small volume extraction kit (Roche, Basel, Switzerland) with an elution volume of 100 uL, as described previously (21).

Viral metagenomic next-generation sequencing (mNGS)

Twist comprehensive viral research panel workflow

Sample preparation was performed using the Comprehensive Viral Research Panel workflow (Twist Bioscience Corp.) according to the manufacturer's instructions. In short, 5 µL of NA was used as input for cDNA synthesis (Protoscript, New England Biolabs, Inc.) followed by purification using magnetic beads, enzymatic fragmentation for 15 min at 30°C, end repair and dA-tailing (Twist EF Library Prep 2.0, Twist Bioscience Corp.). Next, unique molecular identifier (UMI) adapters with unique dual barcodes (Twist UMI Adapter System, Twist Bioscience Corp.) were ligated to the fragments and amplified using PCR (12 cycles). Amplified libraries were pooled equimolar per eight samples and library pools were used for hybridization with the Twist Comprehensive Virus probe panel, consisting of ~1 million 120 bp probes targeting 15,488 different viral strains infecting humans and animals. Hybridization was performed for 16 h of incubation followed by several wash steps. Captured fragments were further amplified by a post-hybridization PCR (15 cycles). Finally, captured libraries were purified by a bead clean up using AmpureXP, and quantity and fragment size were determined using Qubit (Thermo Fisher, Waltham, MA, USA) and Fragment Analyser (Agilent, Santa Clara, CA, USA), respectively. Libraries were clustered, and approximately 1 million 150 bp paired-end reads were generated per sample, according to manufacturer's protocols (Illumina Inc.) at GenomeScan B.V. using the NovaSeq6000.

SeqCap EZ HyperCap (ViroCap design, Roche)

The SeqCap EZ HyperCap workflow (Roche, Madison, USA) was performed as validated and described previously (10, 21, 35). Briefly, 5 uL of NA was used as direct input (without concentration step) for enzymatic fragmentation and cDNA synthesis using the NEBNext Ultra II Directional RNA Library preparation kit V3.0 (New England Biolabs, Ipswich, MA, USA) for Illumina with several in-house adaptations (omission of the poly A mRNA capture isolation, rRNA depletion and DNase steps) to enable simultaneous detection of both DNA and RNA in a single tube per sample (12, 32). After purification, dual barcodes (NEBNext Multiplex oligos for Illumina 96 unique dual index primer pairs) were attached to the fragments and amplified using PCR (21 cycles). Four barcoded samples including controls were pooled, Cot (enriched for repetitive sequences) human DNA and HyperCap Universal Blocking Oligos were added before purification, following incubation for >40 h with the SeqCap EZ HyperCap v1 [ViroCap design, 2015 (14)], a collection of approximately 2 million oligonucleotide probes (70–120 mers) targeting all known vertebrate viruses. A complete list of the viral taxa included can be found in the supplementary tables of the manuscript by Briese *et al*. (14). Captured fragments were further amplified by a post-hybridization PCR (14 cycles). Finally, captured libraries were purified by bead clean up using AmpureXP, and quantity and fragment sizes were determined using Qubit and Fragment Analyser, respectively. Approximately 10 million 150 bp paired-end reads were sequenced per sample according to the manufacturer's protocols (Illumina Inc.) at GenomeScan B.V. using the NovaSeq6000.

Data analysis

Bioinformatic analysis

Image analysis, base calling, and quality check of sequence data were performed with the Illumina data analysis pipelines RTA3.4.4 and bcl2fastq v2.20 (Illumina). Sequence data obtained using both probe capture metagenomics methods were analyzed using a previously validated (10, 21, 36, 37) bioinformatics pipeline. After quality pre-processing and removal of human reads (by mapping them to the human reference genome GRCh38 [https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/ with Bowtie2 (38) version 2.3.4], data sets were analyzed using Genome Detective (39) version 2.48 (accessed April–May 2023) as described previously (36). Genome Detective includes *de novo* assembly and both nucleotide- and amino acid-based classification in combination with a RefSeq/Swiss-Prot Uniref database by Genome Detective (39).

Read counts were normalized for total read count and genome size using the formula: reads per kilobase per million (RPKM) = (number of reads mapped to the virus genome Y * 10^6) / (total number of reads * length of the genome in kB) (37). To enable the analyses of the percentage of genome coverage per 1 million total reads, 1 million raw reads were randomly selected (32) from the 10 million reads generated for the Roche protocol. The random selection from raw FASTQ files was performed with the seqtk tool (https://github.com/lh3/seqtk, version 1.3).

Performance metrics and statistical analyses

Sensitivity and specificity were calculated using the results from the synthetic viral sequences and the ATCC Virome Virus mix. Synthetic and reference material controls were considered as gold standard; however, additional findings were considered false positives, and non-vertebrate viral detections were excluded from analyses. Receiver Operating Characteristic (ROC) curves were generated by varying the number of sequence-read counts used as cut-off for defining a positive result, given a prerequisite of ≥3 genome regions covered (40), and area under the curves (AUCs) were calculated. Spearman correlations of sequence read count with VL, as determined by qPCR and ddPCR, were analyzed.

Limits of detection for both methods were determined by ddPCR using 10-fold serial dilutions of synthetic viral NA in human cfDNA background (see Table S1), and by qPCR

for clinical samples (selected based on VLs with 10-fold differences). Reproducibility was determined by analyzing the coefficient of variance (CV) between runs (standard deviation/mean * 100).

Statistical analyses were performed using SPSS version 25 and 29. Statistics with *P*-values of 0.05 and lower were considered significant.

RESULTS

Analytic sensitivity, specificity, and ROC

Detection of synthetic viral sequences in human cfDNA background, and the Virome Virus Mix, using the Twist Comprehensive Viral Research Panel and the SeqCap EZ HyperCap (ViroCap) workflow is depicted in Table S1 (sheet 1). For both methods, sensitivity was 100% (23/23 target viruses, cycle threshold, C_T, values ranging from 20 to 32). Viral target read counts ranged from 334 to 872,042 RPM for the Twist Comprehensive Viral Research workflow and 2,171–971,610 RPM for the SeqCap EZ HyperCap workflow. Genome coverage ranged from 91.1% to 100% (median 99.8%), and 8.4% to 100% (median 97.5%), for these respective methods. The presence of multiple whole viruses in different concentrations in the Virome Virus mix did not hinder their detection, for both methods. Sensitivity and specificity were calculated for different thresholds for defining a positive result: (i) sequence read counts and (ii) genome coverage percentage, as depicted in Fig. 2 and Tables 1 and 2. For the calculation of the percentage of the viral genomes covered, a random selection of 1 million sequence reads per data set was used. Figure 2 shows that both RPM and genome coverage were distinctive parameters for defining a true positive result, with AUC of 99.8% for both methods when considering RPM as a parameter, and ≥99.7% when considering genome coverage as a parameter. Sensitivity and specificity scores of ≥95% were accomplished for both methods when 500 RPM was set as threshold, on top of a prerequisite of a minimum of three distributed regions of the genome being covered (Table 1). Similarly, when coverage was set at 10% of the genome, both methods reached sensitivity and specificity levels of 95% and higher (Table 2). Increasing the threshold for genome coverage resulted in decreased

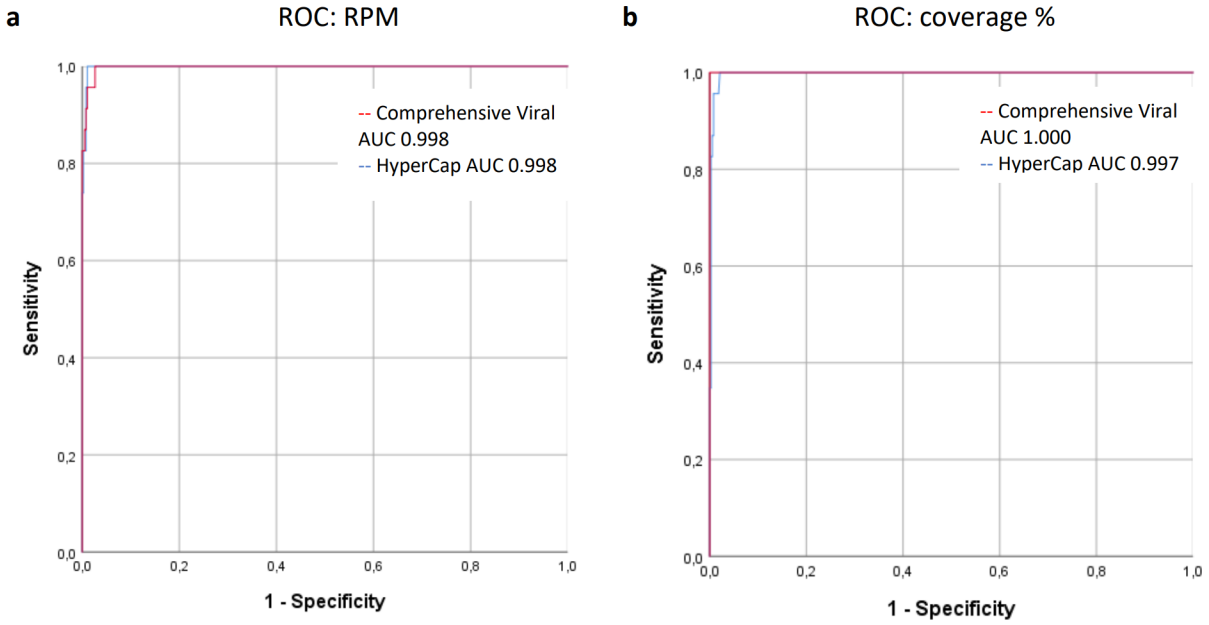

FIG 2 ROC curves for the prediction of detection of viral sequences using the virus capture probe-based metagenomic workflows Twist Comprehensive Viral Research, and SeqCap EZ HyperCap (Roche). The validation panel consisted of synthetic viral sequences spiked in a background of human cell-free DNA (90%–99.92%) and diluted ATCC Virome virus mix standard (copies/mL ranging from 10^4 to 10^7). (A) ROC based on varying threshold of sequence RPM, and (B), based on varying threshold of percentage of genome coverage for defining a positive result, using a random selection of 1 million sequence reads per data set. For all curves (A and B), a minimum of three distributed regions of the genome covered was set as primary parameter for defining detection.

TABLE 1 Sensitivity and specificity resulting from varying thresholds based on sequence read counts using a random selection of 1 million sequence reads per data set, for the capture of probe-based metagenomic workflows SeqCap EZ HyperCap (Roche) and Twist Comprehensive Viral Research Panel workflow[a]

Thresholds based on read counts				
	50 RPM	**500 RPM**	**5,000 RPM**	**50,000 RPM**
Twist Comprehensive Viral Research				
Sensitivity	1.000	0.957	0.870	0.739
Specificity	0.960	0.979	0.995	1.000
Corresponding LOD (c/mL)[b]	RNA: 10^{1}	RNA: 10^{1}	RNA: 10^{1}	RNA: $10^{2\text{-}4}$
	DNA: 10^{2}	DNA: $10^{2\text{-}3}$	DNA: $10^{2\text{-}3}$	DNA: 10^{4}
SeqCap EZ HyperCap workflow (Roche)				
Sensitivity	1.000	1.000	0.913	0.739
Specificity	0.976	0.984	0.992	0.997
Corresponding LOD (c/mL)[b]	RNA viruses: 10^{2}	RNA: 10^{2}	RNA: 10^{2}	RNA: $10^{2\text{-}4}$
	DNA viruses: $10^{2\text{-}3}$	DNA: $10^{2\text{-}3}$	DNA: $10^{2\text{-}4}$	DNA: $10^{4\text{->}}$

[a]The corresponding ROC is shown in Fig. 2. A minimum of three distributed regions of the genome covered was set as primary parameter for defining detection.
[b]Based on LOD of Inf A, SARS-CoV-2, EBV, HBV, and bocavirus.

sensitivity for the SeqCap EZ HyperCap workflow, whereas it did not negatively affect the outcomes of the Twist Comprehensive Viral Research Panel workflow.

Correlation of VL and sequence read counts

VLs, as determined by ddPCR on synthetic viral sequences in human cfDNA background, and by qPCR on clinical plasma samples, were compared with sequence read counts normalized by total library size and genome size (Fig. 3). Read counts correlated moderately with VLs. Outliers were detected for samples with low VLs, likely attributable to the stochastic effect around the limits of detection of PCR and the sequencing protocols.

Limits of detection

The limits of detection of both probe capture methods were analyzed for several ssRNA, dsDNA, and ssDNA viruses, and are shown in Fig. 4 ; Table S1. The limits of detection for the RNA viruses tested were approximately 50 and 500 copies/mL for the Twist Comprehensive Viral Research Panel workflow and the SeqCap EZ HyperCap workflow,

TABLE 2 Sensitivity and specificity resulting from varying thresholds based on genome coverage percentage using a random selection of 1 million sequence reads per data set, for the capture probe-based metagenomic workflows SeqCap EZ HyperCap (Roche) and Twist Comprehensive Viral Research Panel workflow[a]

Thresholds based on genome coverage				
	5% coverage	**10% coverage**	**20% coverage**	**90% coverage**
Twist Comprehensive Viral Research				
Sensitivity	1.000	1.000	1.000	0.957
Specificity	0.950	0.968	0.987	1.000
Corresponding LOD (c/mL)[b]	RNA viruses: 10^{1}	RNA: $10^{1\text{-}2}$	RNA: 10^{2}	RNA: 10^{3}
	DNA viruses: 10^{2}	DNA: 10^{2}	DNA: 10^{2}	DNA: 10^{3}–10^{4}
SeqCap EZ HyperCap workflow (Roche)				
Sensitivity	1.000	0.957	0.870	0.696
Specificity	0.973	0.981	0.995	0.997
Corresponding LOD (c/mL)[b]	RNA: 10^{2}	RNA: 10^{3}–10^{4}	RNA: 10^{3}–10^{4}	RNA: 10^{4}
	DNA: $10^{2\text{-}3}$	DNA: $10^{2\text{-}3}$	DNA: $10^{2\text{-}6}$	DNA: $10^{3\text{->}4}$

[a]The corresponding ROC is shown in Fig. 2. A minimum of three distributed regions of the genome covered was set as primary parameter for defining detection.
[b]Based on LOD of Inf A, SARS-CoV-2, EBV, HBV, and bocavirus.

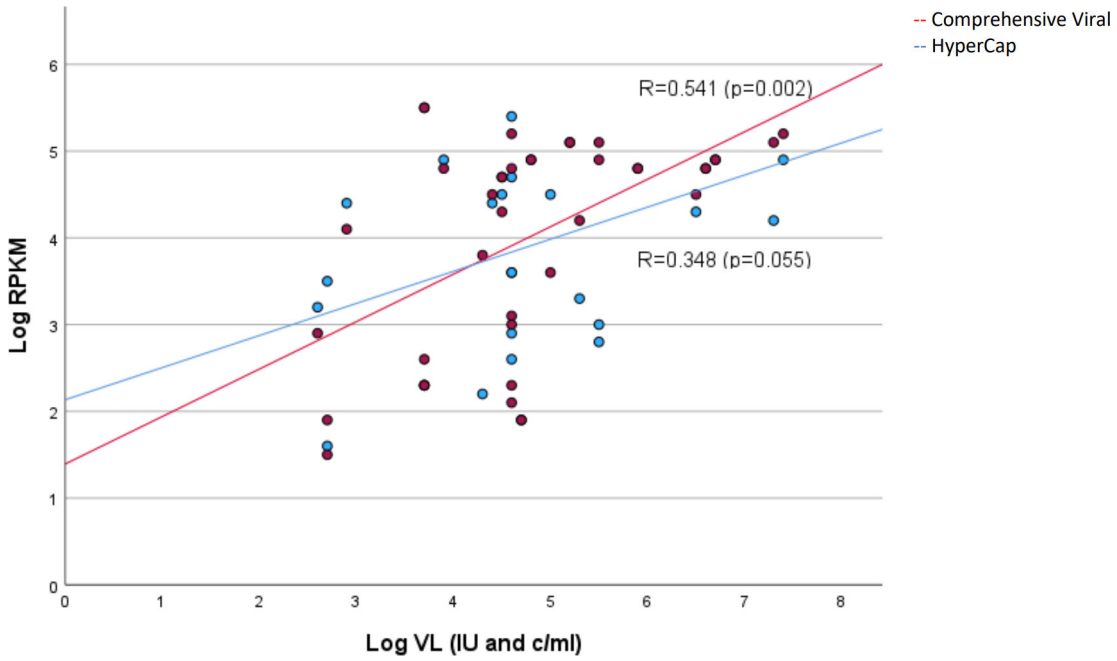

FIG 3 Correlation graph depicting linearity between the VL (log_{10} IU and C/mL, horizontally) and the log_{10} RKPM genome as generated using the virus capture probe-based metagenomic workflows Twist Comprehensive Viral Research and SeqCap EZ HyperCap (Roche). Included are detections by both methods from synthetic viral sequences spiked in a background of human cell-free DNA (90–99.991%), dilution series (see Fig. 4), and clinical samples.

respectively. For the dsDNA viruses tested, the LOD was 50–500 IU/mL for both methods. The LOD for the ssDNA virus human bocavirus (HBoV) was approximately 500 and 5,000 copies/mL for the Twist Comprehensive Viral Research Panel workflow and the

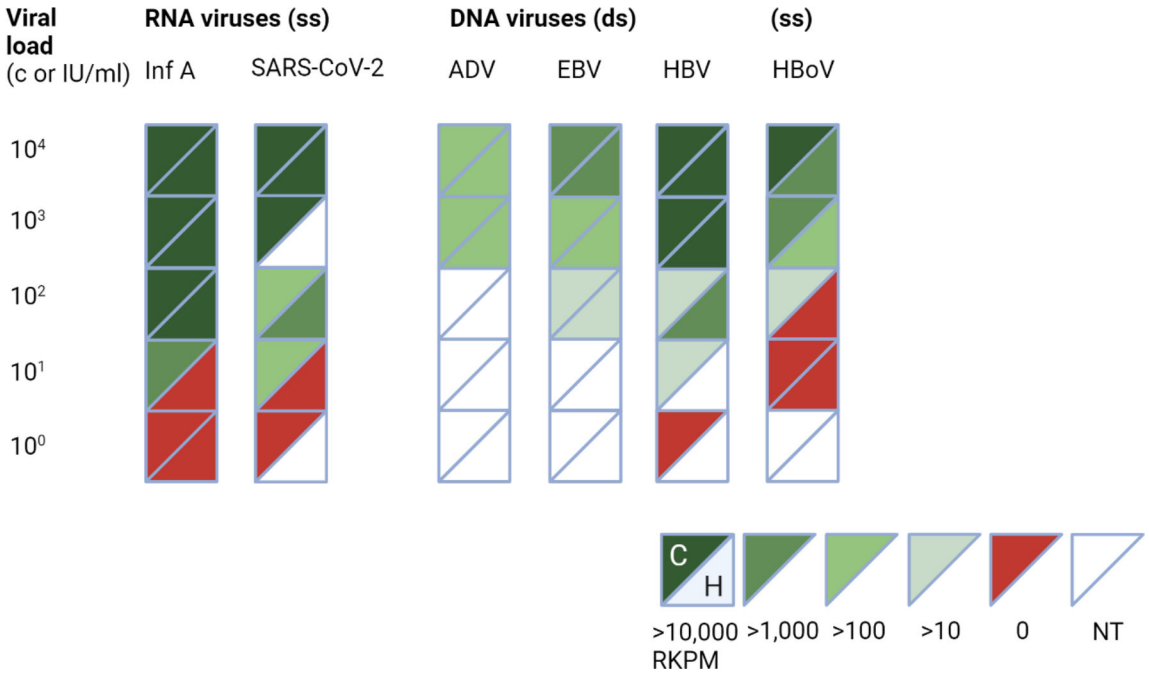

FIG 4 LOD of viral sequences using the virus capture probe-based metagenomic workflows Twist Comprehensive Viral Research (depicted in the left upper corner, "C"), and SeqCap EZ HyperCap (Roche, depicted in the right lower corner, "H"). RPKM genome are shown for different VLs (c or IU/mL, see Table S1). The samples consisted of synthetic viral sequences spiked in a background of human cell-free DNA (90–99.99990%) (Inf A, SARS-CoV-2, HBoV), and clinical EDTA plasma samples (ADV, EBV, HBV). NT, not tested. Created using BioRender.

SeqCap EZ HyperCap workflow, respectively. For the lowest LOD measured, this would correspond to the detection of <1 genome copy (part of a genome copy, or genome fragments) as input for the library preparation, without taking into account a threshold (RPM or coverage) for defining a positive result.

Reproducibility

Between-run variability as generated by both probe hybridization metagenomic workflows was studied by duplicate testing of three materials: two clinical samples and one synthetic viral sequence material in the presence of human cfDNA background (Fig. 5 ; Table S1). Normalized sequence read counts and genome coverage percentage were analyzed. Differences in target virus RPKM between runs were relatively low, ranging up to 4.7% (SE 2.35) coefficients of variance.

Effect of human background sequences

The qualitative and quantitative effects of increased proportion of human background sequences on the detection of viral target sequences were studied using synthetic viral sequences spiked in a varying amount of human cfDNA background sequences (90% vs. 99.999%, Fig. S1). No qualitative negative effect was found when the human cfDNA background proportion was increased up to 5 ng/100 uL, suggesting effective capturing of viral sequences. Quantitative target virus read counts were reduced in a single sample, in which non-human reads were accounted for the largest proportion of the read count. Overall, these data indicated effective capture of target sequences.

Application of determined thresholds to clinical samples

Optimal thresholds for defining a positive result, based on ROC as described above (using synthetic viral sequences and the Virome Virus mixture), were applied to the eight clinical plasma samples with known VLs (Table S1). All qPCR positive findings were positive by mNGS, for both methods. Additional findings when applying a threshold of minimal 500 RPM in combination with 10% coverage of at least three regions of the genome were: torque teno viruses (TTVs), adeno-associated dependoparvovirus A (AAV), and polyomaviruses. The TTV and AAV detections were consistent for both methods and undetected in the synthetic controls (Table S1). In contrast, Merkel cell polyomavirus was detected using the SeqCap EZ HyperCap workflow in two samples and synthetic controls, indicating environmental contamination. The spiked internal control viruses used in our laboratory, EAV and PhHV (both with C_T-values of approximately 33), were

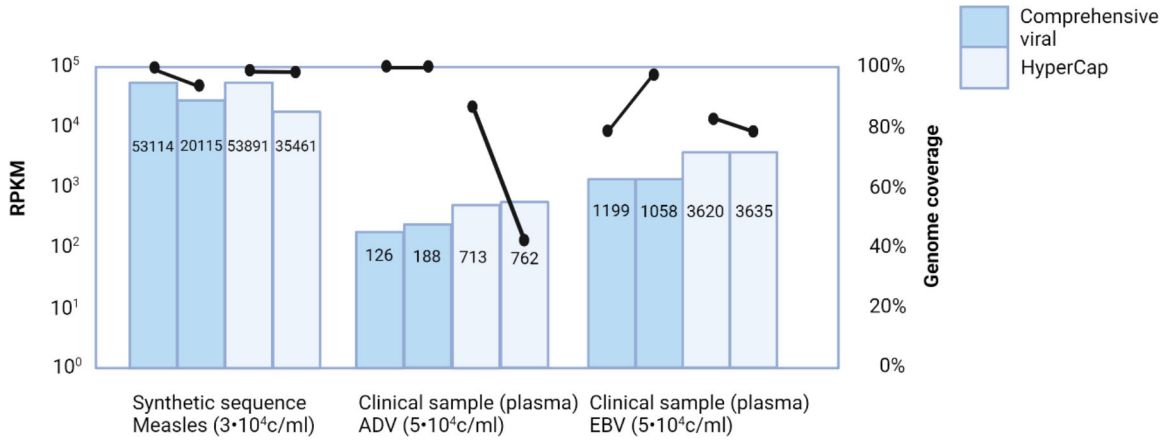

FIG 5 Reproducibility of read counts and genome coverage percentages. Between-run variability in three samples using the parameters RPKM (bars, left axis) and genome coverage percentage (black dots, right axis) as generated using the virus capture probe-based metagenomic workflows Twist Comprehensive Viral Research and SeqCap EZ HyperCap (Roche). The percentage of genome coverage was based on a random selection of 1 million sequence reads per data set. Coefficients of variance in RPKM ranged from 0.0% to 4.7% (see Table S1). Created using BioRender.

not detected and were not part of the design of the Twist method, in contrast to the Roche method.

DISCUSSION

This study provides the first one-to-one comparison of two pan-viral metagenomic probe capture workflows. The current data show analytical performances in ranges acceptable for clinical samples, for both probe hybridization-targeted metagenomic approaches. A combination of RPM and percentage of genome coverage was optimal for defining a positive result, accompanied by sensitivity and specificity well over 95% for both methods. Limits of detection were within ranges applicable to clinical settings: 50–500 c/mL for the Twist protocol when thresholds of 500 RPM and 10% were considered. While untargeted methods are intrinsically affected by the amount of background human DNA present in (tissue) samples (41), the results of this study show effective capturing in increasing proportions of human cell-free DNA without significantly affecting the read counts and the coverage of the virus genome.

A recent report has focused on a smaller probe panel targeting 29 human respiratory pathogenic viruses (Twist Respiratory Virus Panel) in comparison to the VirCapSeq (Roche) (42). The authors concluded that the Twist Respiratory Virus Panel workflow was suitable for the detection of both respiratory co-infections and SARS-CoV-2 variants with >90% 10-fold genome coverage. The latter is in line with our current data: genome coverage was generally 90%–100% for samples ≥1,000 c/mL, for a range of RNA and DNA viruses. It must be noted that the required pooling of samples before hybridization leads to lower amounts of total reads generated for lower biomass samples. Though this may potentially result in underestimation of the performance, in practice, the sensitivity was 100% despite lower total counts in some cases using the Twist workflow. Another report was recently published on the use of the Twist Comprehensive Viral Research Panel aiming at the detection of viruses involved in pediatric hepatitis cases of unknown origin, with an association with AAV2 was hypothesized (43). In 17 cases, AAV2 was detected using targeted sequencing, while in 7 of these pediatric cases, AAV2 was missed by untargeted metagenomic sequencing, illustrating the significance of the use of enrichment by hybridization. With regard to cost-efficiency, a recent study compared PCR, sequence-independent single primer amplification, and the Twist Comprehensive Viral Research Panel for the detection of Japanese encephalitis (44). The authors concluded that the PCR panels were not able to detect all genotypes, whereas broader surveillance of vector-borne pathogens would be more effective though costly (44). Hybridization capture has been approved by the Food and Drug Administration (FDA) for SARS-CoV-2 variant monitoring, illustrating the acknowledged significance of this type of enrichment. The LOD of the SARS-CoV-2 specific hybridization method in their study was 800 copies/mL (45), in line with our current and previous (10, 46) findings when using the broader panel. Even using the panel designed in 2015, (14) resulted in excellent genome coverage of SARS-CoV-2 due to sequence homology with animal coronaviruses and the variability in the probe design allowing for sequence mismatches (10).

In the current study, target read counts were normalized for 1 million total reads (RPM) and by subsampling, while 10 million reads were aimed at when using the SeqCap EZ HyperCap workflow based on the original publication (14). Importantly, subsampling did not result in reduced detection, which may offer an opportunity to reduce sequence costs.

Of note, uneven distribution of total viral read counts per sample resulted from the pre-hybridization pooling step present in both protocols, leading to proportionally higher counts for samples with higher VLs, up to several $\log_{10}$-folds. Despite these differences in total counts, all viral sequences in the panel were detected. Per protocol, equimolar pooling of sequences present before probe hybridization did not prevent uneven distribution of viral sequence reads due to the varying amount of non-viral sequences. Alternatively, probe hybridization per single sample may eliminate uneven

distribution of reads, though against higher material costs due to up to eight-fold increased probe consumption. Importantly, since all viral sequences in the current panel were detected, this would not have resulted in a different outcome.

Contaminating sequences are a challenge when implementing metagenomics in diagnostic settings. In the current study, both environmental sequences (for instance EBV, ADV C) and within-batch cross-contamination of target virus sequences (measles, influenza A, SARS-CoV-2) were detected in negative controls (cfDNA) and positive reference materials (Table S1). In almost all cases of additional findings, the RPM count and the percentage of genome coverage were low, enabling a relatively clear distinction between true and false positive findings when the determined optimal thresholds for defining a positive result were considered.

This study has several limitations. The synthetic sequences spiked in cell-free human DNA did not contain other background NAs such as bacterial and human RNA, though the latter proportion is generally low [<5% (47) dependent on the sample type]. The amount of cfDNA background in the synthetic materials may not perfectly represent the amount in all types of clinical samples, though it was a close representation for certain types (12, 31–33). Furthermore, though ssRNA, dsDNA, and ssDNA viruses were analyzed, detection and LOD results cannot be directly extrapolated to every single virus. These parameters may vary to some extent for different viruses, particularly those not included in the synthetic controls (manufactured by Twist). This was also exemplified by the lack of detection of EAV and PhHV using the Twist Comprehensive Viral Research panel. Though these viruses are not considered human pathogens, this illustrates the presence of certain restrictions with regard to the animal viruses included in the panel. Further analyses of the lists of viruses delivered by the probe designers showed that all pathogens on the WHO list of diseases with pandemic potential (https://www.who.int/news/item/21-11-2022-who-to-identify-pathogens-that-could-cause-future-outbreaks-and-pandemics) are present, in both probe panels.

To summarize, this study provides data supporting further steps toward widespread introduction of viral metagenomics for pathogen detection in clinical settings. In addition, it provides guidance for the integration of probe hybridization methods in surveillance to track pathogens of pandemic potential in low biomass samples such as wastewater (48) and wild-life swabs (49).

ACKNOWLEDGMENTS

This study was partially funded by Corona accelerated R&D in Europe (CARE Innovative Medicines Initiative, IMI).

Conceptualization: J.J.C.D.V. Software: I.S. Investigation; K.M., A.B. Methodology; K.M., E.C.J.C., J.J.C.D.V. Software; I.S. Formal analysis: K.M., I.S. Supervision; E.C.C., A.B., S.A.B., J.J.C.D.V. Visualization: K.M., J.J.C.D.V. Roles/Writing-original draft; K.M., J.J.C.D.V. Writing-review and editing: all authors.

AUTHOR AFFILIATIONS

[1]Department of Medical Microbiology, Leiden University Center for Infectious Diseases, Leiden University Medical Center, Leiden, the Netherlands
[2]GenomeScan B.V., Leiden, the Netherlands

AUTHOR ORCIDs

Kees Mourik http://orcid.org/0000-0001-7081-1237
Stefan A. Boers http://orcid.org/0000-0002-1560-6799
Jutte J. C. de Vries http://orcid.org/0000-0003-2530-6260

FUNDING

Funder	Grant(s)	Author(s)
Corona accelerated R&D in Europe		Kees Mourik

AUTHOR CONTRIBUTIONS

Kees Mourik, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Igor Sidorov, Formal analysis, Software, Writing – review and editing | Ellen C. Carbo, Writing – review and editing | David van der Meer, Writing – review and editing | Arnoud Boot, Investigation, Supervision, Writing – review and editing | Aloysius C. M. Kroes, Writing – review and editing | Eric C. J. Claas, Methodology, Writing – review and editing | Stefan A. Boers, Supervision, Writing – review and editing | Jutte J. C. de Vries, Conceptualization, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing

ETHICS APPROVAL

This study was approved by the medical ethics review committee Leiden/The Hague/Delft (CME number B20.002, 2020/2022).

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Figure S1 (JCM00345-24-S0001.pdf). Limited effect of spiked-in human cell-free DNA background sequences (90 to 99.92%) on target virus read counts.
Table S1 (JCM00345-24-S0002.xlsx). Sample characteristics, sequence read counts, and percentages of coverage of target virus genomes as obtained using the Twist Comprehensive Viral Research Panel and SeqCap EZ HyperCap (Roche).

REFERENCES

1. Gauthier NPG, Chorlton SD, Krajden M, Manges AR. 2023. Agnostic sequencing for detection of viral pathogens. Clin Microbiol Rev 36:e0011922. https://doi.org/10.1128/cmr.00119-22
2. Chiu CY, Miller SA. 2019. Clinical metagenomics. Nat Rev Genet 20:341–355. https://doi.org/10.1038/s41576-019-0113-7
3. Deng X, Achari A, Federman S, Yu G, Somasekar S, Bártolo I, Yagi S, Mbala-Kingebeni P, Kapetshi J, Ahuka-Mundeke S, et al. 2020. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. Nat Microbiol 5:443–454. https://doi.org/10.1038/s41564-019-0637-9
4. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, Federman S, Stryke D, Briggs B, Langelier C, et al. 2019. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med 380:2327–2340. https://doi.org/10.1056/NEJMoa1803396
5. Carbo EC, Blankenspoor I, Goeman JJ, Kroes ACM, Claas ECJ, De Vries JJC. 2021. Viral metagenomic sequencing in the diagnosis of meningoencephalitis: a review of technical advances and diagnostic yield. Expert Rev Mol Diagn 21:1139–1146. https://doi.org/10.1080/14737159.2021.1985467
6. Carr VR, Chaguza C. 2021. Metagenomics for surveillance of respiratory pathogens. Nat Rev Microbiol 19:285. https://doi.org/10.1038/s41579-021-00541-8
7. Kafetzopoulou LE, Pullan ST, Lemey P, Suchard MA, Ehichioya DU, Pahlmann M, Thielebein A, Hinzmann J, Oestereich L, Wozniak DM, et al. 2019. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science 363:74–77. https://doi.org/10.1126/science.aau9343
8. Holmes EC. 2022. COVID-19-lessons for zoonotic disease. Science 375:1114–1115. https://doi.org/10.1126/science.abn2222
9. Morfopoulou S, Buddle S, Torres Montaguth OE, Atkinson L, Guerra-Assunção JA, Moradi Marjaneh M, Zennezini Chiozzi R, Storey N, Campos L, Hutchinson JC, et al. 2023. Genomic investigations of unexplained acute hepatitis in children. Nature 617:564–573. https://doi.org/10.1038/s41586-023-06003-w
10. Carbo EC, Sidorov IA, Zevenhoven-Dobbe JC, Snijder EJ, Claas EC, Laros JFJ, Kroes ACM, de Vries JJC. 2020. Coronavirus discovery by metagenomic sequencing: a tool for pandemic preparedness. J Clin Virol 131:104594. https://doi.org/10.1016/j.jcv.2020.104594
11. López-Labrador FX, Brown JR, Fischer N, Harvala H, Van Boheemen S, Cinek O, Sayiner A, Madsen TV, Auvinen E, Kufner V, et al. 2021. Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: wet lab procedure. J Clin Virol 134:104691. https://doi.org/10.1016/j.jcv.2020.104691
12. van Rijn AL, van Boheemen S, Sidorov I, Carbo EC, Pappas N, Mei H, Feltkamp M, Aanerud M, Bakke P, Claas ECJ, Eagan TM, Hiemstra PS, Kroes ACM, de Vries JJC. 2019. The respiratory virome and exacerbations in patients with chronic obstructive pulmonary disease. PLoS One 14:e0223952. https://doi.org/10.1371/journal.pone.0223952
13. Metsky HC, Siddle KJ, Gladden-Young A, Qu J, Yang DK, Brehio P, Goldfarb A, Piantadosi A, Wohl S, Carter A, et al. 2019. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. Nat Biotechnol 37:160–168. https://doi.org/10.1038/s41587-018-0006-x
14. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI. 2015. Virome capture sequencing enables sensitive viral diagnosis and

comprehensive virome analysis. mBio 6:e01491-15. https://doi.org/10.1128/mBio.01491-15
15. Wylie TN, Wylie KM, Herter BN, Storch GA. 2015. Enhanced virome sequencing using targeted sequence capture. Genome Res 25:1910–1920. https://doi.org/10.1101/gr.191049.115
16. Kuchinski KS, Loos KD, Suchan DM, Russell JN, Sies AN, Kumakamba C, Muyembe F, Mbala Kingebeni P, Ngay Lukusa I, N'Kawa F, et al. 2022. Targeted genomic sequencing with probe capture for discovery and surveillance of coronaviruses in bats. Elife 11:e79777. https://doi.org/10.7554/eLife.79777
17. Yamaguchi J, Olivo A, Laeyendecker O, Forberg K, Ndembi N, Mbanya D, Kaptue L, Quinn TC, Cloherty GA, Rodgers MA, Berg MG. 2018. Universal target capture of HIV sequences from NGS libraries. Front Microbiol 9:2150. https://doi.org/10.3389/fmicb.2018.02150
18. Wang H, Pian H, Fan L, Li J, Yang J, Zheng Z. 2023. Multiple-probe-assisted DNA capture and amplification for high-throughput African swine fever virus detection. Appl Microbiol Biotechnol 107:797–805. https://doi.org/10.1007/s00253-022-12334-x
19. Doddapaneni H, Cregeen SJ, Sucgang R, Meng Q, Qin X, Avadhanula V, Chao H, Menon V, Nicholson E, Henke D, Piedra F-A, Rajan A, Momin Z, Kottapalli K, Hoffman KL, Sedlazeck FJ, Metcalf G, Piedra PA, Muzny DM, Petrosino JF, Gibbs RA. 2021. Oligonucleotide capture sequencing of the SARS-Cov-2 genome and subgenomic fragments from COVID-19 individuals. PLoS One 16:e0244468. https://doi.org/10.1371/journal.pone.0244468
20. Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, Andersson M, Otecko N, Wise EL, Moore N, et al. 2021. SARS-Cov-2 within-host diversity and transmission. Science 372:eabg0821. https://doi.org/10.1126/science.abg0821
21. Carbo EC, Buddingh EP, Karelioti E, Sidorov IA, Feltkamp MCW, Borne P von dem, Verschuuren J, Kroes ACM, Claas ECJ, de Vries JJC. 2020. Improved diagnosis of viral encephalitis in adult and pediatric hematological patients using viral metagenomics. J Clin Virol 130:104566. https://doi.org/10.1016/j.jcv.2020.104566
22. Schuele L, Cassidy H, Lizarazo E, Strutzberg-Minder K, Schuetze S, Loebert S, Lambrecht C, Harlizius J, Friedrich AW, Peter S, Niesters HGM, Rossen JWA, Couto N. 2020. Assessment of viral targeted sequence capture using Nanopore sequencing directly from clinical samples. Viruses 12:1358. https://doi.org/10.3390/v12121358
23. Wylie KM, Wylie TN, Buller R, Herter B, Cannella MT, Storch GA. 2018. Detection of viruses in clinical samples by use of metagenomic sequencing and targeted sequence capture. J Clin Microbiol 56:e01123-18. https://doi.org/10.1128/JCM.01123-18
24. Jansen SA, Nijhuis W, Leavis HL, Riezebos-Brilman A, Lindemans CA, Schuurman R. 2020. Broad virus detection and variant discovery in fecal samples of hematopoietic transplant recipients using targeted sequence capture metagenomics. Front Microbiol 11:560179. https://doi.org/10.3389/fmicb.2020.560179
25. Stout MJ, Brar AK, Herter BN, Rankin A, Wylie KM. 2023. The plasma virome in longitudinal samples from pregnant patients. Front Cell Infect Microbiol 13:1061230. https://doi.org/10.3389/fcimb.2023.1061230
26. Garand M, Huang SSY, Goessling LS, Wan F, Santillan DA, Santillan MK, Brar A, Wylie TN, Wylie KM, Eghtesady P. 2023. Virome analysis and association of positive coxsackievirus B serology during pregnancy with congenital heart disease. Microorganisms 11:262. https://doi.org/10.3390/microorganisms11020262
27. Flerlage T, Crawford JC, Allen EK, Severns D, Tan S, Surman S, Ridout G, Novak T, Randolph A, West AN, Thomas PG. 2023. Single cell transcriptomics identifies distinct profiles in pediatric acute respiratory distress syndrome. Nat Commun 14:3870. https://doi.org/10.1038/s41467-023-39593-0
28. Cassidy H, Schuele L, Lizarazo-Forero E, Couto N, Rossen JWA, Friedrich AW, van Leer-Buter C, Niesters HGM. 2022. Exploring a prolonged enterovirus C104 infection in a severely ill patient using nanopore sequencing. Virus Evol 8:veab109. https://doi.org/10.1093/ve/veab109
29. Esnault G, Earley B, Cormican P, Waters SM, Lemon K, Cosby SL, Lagan P, Barry T, Reddington K, McCabe MS. 2022. Assessment of rapid MinION Nanopore DNA virus meta-genomics using calves experimentally infected with bovine herpes virus-1. Viruses 14:1859. https://doi.org/10.3390/v14091859
30. Zhang C, Burch M, Wylie K, Herter B, Franklin CL, Ericsson AC. 2021. Characterization of the eukaryotic virome of mice from different sources. Microorganisms 9:2064. https://doi.org/10.3390/microorganisms9102064
31. Junier T, Huber M, Schmutz S, Kufner V, Zagordi O, Neuenschwander S, Ramette A, Kubacki J, Bachofen C, Qi W, Laubscher F, Cordey S, Kaiser L, Beuret C, Barbié V, Fellay J, Lebrand A. 2019. Viral metagenomics in the clinical realm: lessons learned from a swiss-wide ring trial. Genes (Basel) 10:655. https://doi.org/10.3390/genes10090655
32. van Boheemen S, van Rijn AL, Pappas N, Carbo EC, Vorderman RHP, Sidorov I, van T Hof PJ, Mei H, Claas ECJ, Kroes ACM, de Vries JJC. 2020. Retrospective validation of a metagenomic sequencing protocol for combined detection of RNA and DNA viruses using respiratory samples from pediatric patients. J Mol Diagn 22:196–207. https://doi.org/10.1016/j.jmoldx.2019.10.007
33. Morfopoulou S, Mee ET, Connaughton SM, Brown JR, Gilmour KChongWK "Kling"Duprex WP, Ferguson D, Hubank M, Hutchinson C, Kaliakatsos M, McQuaid S, Paine S, Plagnol V, Ruis C, Virasami A, Zhan H, Jacques TS, Schepelmann S, Qasim W, Breuer J. 2017. Deep sequencing reveals persistence of cell-associated mumps vaccine virus in chronic encephalitis. Acta Neuropathol 133:139–147. https://doi.org/10.1007/s00401-016-1629-y
34. Carbo EC, Russcher A, Kraakman MEM, de Brouwer CS, Sidorov IA, Feltkamp MCW, Kroes ACM, Claas ECJ, de Vries JJC. 2022. Longitudinal monitoring of DNA viral loads in transplant patients using quantitative metagenomic next-generation sequencing. Pathogens 11:236. https://doi.org/10.3390/pathogens11020236
35. Reyes A, Carbo EC, Harinxma Thoe Slooten JS van, Kraakman MEM, Sidorov IA, Claas ECJ, Kroes ACM, Visser LG, de JJCV. 2021. Viral metagenomic sequencing in a cohort of international travellers returning with febrile illness. J Clin Virol 143:104940. https://doi.org/10.1016/j.jcv.2021.104940
36. Carbo EC, Sidorov IA, van Rijn-Klink AL, Pappas N, van Boheemen S, Mei H, Hiemstra PS, Eagan TM, Claas ECJ, Kroes ACM, de Vries JJC. 2022. Performance of five metagenomic classifiers for virus pathogen detection using respiratory samples from a clinical cohort. Pathogens 11:340. https://doi.org/10.3390/pathogens11030340
37. de Vries JJC, Brown JR, Fischer N, Sidorov IA, Morfopoulou S, Huang J, Munnink BBO, Sayiner A, Bulgurcu A, Rodriguez C, et al. 2021. Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples. J Clin Virol 141:104908. https://doi.org/10.1016/j.jcv.2021.104908
38. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
39. Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, Pauwels R, Alcantara LC, Vanden Eynden E, Vandamme A-M, Deforche K, de Oliveira T. 2019. Genome detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics 35:871–873. https://doi.org/10.1093/bioinformatics/bty695
40. de Vries JJC, Brown JR, Couto N, Beer M, Le Mercier P, Sidorov I, Papa A, Fischer N, Oude Munnink BB, Rodriquez C, et al. 2021. Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting. J Clin Virol 138:104812. https://doi.org/10.1016/j.jcv.2021.104812
41. Lewandowska DW, Zagordi O, Geissberger F-D, Kufner V, Schmutz S, Böni J, Metzner KJ, Trkola A, Huber M. 2017. Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. Microbiome 5:94. https://doi.org/10.1186/s40168-017-0317-z
42. Kim KW, Deveson IW, Pang CNI, Yeang M, Naing Z, Adikari T, Hammond JM, Stevanovski I, Beukers AG, Verich A, Yin S, McFarlane D, Wilkins MR, Stelzer-Braid S, Bull RA, Craig ME, van Hal SJ, Rawlinson WD. 2021. Respiratory viral co-infections among SARS-Cov-2 cases confirmed by virome capture sequencing. Sci Rep 11:3934. https://doi.org/10.1038/s41598-021-83642-x
43. Servellita V, Sotomayor Gonzalez A, Lamson DM, Foresythe A, Huh HJ, Bazinet AL, Bergman NH, Bull RL, Garcia KY, Goodrich JS, et al. 2023.

Adeno-associated virus type 2 in US children with acute severe hepatitis. Nature 617:574–580. https://doi.org/10.1038/s41586-023-05949-1
44. Crispell G, Williams K, Zielinski E, Iwami A, Homas Z, Thomas K. 2022. Method comparison for Japanese encephalitis virus detection in samples collected from the Indo-Pacific region. Front Public Health 10:1051754. https://doi.org/10.3389/fpubh.2022.1051754
45. Nagy-Szakal D, Couto-Rodriguez M, Wells HL, Barrows JE, Debieu M, Butcher K, Chen S, Berki A, Hager C, Boorstein RJ, Taylor MK, Jonsson CB, Mason CE, O'Hara NB. 2021. Targeted hybridization capture of SARS-Cov-2 and metagenomics enables genetic variant discovery and nasal microbiome insights. Microbiol Spectr 9:e0019721. https://doi.org/10.1128/Spectrum.00197-21
46. Carbo EC, Mourik K, Boers SA, Munnink BO, Nieuwenhuijse D, Jonges M, Welkers MRA, Matamoros S, van Harinxma Thoe Slooten J, Kraakman MEM, Karelioti E, van der Meer D, Veldkamp KE, Kroes ACM, Sidorov I, de Vries JJC. 2023. A comparison of five Illumina, ion torrent, and Nanopore sequencing technology-based approaches for whole genome sequencing of SARS-Cov-2. Eur J Clin Microbiol Infect Dis 42:701–713. https://doi.org/10.1007/s10096-023-04590-0
47. Wu J, Xiao J, Zhang Z, Wang X, Hu S, Yu J. 2014. Ribogenomics: the science and knowledge of RNA. Genom Proteom Bioinform 12:57–63. https://doi.org/10.1016/j.gpb.2014.04.002
48. Clark JR, Terwilliger A, Avadhanula V, Tisza M, Cormier J, Javornik-Cregeen S, Ross MC, Hoffman KL, Troisi C, Hanson B, Petrosino J, Balliew J, Piedra PA, Rios J, Deegan J, Bauer C, Wu F, Mena KD, Boerwinkle E, Maresso AW. 2023. Wastewater pandemic preparedness: toward an end-to-end pathogen monitoring program. Front Public Health 11:1137881. https://doi.org/10.3389/fpubh.2023.1137881
49. Lwande OW, Thalin T, de Jong J, Sjödin A, Näslund J, Evander M, Ecke F. 2022. Alphacoronavirus in a Daubenton's myotis bat (*Myotis daubentonii*) in Sweden. Viruses 14:556. https://doi.org/10.3390/v14030556