

1 Comparison of the performance of two targeted metagenomic virus capture probe-based methods  
2 using synthetic viral sequences and clinical samples

3

4 Kees Mourik<sup>1\*</sup>, Igor Sidorov<sup>1</sup>, Ellen C. Carbo<sup>1</sup>, David van der Meer<sup>2</sup>, Arnoud Boot<sup>2</sup>, Aloysius C. M. Kroes<sup>1</sup>,  
5 Eric C.J. Claas<sup>1</sup>, Stefan A. Boers<sup>1</sup>, Jutte J.C. de Vries<sup>1\*</sup>

6 <sup>1</sup> Department of Medical Microbiology, Leiden University Center for Infectious Diseases, Leiden  
7 University Medical Center, Leiden, the Netherlands

8 <sup>2</sup> GenomeScan B.V., Leiden, the Netherlands

9 \*Corresponding authors

10

11 Abstract

12 Viral enrichment by probe hybridization has been reported to significantly increase the sensitivity of  
13 viral metagenomics.

14 This study compares the analytical performance of two targeted metagenomic virus capture probe-  
15 based methods: i) SeqCap EZ HyperCap by Roche (ViroCap) and ii) Twist Comprehensive Viral Research  
16 Panel workflow, for diagnostic use. Sensitivity, specificity, limit of detection, and effect of human  
17 background DNA were analysed, using synthetic viral sequences, clinical and reference samples with  
18 known viral loads.

19 Sensitivity and specificity were 95% and higher for both methods. Combining thresholds for viral  
20 sequence read counts and genome coverage (respectively 500 reads per million and 10% coverage)  
21 resulted in optimal prediction of true positive results. Limits of detection were approximately 50-500  
22 copies/ml for both methods. Increasing proportions of spike-in cell free human background sequences  
23 did not negatively affect viral detection.

24 These data show analytical performances in ranges applicable to clinical samples, for both probe  
25 hybridization metagenomic approaches. This study supports further steps towards more widespread  
26 use of viral metagenomics for pathogen detection, in clinical and surveillance settings using low  
27 biomass samples.

28

29 **Keywords:** viral metagenomics, capture probes, targeted metagenomics, viral diagnostics

**NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## 30 Introduction

31 Viral metagenomics has been gradually applied for broad-spectrum pathogen detection of infectious  
32 diseases<sup>1-5</sup>, surveillance of emerging diseases<sup>3,6-8</sup>, and pathogen discovery<sup>9,10</sup>. Though metagenomic  
33 approaches have been practiced for decades in the field of marine environments and the human  
34 microbiome, this approach is nowadays changing how physicians diagnose infectious diseases<sup>2</sup>.  
35 Whereas amplicon based metagenomics has been successfully adopted for bacterial diagnostics, no  
36 amplicon based pan-viral approach is available and therefore viral metagenomics has not yet been  
37 widely deployed as diagnostic tool in clinical laboratories<sup>1</sup>. One of the main challenges for application  
38 in clinical settings is the low level of viral genomes in the presence of high levels of host material in  
39 patient samples. Several methods for depletion of host sequences and enrichment of viral sequences  
40 have been studied with varying success rates<sup>11</sup>. For example, depletion of host cells prior to extraction  
41 of nucleic acids (NA) has been reported to be not advantageous in clinical samples as also intracellular  
42 viral particles or NA will be removed<sup>11,12</sup>. In contrast, viral enrichment by probe hybridization methods  
43 has been reported to significantly increase sensitivity in various sample types<sup>1,13-19</sup>, up to the level  
44 required for accurate detection of low frequency virus variants<sup>20</sup>.

45 Previously, the performance of a hybridization capture probe panel targeting vertebrate viruses in  
46 cerebrospinal fluids from patients with meningo-encephalitis has been analysed<sup>21</sup>. Viral target  
47 sequence read counts increased 100-10.000 fold compared to unenriched metagenomic sequencing,  
48 and sensitivity by enrichment was comparable with polymerase chain reaction (PCR)<sup>21</sup>. Moreover,  
49 these earlier data showed that this hybridisation panel of approximately two million capture probes  
50 designed in 2015 was suited for the detection of novel coronaviruses by reactivity with other  
51 vertebrate betacoronavirus probes<sup>10</sup>. During the past years, this specific hybridisation panel  
52 distributed by Roche has been adopted in a broad range of different clinical<sup>15,22-29</sup> and zoonotic  
53 settings<sup>22,29,30</sup>. Recently, Twist Bioscience has released a new hybridisation panel containing  
54 approximately one million size capture probes targeting human and animal viruses. Reports  
55 comparing the performance of viral metagenomic hybridisation panels are lacking.

56 Here, we compare the analytical performance of two targeted metagenomic virus capture probe-  
57 based methods: i) SeqCap EZ HyperCap by Roche (ViroCap) and ii) Twist Comprehensive Viral Research  
58 Panel workflow, for clinical diagnostic use. Sensitivity, specificity, and limit of detection were analysed  
59 using synthetic viral sequences, clinical and reference samples with known viral loads.

60

## 61 **Methods**

### 62 Synthetic sequences and clinical samples

63 An overview of the validation panels and study design is shown in **Figure 1**.

64 In order to mimic the complexity of clinical samples while reducing the number of additional viral  
65 sequences, enabling sensitivity and specificity analyses, a panel was prepared of synthetic viral  
66 sequences ( $10^0$  -  $10^7$  copies/ml) spiked in human cell free DNA (cf DNA) background sequences (Twist  
67 Bioscience, San Francisco, USA). Synthetic viral sequences covered >99.9% of the viral genomes of  
68 SARS-CoV-2, influenza A virus (Inf A), measles, enterovirus D68, and bocavirus and were mixed with  
69 several proportions of human cfDNA (90-99.999% of weight, corresponding with proportions of up to  
70 10-0.001%<sup>31</sup> of viral nucleotides in a clinical sample<sup>12,32,33</sup>). Concentrations of viral sequences were  
71 determined by digital droplet PCR in triplicate (BioRad QX200). A total of 25 of synthetic mixtures were  
72 included (see **Suppl. Table 1**).

73 Clinical EDTA plasma samples (n=8), previously submitted to the Clinical Microbiological Laboratory  
74 for routine diagnostic testing and tested positive by qPCR<sup>21,34</sup> for adenovirus (ADV), Epstein-Barr virus  
75 (EBV) and Hepatitis B virus (HBV), were included in the comparison. Viral loads ranged from 500 to  
76 50,000 International Units (IU)/ml.

77 In addition, a dilution of ATCC Virome Virus Mix (MSA-2008™, ATCC, Manassas, USA) of cultivated  
78 adenovirus type F (ADV), cytomegalovirus (CMV), respiratory syncytial virus (RSV), influenza B virus,  
79 reovirus 3, and zika virus was included.

80

### 81 Ethical approval

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

84

### 85 Nucleic acids extraction

86 Clinical samples and the ATCC whole virus mixture were subjected to extraction of total nucleic acids  
87 (NA) using the MagNAPure 96 DNA and Viral NA Small volume extraction kit (Roche, Basel,  
88 Switzerland) as described previously<sup>21</sup>.

89

90 Viral metagenomic next-generation sequencing (mNGS)

91 Twist Comprehensive Viral Research Panel workflow

92 Sample preparation was performed using the Comprehensive Viral Research Panel workflow (Twist  
93 Bioscience Corp.) according to the manufacturer's instructions. In short, 5 µl of NA was used as input  
94 for cDNA synthesis (Protoscript, New England Biolabs, Inc) followed by purification using magnetic  
95 beads, enzymatic fragmentation for 15 minutes at 30 °C, end repair and dA-tailing (Twist EF Library  
96 Prep 2.0, Twist Bioscience Corp.). Next, unique molecular identifier (UMI) adapters with unique dual  
97 barcodes (Twist UMI Adapter System, Twist Bioscience Corp.) were ligated to the fragments and  
98 amplified using PCR (12 cycles). Amplified libraries were pooled per 8 samples and library pools were  
99 used for hybridization with the Twist Comprehensive Virus probe panel, consisting of ~1 million 120  
100 bp probes targeting 15,488 different viral strains infecting human and animals. Hybridization was  
101 performed for 16 hours of incubation followed by several wash steps. Captured fragments were  
102 further amplified by a post-hybridization PCR (15 cycles). Finally, captured libraries were purified by a  
103 bead clean up using AmpureXP, and quantity and fragment size were determined using Qubit (Thermo  
104 Fisher, Waltham, MA, USA) and Fragment Analyser (Agilent, Santa Clara, CA, USA) respectively.  
105 Libraries were clustered and approximately 1 million 150bp paired-end reads were generated per  
106 sample, according to manufacturer's protocols (Illumina Inc.) at GenomeScan B.V. using the  
107 NovaSeq6000.

108

109 SeqCap EZ HyperCap (ViroCap design, Roche)

110 The SeqCap EZ HyperCap workflow (Roche, Madison, USA) was performed as validated and described  
111 previously<sup>10,21,35</sup>. Briefly, 5 ul of NA was used as direct input (without concentration step) for enzymatic  
112 fragmentation and cDNA synthesis using the NEBNext Ultra II Directional RNA Library preparation kit  
113 V3.0 (New England Biolabs, Ipswich, MA, USA) for Illumina with several in-house adaptations to enable  
114 simultaneous detection of both DNA and RNA in a single tube per sample<sup>12,32</sup>. After purification, dual  
115 barcodes (NEBNext Multiplex oligos for illumina 96 unique dual index primer pairs) were attached to  
116 the fragments and amplified using PCR (21 cycles). Four barcoded samples including controls were  
117 pooled, COT (enriched for repetitive sequences) human DNA and HyperCap Universal Blocking Oligos  
118 were added before purification, following incubation for >40 hours with the SeqCap EZ HyperCap v1  
119 (ViroCap design, 2015<sup>14</sup>), a collection of approximately two million oligonucleotide probes (70–120  
120 mers) targeting all known vertebrate viruses. A complete list of the viral taxa included can be found in  
121 the supplementary tables of the manuscript by Briese et al<sup>14</sup>. Captured fragments were further

122 amplified by a post-hybridization PCR (14 cycles). Finally, captured libraries were purified by bead  
123 clean up using AmpureXP, and quantity and fragment sizes were determined using Qubit (Thermo  
124 Fisher, Waltham, MA, USA) and Fragment Analyser (Agilent, Santa Clara, CA, USA), respectively.  
125 Approximately 10 million 150 bp paired-end reads were sequenced per sample according to  
126 manufacturer's protocols (Illumina Inc.) at GenomeScan B.V. using the NovaSeq6000.

127

128 Data analysis

129 Bioinformatic analysis

130 Image analysis, base calling, and quality check of sequence data were performed with the Illumina  
131 data analysis pipelines RTA3.4.4 and bcl2fastq v2.20 (Illumina). Sequence data obtained using both  
132 probe capture metagenomics methods were analyzed using a previously validated<sup>10,21,36,37</sup>  
133 bioinformatics pipeline. After quality pre-processing and removal of human reads (by mapping them  
134 to the human reference genome GRCh38  
135 ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.26/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/) with Bowtie2<sup>38</sup> version 2.3.4),  
136 datasets were analyzed using Genome Detective<sup>39</sup> version 2.48 (accessed April – May 2023) as  
137 described previously<sup>36</sup>. Genome Detective includes *de novo* assembly and both nucleotide and amino  
138 acid based classification in combination with a RefSeq / Swiss-Prot Uniref database by Genome  
139 Detective<sup>39</sup>

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

146

147 Performance metrics and statistical analyses

148 Sensitivity and specificity were calculated using the results from the synthetic viral sequences and the  
149 ATCC Virome Virus mix. Additional findings were considered false positives, and non-vertebrate  
150 viruses were excluded from analyses. Receiver Operating Characteristic (ROC) curves were generated  
151 by varying the number of sequence-read counts used as cut-off for defining a positive result, given a  
152 prerequisite of  $\geq 3$  genome regions covered<sup>40</sup>, and area under the curves (AUC) were calculated.

153 Spearman correlations of sequence read counts with viral load, as determined by qPCR and ddPCR,  
154 were analyzed.

155 Limits of detection for both methods were determined using 10-fold serial dilutions of synthetic viral  
156 NA in human cfDNA background, and undiluted clinical samples. Reproducibility was determined by  
157 analyzing the coefficient of variance between runs.

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

160

## 161 Results

### 162 Analytic sensitivity, specificity, and ROC

163 Detection of synthetic viral sequences in human cfDNA background, and the Virome Virus Mix, using  
164 the Twist Comprehensive Viral Research Panel and the SeqCap EZ HyperCap (ViroCap) workflow is  
165 depicted in **Suppl. Table 1**. For both methods, sensitivity was 100% (23/23, cycle threshold,  $C_t$ , values  
166 ranging from 20 to 32). Viral target read counts ranged from 334-872,042 reads per million (RPM) for  
167 the Twist Comprehensive Viral Research workflow, and 2,171- 971,610 RPM for the SeqCap EZ  
168 HyperCap workflow. Genome coverage ranged from 91.1-100% (median 99.8%), and 8.4-100%  
169 (median 97.5%), for these respective methods. Sensitivity and specificity were calculated for different  
170 thresholds for defining a positive result: i) sequence read counts and ii) genome coverage percentage,  
171 as depicted in **Figure 2** and **Table 1**. For calculation of the percentage of the viral genomes covered, a  
172 random selection of 1 million sequence reads per dataset were used. **Figure 2** shows that both RPM  
173 and genome coverage were distinctive parameters for defining a true positive result, with AUC of  
174 99.8% for both methods when considering RPM as parameter, and  $\geq 99.7\%$  when considering genome  
175 coverage as parameter. Sensitivity and specificity scores of  $\geq 95\%$  were accomplished for both methods  
176 when 500 RPM was set as threshold, on top of a prerequisite of minimum of three distributed regions  
177 of the genome being covered (**Table 1**). Similarly, when coverage was set at 10% of the genome, both  
178 methods reached sensitivity and specificity levels of 95% and higher. Increasing the threshold for  
179 genome coverage resulted in decreased sensitivity for the SeqCap EZ HyperCap workflow, whereas it  
180 did not negatively affect the outcomes of the Twist Comprehensive Viral Research Panel workflow.

181

### 182 Correlation of viral load and sequence read counts

183 Viral loads as determined by ddPCR on synthetic viral sequences in human cfDNA background, and by  
184 qPCR on clinical plasma samples, were compared with sequence read counts normalized by total  
185 library size and genome size (**Figure 3**). Read counts were significantly correlated with viral loads, for  
186 both methods. Outliers were detected for samples with low viral loads, likely attributable to the  
187 stochastic effect around the limits of detection of PCR and the sequencing protocols.

188

### 189 Limits of detection

190 The limits of detection of both probe capture methods were analysed for several ssRNA, dsDNA, and  
191 ssDNA viruses, and is shown in **Figure 4** and **Suppl. Table 1**. The limits of detection for the RNA viruses

192 tested were approximately 50 and 500 copies/mL for the Twist Comprehensive Viral Research Panel  
193 workflow and the SeqCap EZ HyperCap workflow, respectively. For the DNA viruses tested, the LOD  
194 was approximately 500 IU/ml for both methods, apart from the limit of detection of HBoV, which was  
195 approximately 5,000 c/ml using the SeqCap EZ HyperCap workflow.

196

#### 197 Reproducibility

198 Between-run variability as generated by both probe hybridization metagenomic workflows was  
199 studied by repeated testing of clinical samples and synthetic sequences in presence of human cfDNA  
200 background (**Figure 5** and **Suppl. Table 1**). Normalized sequence read counts and genome coverage  
201 percentage were analysed. Differences in target virus RPKM between runs were relatively low, ranging  
202 from 0.0 to 4.7% coefficients of variance.

203

#### 204 Effect of human background sequences

205 The qualitative and quantitative effects of increased proportion of human background sequences on  
206 the detection of viral target sequences was studied using synthetic viral sequences spiked in a varying  
207 amount of human cfDNA background sequences (90% versus 99.999%, **Suppl. Figure 1**). No qualitative  
208 negative effect was found when the human cfDNA background proportion was increased. Quantitative  
209 target virus read counts were reduced in a single sample in which non-human reads were accounted  
210 for the largest proportion of the read count. Overall, these data indicated effective capture of target  
211 sequences.

212

#### 213 Application of determined thresholds to clinical samples

214 Optimal thresholds for defining a positive result, determined as described above (using synthetic viral  
215 sequences and the Virome Virus mixture) were applied to the eight clinical plasma samples with  
216 known viral loads (**Suppl. Table 1**). All qPCR positive findings were positive by mNGS, for both  
217 methods. Additional findings when applying a threshold of minimal 500 RPM in combination with 10%  
218 coverage of at least three regions of the genome were: torque teno viruses, adeno-associated  
219 dependoparvovirus A, and polyomaviruses. The additional findings were consistent for both methods,  
220 except for Merkel cell polyomavirus which was detected using the SeqCap EZ HyperCap workflow in  
221 two samples, indicating environmental contamination. The internal control sequences used in our  
222 laboratory, equine arteritis virus (EAV) and phocid herpes virus (PhHV, both with  $C_T$ -values of



223 approximately 33) were not detected and not part of the design of the Twist method, in contrast to  
224 the Roche method.

225

## 226 Discussion

227 These data show analytical performances in ranges acceptable for clinical samples, for both probe  
228 hybridization targeted metagenomic approaches. A combination of RPM and percentage of genome  
229 coverage were optimal for defining a positive result, accompanied by sensitivity and specificity well  
230 over 95% for both methods. Limits of detection were within ranges applicable to clinical settings: 50-  
231 500 c/ml for the Twist protocol when thresholds of 500 RPM and 10% were considered. While  
232 untargeted methods are intrinsically affected by the amount of background human DNA present in  
233 (tissue) samples<sup>41</sup>, the results of this study show effective capturing in increasing proportions of  
234 human cell free DNA without significantly affecting the read counts and the coverage of the virus  
235 genome. This study provides the first one-to-one comparison of two pan-viral metagenomic probe  
236 capture workflows.

237 A recent report has studied a smaller probe panel targeting 29 human respiratory pathogenic viruses  
238 in comparison to the VirCapSeq (Roche)<sup>42</sup>. The authors conclude that the Twist Respiratory Virus Panel  
239 workflow was suited for detection of both respiratory co-infections and SARS-CoV-2 variants with  
240 >90% tenfold genome coverage. The latter is in line with our current data: genome coverage was  
241 generally 90-100% for samples  $\geq 1,000$  C/ml, for a range of RNA and DNA viruses. It must be noted that  
242 the required pooling of samples prior to hybridization lead to lower amounts of total reads generated  
243 for lower biomass samples. Though this potentially may result in underestimation of the performance,  
244 in practice, the sensitivity was 100% despite lower total counts in some cases using the Twist  
245 workflow. Another report was recently published on the use of the Twist Comprehensive Viral  
246 Research Panel aiming at detection of viruses involved in pediatric hepatitis cases of unknown origin,  
247 while an association with AAV2 was hypothesised<sup>43</sup>. In 17 cases, AAV2 was detected using targeted  
248 sequencing, while in seven of these pediatric cases AAV2 was missed by untargeted metagenomic  
249 sequencing, illustrating the significance of the use of enrichment by hybridization. With regard to cost-  
250 efficiency, a recent study compared PCR, sequence-independent single primer amplification (SISPA),  
251 and the Twist Comprehensive Viral Research Panel for the detection of Japanese encephalitis<sup>44</sup>. The  
252 authors concluded that the PCR panels were not able to detect all genotypes, whereas broader  
253 surveillance of vector-borne pathogens would be more effective though costly<sup>44</sup>. Hybridization  
254 capture has been approved by the FDA for SARS-CoV-2 variant monitoring, illustrating the  
255 acknowledged significance of this type of enrichment. The limit of detection of the SARS-CoV-2 specific  
256 hybridization method in their study was 800 copies/ml<sup>45</sup>, in line with our current and previous<sup>10,46</sup>  
257 findings when using the broader panel. Even using the panel designed in 2015<sup>14</sup> resulted in excellent  
258 genome coverage of SARS-CoV-2 due to sequence homology with animal coronaviruses and the  
259 variability in the probe design allowing for sequence mismatches<sup>10</sup>.

260 This study has several limitations. The synthetic sequences spiked in cell free human DNA did not  
261 contain other background nucleic acids such as bacterial and human RNA, though the latter proportion  
262 is generally low (<5%<sup>47</sup> dependent on the sample type). Furthermore, though ssRNA, dsDNA, and  
263 ssDNA viruses were analyzed, detection and LOD results cannot be directly extrapolated to every  
264 single virus. These parameters may vary to some extent for different viruses, particularly those not  
265 included in the synthetic controls (manufactured by Twist). This was also exemplified by the lack of  
266 detection of EAV and PhHV using the Twist Comprehensive Viral Research panel. Though these viruses  
267 are not considered human pathogens, this illustrates the presence of certain restrictions with regard  
268 to the animal viruses included in the panel. Further analyses of the lists of viruses delivered by the  
269 probe designers showed that all pathogens on the WHO list of diseases with pandemic potential  
270 ([https://www.who.int/news/item/21-11-2022-who-to-identify-pathogens-that-could-cause-future-](https://www.who.int/news/item/21-11-2022-who-to-identify-pathogens-that-could-cause-future-outbreaks-and-pandemics)  
271 [outbreaks-and-pandemics](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.

272 To summarize, this study provides data supporting further steps towards widespread introduction of  
273 viral metagenomics for pathogen detection in clinical settings. In addition, it provides guidance for  
274 integration of probe hybridization methods in surveillance to track pathogens of pandemic potential  
275 in low biomass samples such as wastewater<sup>48</sup> and wild life swabs<sup>49</sup>.

276

277 Author's contribution

278 Conceptualization: JJC.V. Software: IS, Investigation; KM, AB. Methodology; KM, ECJC, JJC.V. Software;  
279 IS. Formal analysis: KM, IS. Supervision; ECC, AB, SAB, JJC.V. Visualization: KM, JJC.V. Roles/Writing -  
280 original draft; KM, JJC.V. Writing - review & editing: all authors.

281

282 Declaration of potential competing interest

283 DM and AB are employees of GenomeScan B.V. and provided the sequencing service. They were not  
284 involved in bioinformatic/statistical data analysis, nor interpretation of results.

285

286 Funding

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

289

290 **Table 1.** Sensitivity and specificity resulting from varying thresholds based on **a**, sequence read counts  
 291 and **b**, genome coverage percentage using a random selection of 1 million sequence reads per dataset,  
 292 for the capture probe based metagenomic workflows SeqCap EZ HyperCap (Roche) and Twist  
 293 Comprehensive Viral Research Panel workflow. Corresponding ROCs are shown in **Fig. 2**. RPM; read  
 294 counts per million, LOD; limit of detection. For all tables, a minimum of three distributed regions of  
 295 the genome covered was set as primary parameter for defining detection.

296 **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)*	RNA: 10 <sup>1</sup> DNA: 10 <sup>2</sup>	RNA: 10 <sup>1</sup> DNA: 10 <sup>2-3</sup>	RNA: 10 <sup>1</sup> DNA: 10 <sup>2-3</sup>	RNA: 10 <sup>2-4</sup> DNA: 10 <sup>4</sup>
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)*	RNA viruses: 10 <sup>2</sup> DNA viruses: 10 <sup>2-3</sup>	RNA: 10 <sup>2</sup> DNA: 10 <sup>2-3</sup>	RNA: 10 <sup>2</sup> DNA: 10 <sup>2-4</sup>	RNA: 10 <sup>2-4</sup> DNA: 10 <sup>4-&gt;</sup>

297 **b**

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)*	RNA viruses: 10 <sup>1</sup> DNA viruses: 10 <sup>2</sup>	RNA: 10 <sup>1-2</sup> DNA: 10 <sup>2</sup>	RNA: 10 <sup>2</sup> DNA: 10 <sup>2</sup>	RNA: 10 <sup>3</sup> DNA: 10 <sup>3-10<sup>4</sup></sup>
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)*	RNA: 10 <sup>2</sup> DNA: 10 <sup>2-3</sup>	RNA: 10 <sup>3-10<sup>4</sup></sup> DNA: 10 <sup>2-3</sup>	RNA: 10 <sup>3-10<sup>4</sup></sup> DNA: 10 <sup>2-6</sup>	RNA: 10 <sup>4</sup> DNA: 10 <sup>3-&gt;4</sup>

298 \*Based on LOD of Inf A, SARS-CoV-2, EBV, HBV, and bocavirus. Inf A; influenza A virus, EBV, Epstein-  
 299 Barr virus, HBV, hepatitis B virus.

## 300 References

301

302 1 Gauthier, N. P. G., Chorlton, S. D., Kraiden, M. & Manges, A. R. Agnostic Sequencing for  
303 Detection of Viral Pathogens. *Clin Microbiol Rev* **36**, e0011922 (2023).  
304 <https://doi.org/10.1128/cmr.00119-22>

305 2 Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat Rev Genet* **20**, 341-355 (2019).  
306 <https://doi.org/10.1038/s41576-019-0113-7>

307 3 Deng, X. *et al.* Metagenomic sequencing with spiked primer enrichment for viral diagnostics  
308 and genomic surveillance. *Nat Microbiol* **5**, 443-454 (2020). [https://doi.org/10.1038/s41564-](https://doi.org/10.1038/s41564-019-0637-9)  
309 [019-0637-9](https://doi.org/10.1038/s41564-019-0637-9)

310 4 Wilson, M. R. *et al.* Clinical Metagenomic Sequencing for Diagnosis of Meningitis and  
311 Encephalitis. *N Engl J Med* **380**, 2327-2340 (2019). <https://doi.org/10.1056/NEJMoa1803396>

312 5 Carbo, E. C. *et al.* Viral metagenomic sequencing in the diagnosis of meningoencephalitis: a  
313 review of technical advances and diagnostic yield. *Expert Rev Mol Diagn* **21**, 1139-1146 (2021).  
314 <https://doi.org/10.1080/14737159.2021.1985467>

315 6 Carr, V. R. & Chaguza, C. Metagenomics for surveillance of respiratory pathogens. *Nat Rev*  
316 *Microbiol* **19**, 285 (2021). <https://doi.org/10.1038/s41579-021-00541-8>

317 7 Kafetzopoulou, L. E. *et al.* Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa  
318 fever outbreak. *Science* **363**, 74-77 (2019). <https://doi.org/10.1126/science.aau9343>

319 8 Holmes, E. C. COVID-19-lessons for zoonotic disease. *Science* **375**, 1114-1115 (2022).  
320 <https://doi.org/10.1126/science.abn2222>

321 9 Morfopoulou, S. *et al.* Genomic investigations of unexplained acute hepatitis in children.  
322 *Nature* **617**, 564-573 (2023). <https://doi.org/10.1038/s41586-023-06003-w>

323 10 Carbo, E. C. *et al.* Coronavirus discovery by metagenomic sequencing: a tool for pandemic  
324 preparedness. *J Clin Virol* **131**, 104594 (2020). <https://doi.org/10.1016/j.jcv.2020.104594>

325 11 Lopez-Labrador, F. X. *et al.* Recommendations for the introduction of metagenomic high-  
326 throughput sequencing in clinical virology, part I: Wet lab procedure. *J Clin Virol* **134**, 104691  
327 (2021). <https://doi.org/10.1016/j.jcv.2020.104691>

328 12 van Rijn, A. L. *et al.* The respiratory virome and exacerbations in patients with chronic  
329 obstructive pulmonary disease. *PLoS One* **14**, e0223952 (2019).  
330 <https://doi.org/10.1371/journal.pone.0223952>

331 13 Metsky, H. C. *et al.* Capturing sequence diversity in metagenomes with comprehensive and  
332 scalable probe design. *Nat Biotechnol* **37**, 160-168 (2019). [https://doi.org/10.1038/s41587-](https://doi.org/10.1038/s41587-018-0006-x)  
333 [018-0006-x](https://doi.org/10.1038/s41587-018-0006-x)

334 14 Briese, T. *et al.* Virome Capture Sequencing Enables Sensitive Viral Diagnosis and  
335 Comprehensive Virome Analysis. *mBio* **6**, e01491-01415 (2015).  
336 <https://doi.org/10.1128/mBio.01491-15>

337 15 Wylie, T. N., Wylie, K. M., Herter, B. N. & Storch, G. A. Enhanced virome sequencing using  
338 targeted sequence capture. *Genome Res* **25**, 1910-1920 (2015).  
339 <https://doi.org/10.1101/gr.191049.115>

340 16 Kuchinski, K. S. *et al.* Targeted genomic sequencing with probe capture for discovery and  
341 surveillance of coronaviruses in bats. *Elife* **11** (2022). <https://doi.org/10.7554/eLife.79777>

342 17 Yamaguchi, J. *et al.* Universal Target Capture of HIV Sequences From NGS Libraries. *Front*  
343 *Microbiol* **9**, 2150 (2018). <https://doi.org/10.3389/fmicb.2018.02150>

344 18 Wang, H. *et al.* Multiple-probe-assisted DNA capture and amplification for high-throughput  
345 African swine fever virus detection. *Appl Microbiol Biotechnol* **107**, 797-805 (2023).  
346 <https://doi.org/10.1007/s00253-022-12334-x>

347 19 Doddapaneni, H. *et al.* Oligonucleotide capture sequencing of the SARS-CoV-2 genome and  
348 subgenomic fragments from COVID-19 individuals. *PLoS One* **16**, e0244468 (2021).  
349 <https://doi.org/10.1371/journal.pone.0244468>

- 350 20 Lythgoe, K. A. *et al.* SARS-CoV-2 within-host diversity and transmission. *Science* **372** (2021).  
351 <https://doi.org/10.1126/science.abg0821>
- 352 21 Carbo, E. C. *et al.* Improved diagnosis of viral encephalitis in adult and pediatric hematological  
353 patients using viral metagenomics. *J Clin Virol* **130**, 104566 (2020).  
354 <https://doi.org/10.1016/j.jcv.2020.104566>
- 355 22 Schuele, L. *et al.* Assessment of Viral Targeted Sequence Capture Using Nanopore Sequencing  
356 Directly from Clinical Samples. *Viruses* **12** (2020). <https://doi.org/10.3390/v12121358>
- 357 23 Wylie, K. M. *et al.* Detection of Viruses in Clinical Samples by Use of Metagenomic Sequencing  
358 and Targeted Sequence Capture. *J Clin Microbiol* **56** (2018).  
359 <https://doi.org/10.1128/JCM.01123-18>
- 360 24 Jansen, S. A. *et al.* Broad Virus Detection and Variant Discovery in Fecal Samples of  
361 Hematopoietic Transplant Recipients Using Targeted Sequence Capture Metagenomics. *Front*  
362 *Microbiol* **11**, 560179 (2020). <https://doi.org/10.3389/fmicb.2020.560179>
- 363 25 Stout, M. J., Brar, A. K., Herter, B. N., Rankin, A. & Wylie, K. M. The plasma virome in  
364 longitudinal samples from pregnant patients. *Front Cell Infect Microbiol* **13**, 1061230 (2023).  
365 <https://doi.org/10.3389/fcimb.2023.1061230>
- 366 26 Garand, M. *et al.* Virome Analysis and Association of Positive Coxsackievirus B Serology during  
367 Pregnancy with Congenital Heart Disease. *Microorganisms* **11** (2023).  
368 <https://doi.org/10.3390/microorganisms11020262>
- 369 27 Flerlage, T. *et al.* Single cell transcriptomics identifies distinct profiles in pediatric acute  
370 respiratory distress syndrome. *Nat Commun* **14**, 3870 (2023).  
371 <https://doi.org/10.1038/s41467-023-39593-0>
- 372 28 Cassidy, H. *et al.* Exploring a prolonged enterovirus C104 infection in a severely ill patient using  
373 nanopore sequencing. *Virus Evol* **8**, veab109 (2022). <https://doi.org/10.1093/ve/veab109>
- 374 29 Esnault, G. *et al.* Assessment of Rapid MinION Nanopore DNA Virus Meta-Genomics Using  
375 Calves Experimentally Infected with Bovine Herpes Virus-1. *Viruses* **14** (2022).  
376 <https://doi.org/10.3390/v14091859>
- 377 30 Zhang, C. *et al.* Characterization of the Eukaryotic Virome of Mice from Different Sources.  
378 *Microorganisms* **9** (2021). <https://doi.org/10.3390/microorganisms9102064>
- 379 31 Junier, T. *et al.* Viral Metagenomics in the Clinical Realm: Lessons Learned from a Swiss-Wide  
380 Ring Trial. *Genes (Basel)* **10** (2019). <https://doi.org/10.3390/genes10090655>
- 381 32 van Boheemen, S. *et al.* Retrospective Validation of a Metagenomic Sequencing Protocol for  
382 Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric  
383 Patients. *J Mol Diagn* **22**, 196-207 (2020). <https://doi.org/10.1016/j.jmoldx.2019.10.007>
- 384 33 Morfopoulou, S. *et al.* Deep sequencing reveals persistence of cell-associated mumps vaccine  
385 virus in chronic encephalitis. *Acta Neuropathol* **133**, 139-147 (2017).  
386 <https://doi.org/10.1007/s00401-016-1629-y>
- 387 34 Carbo, E. C. *et al.* Longitudinal Monitoring of DNA Viral Loads in Transplant Patients Using  
388 Quantitative Metagenomic Next-Generation Sequencing. *Pathogens* **11** (2022).  
389 <https://doi.org/10.3390/pathogens11020236>
- 390 35 Reyes, A. *et al.* Viral metagenomic sequencing in a cohort of international travellers returning  
391 with febrile illness. *J Clin Virol* **143**, 104940 (2021). <https://doi.org/10.1016/j.jcv.2021.104940>
- 392 36 Carbo, E. C. *et al.* Performance of Five Metagenomic Classifiers for Virus Pathogen Detection  
393 Using Respiratory Samples from a Clinical Cohort. *Pathogens* **11** (2022).  
394 <https://doi.org/10.3390/pathogens11030340>
- 395 37 de Vries, J. J. C. *et al.* Benchmark of thirteen bioinformatic pipelines for metagenomic virus  
396 diagnostics using datasets from clinical samples. *J Clin Virol* **141**, 104908 (2021).  
397 <https://doi.org/10.1016/j.jcv.2021.104908>
- 398 38 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-  
399 359 (2012). <https://doi.org/10.1038/nmeth.1923>

400 39 Vilsker, M. *et al.* Genome Detective: an automated system for virus identification from high-  
401 throughput sequencing data. *Bioinformatics* **35**, 871-873 (2019).  
402 <https://doi.org/10.1093/bioinformatics/bty695>

403 40 de Vries, J. J. C. *et al.* Recommendations for the introduction of metagenomic next-generation  
404 sequencing in clinical virology, part II: bioinformatic analysis and reporting. *J Clin Virol* **138**,  
405 104812 (2021). <https://doi.org/10.1016/j.jcv.2021.104812>

406 41 Lewandowska, D. W. *et al.* Optimization and validation of sample preparation for  
407 metagenomic sequencing of viruses in clinical samples. *Microbiome* **5**, 94 (2017).  
408 <https://doi.org/10.1186/s40168-017-0317-z>

409 42 Kim, K. W. *et al.* Respiratory viral co-infections among SARS-CoV-2 cases confirmed by virome  
410 capture sequencing. *Sci Rep* **11**, 3934 (2021). <https://doi.org/10.1038/s41598-021-83642-x>

411 43 Servellita, V. *et al.* Adeno-associated virus type 2 in US children with acute severe hepatitis.  
412 *Nature* **617**, 574-580 (2023). <https://doi.org/10.1038/s41586-023-05949-1>

413 44 Crispell, G. *et al.* Method comparison for Japanese encephalitis virus detection in samples  
414 collected from the Indo-Pacific region. *Front Public Health* **10**, 1051754 (2022).  
415 <https://doi.org/10.3389/fpubh.2022.1051754>

416 45 Nagy-Szakal, D. *et al.* Targeted Hybridization Capture of SARS-CoV-2 and Metagenomics  
417 Enables Genetic Variant Discovery and Nasal Microbiome Insights. *Microbiol Spectr* **9**,  
418 e0019721 (2021). <https://doi.org/10.1128/Spectrum.00197-21>

419 46 Carbo, E. C. *et al.* A comparison of five Illumina, Ion Torrent, and nanopore sequencing  
420 technology-based approaches for whole genome sequencing of SARS-CoV-2. *Eur J Clin*  
421 *Microbiol Infect Dis* **42**, 701-713 (2023). <https://doi.org/10.1007/s10096-023-04590-0>

422 47 Wu, J. *et al.* Ribogenomics: the science and knowledge of RNA. *Genomics Proteomics*  
423 *Bioinformatics* **12**, 57-63 (2014). <https://doi.org/10.1016/j.gpb.2014.04.002>

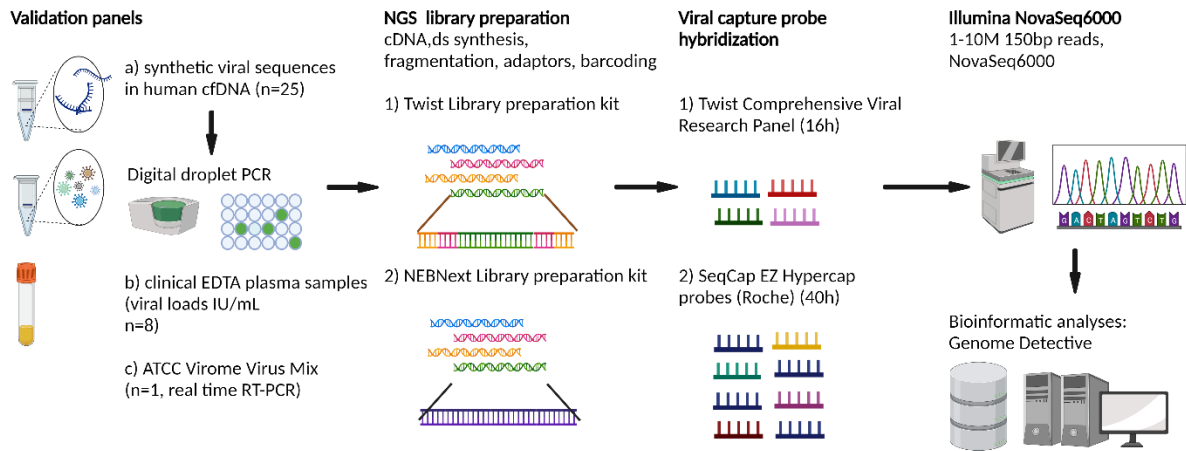
424 48 Clark, J. R. *et al.* Wastewater pandemic preparedness: Toward an end-to-end pathogen  
425 monitoring program. *Front Public Health* **11**, 1137881 (2023).  
426 <https://doi.org/10.3389/fpubh.2023.1137881>

427 49 Lwande, O. W. *et al.* Alphacoronavirus in a Daubenton's Myotis Bat (*Myotis daubentonii*) in  
428 Sweden. *Viruses* **14** (2022). <https://doi.org/10.3390/v14030556>

429

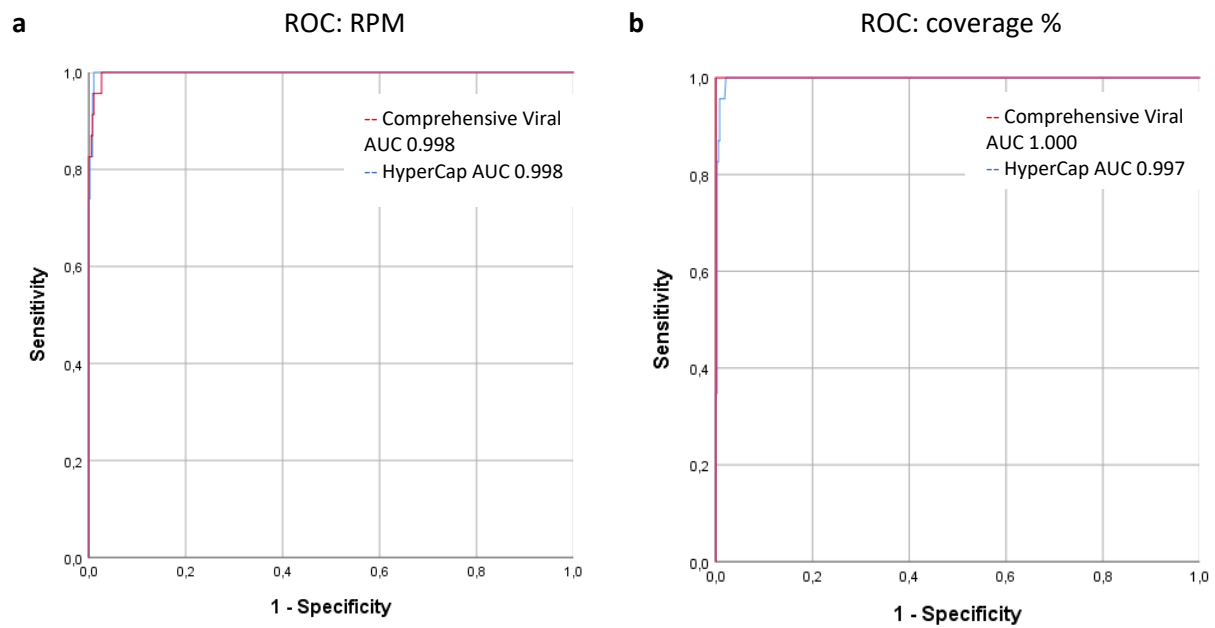


## Study design and workflow

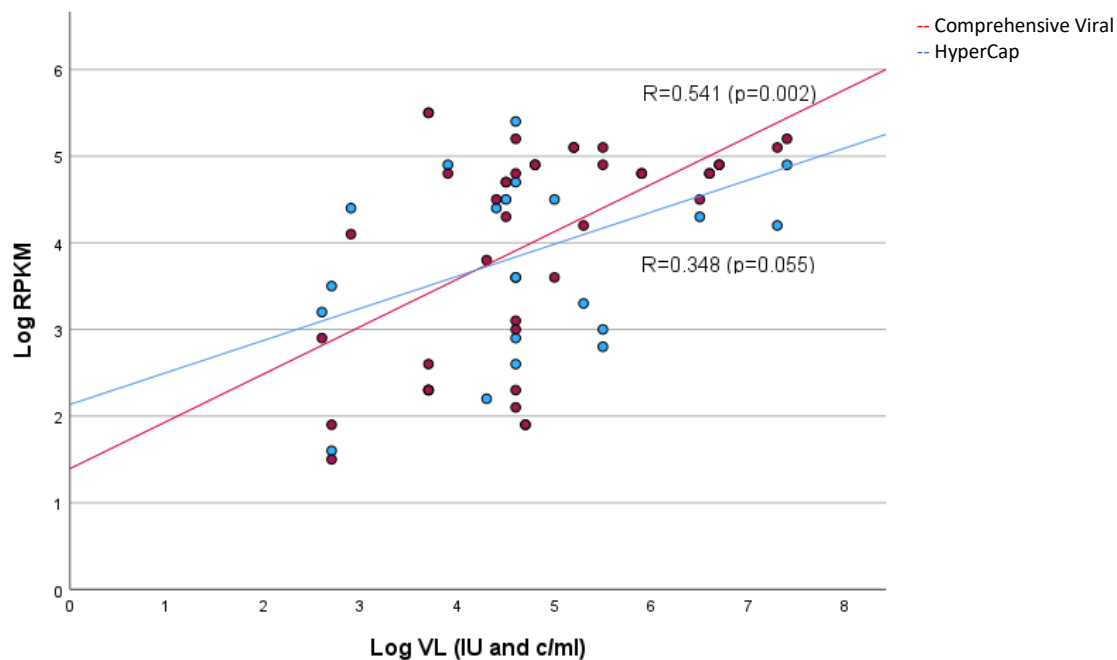


**Fig. 1. Workflows of the capture probe-based targeted metagenomic protocols compared in this study, Twist Comprehensive Viral Research, and the SeqCap EZ HyperCap (ViroCap, Roche) and, both in combination with identical bioinformatic analyses pipeline. Created using BioRender.**

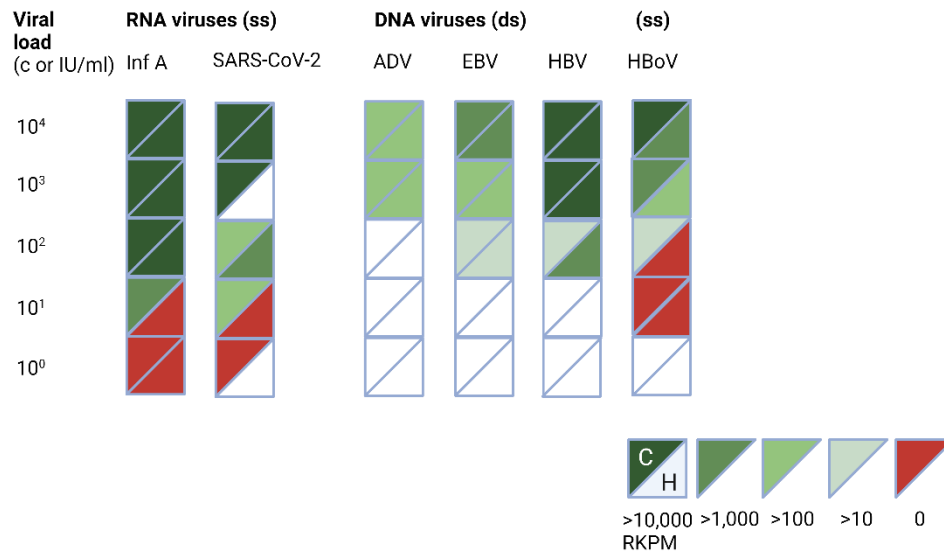




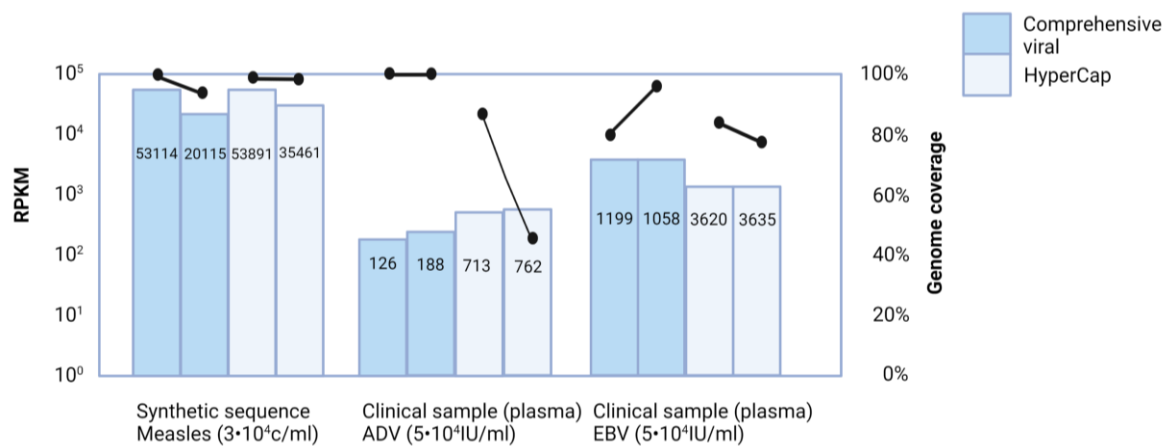
**Fig. 2. Receiver operating characteristic curves (ROC) for prediction of detection of viral sequences** using the virus capture probe based metagenomic workflows Twist Comprehensive Viral Research, and SeqCap EZ HyperCap (Roche). The validation panel consisted of synthetic viral sequences spiked in a background of human cell-free DNA (90-99.999%) and diluted ATCC Virome virus mix standard (copies/mL ranging from  $10^4$  to  $10^7$ ). **a**, ROC based on sequence read counts per million (RPM), and **b**, percentage of genome coverage, using a random selection of 1 million sequence reads per dataset. For all curves a minimum of three distributed regions of the genome covered was set as primary parameter for defining detection.



**Fig. 3. Correlation graph depicting linearity between the viral load (VL,  $\log_{10}$  IU and C/ml, horizontally) and the  $\log_{10}$  read counts per million per kb genome (RPKM) 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%), dilution series (see **Fig. 4**), and clinical samples.**



**Fig. 4. Limit of detection 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'). Read counts per million per kb genome (RPKM) are shown for different viral loads (C/ml). The samples consisted of synthetic viral sequences spiked in a background of human cell-free DNA (90-99,999%) (Inf A, SARS-CoV-2, HBoV), and clinical EDTA plasma samples (ADV, EBV, HBV). Created using BioRender.



**Fig. 5. Reproducibility of read counts and genome coverage percentages.** Between-run variability in RPKM (left axis) and genome coverage percentage (right axis) as generated using the virus capture probe based metagenomic workflows Twist Comprehensive Viral Research and SeqCap EZ HyperCap (Roche). Percentage of genome coverage was based on a random selection of 1 million sequence reads per dataset. Coefficients of variance in RPKM ranged from 0.0-4.7% (see **Suppl. Table 1**). Created using BioRender.