Next-Generation Sequencing Methods for Sensitive Characterisation of Hepatitis B Viral Genomes: A European Multicentre Study

Comparison of NGS Workflows for Detection of HBV

Michael X. Fu[1]*, Maria F. Perdomo[2], Sheila F. Lumley[1,3], Johan Ringlander[4,5], Kai Kean[1], Kaitlin Reid[1], Richard Mayne[1], Oscar Enrique Torres Montaguth[6], Leysa Forrest[7], Sarah Buddle[6], Johannes C. Botha[6,8], Joakim B. Stenbäck[4,5], Zachery Dickson[2], Chris Kent[9], Haiting Chai[1], Matthew Byott[6,8], Leo Hannolainen[2], Shannah Secret[1], George Airey[1], Klaus Hedman[2], Monique I. Andersson[3,10], M. Azim Ansari[1], Eleni Nastouli[6,8], Judith Breuer[6], Philippa C. Matthews[11,12], Tanya Golubchik[1,13], William L. Irving[14], Peter Simmonds[1], Heli Harvala[10,15,16]*

1. Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
2. Department of Virology, University of Helsinki, Helsinki, Finland
3. Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
4. Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Sweden
5. Institute of Biomedicine, Department of Infectious Diseases, University of Gothenburg, Gothenburg, Sweden
6. Infection, Immunity and Inflammation Department, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom
7. Genetics and Genomic Medicine Department, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

8. Advanced Pathogen Diagnostics Unit, University College London Hospitals NHS Trust, London, United Kingdom
9. Institute of Microbiology and Infection, University of Birmingham, Birmingham, United Kingdom
10. Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom
11. Division of Infection and Immunity, University College London, London, United Kingdom
12. The Francis Crick Institute, London, United Kingdom
13. Sydney Infectious Diseases Institute, School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
14. NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, United Kingdom
15. Institute of Biomedicine, University of Turku and University Hospital of Turku, Turku, Finland
16. Microbiology Services, NHS Blood and Transplant, Colindale, United Kingdom

*Corresponding authors: Michael X Fu, michael.fu@ndm.ox.ac.uk, Heli Harvala, heli.harvala@utu.fi

Keywords: high-throughput sequencing; sensitivity; virus detection; viral metagenomics; diagnostics; clinical microbiology; Nanopore; Illumina.

Acknowledgements: The authors thank Dr Belinda Singleton (NHS Blood and Transplant) for helping develop the anellovirus PCR protocol. This study was primarily supported by the National Institute for Health and Care Research [grant number NIHR203338], which supported these authors under the Blood and Transplant Research Unit ‘Genomics to

Enhance Microbiology Screening': MXF, KK, KR, RM, OETM, SB, SS, JBr, TG, PS, and HH. The study was also supported by the Swedish Society for Medical Research (grant number PG-23-0435) and the Gothenburg Society for Medicine (grant number GLS-1001038). SFL is funded by a Wellcome Doctoral Training Fellowship (grant number 220549/Z/20/Z); this research was funded in part by the Wellcome Trust. MFP and LH acknowledge funding from Suomen Lääketieteen säätiö and Medicinska Understödsföreningen Liv & Hälsa. KH and ZD acknowledge funding from the Jane and Aatos Erkko Foundation. TG is supported by an Investigator Grant (GNT2025445) from the National Health and Medical Research Council, Australia (NHMRC). PCM receives core funding from the Francis Crick Institute (ref CC2223) and acknowledges funding support from the University College London NIHR Biomedical Research Centre (BRC). MAA is supported by a Sir Henry Dale Fellowship jointly funded by the Royal Society and Wellcome Trust (220171/Z/20/Z). The funding bodies had no role in the study's design, data collection, analysis, or manuscript writing. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions: Conceptualisation: HH, PS, MXF. Formal analysis: MXF, HH, PS, WLI. Investigation: MXF, MFP, SFL, JR, KK, KR, RM, OETM, LF, SB, JBo, JBS, ZD, CK, HC, MB, LH, SS, GA, KH, MAA, EN, JBr, PM, TG. Writing – Original Draft: MXF. Writing – Review & Editing: all authors. Supervision: HH, PS, MA.

Conflicts of Interest: The authors declare no conflicts of interest that pertain to this work.

Abstract

Background: Choosing a next-generation sequencing (NGS) workflow to enable sensitive and rapid detection of infectious agents remains critical. This collaborative study compared multiple NGS methods to detect and generate hepatitis B virus (HBV) genomes in samples of low viral load.

Methods: 23 HBV DNA-positive plasma samples of genotypes A-E (0.2 to 6207 IU/ml) and one control sample were assayed blindly via 9 NGS methods from 6 European laboratories. Methods included untargeted metagenomics, pre-enrichment by probe-capture followed by Illumina sequencing, and HBV-specific PCR pre-amplification followed by sequencing with Nanopore or Illumina. Construction of consensus sequences was performed at the coordinating centre.

Results: Full HBV genomes were constructed at viral loads >1000 IU/ml for probe-capture methods, >200 IU/ml for PCR-Illumina methods, >10 IU/ml for PCR-Nanopore methods, and in no samples for metagenomic methods. Contamination was observed in the negative control and samples with very low viral loads for PCR-based methods. Probe-capture and metagenomic methods detected additional viruses not routinely screened in blood donations; positive results were confirmed by PCR. Costs were lowest for PCR-Nanopore, and turnaround time was highest for probe-capture methods.

Conclusion: Different methods have different advantages, and the optimal method depends on the context. NGS has the potential to delineate whole-genome sequences at low viral loads if supported by a PCR pre-amplification step. Probe-capture methods also reliably detect HBV at low viral loads but limit genome characterisation while accommodating the

incidental detection of other virus species. Stringent steps are needed to prevent cross-contamination or bioinformatic noise to maximise diagnostic accuracy.

Importance

There is a great need in clinical microbiology and public health to sequence whole genomes of different viruses. Because so many different methods can be used for sequencing, it can be difficult to choose a suitable method. We compared commonly used methods on blood donor samples with low levels of hepatitis B virus DNA. We found that each method had its own benefits; however, our study highlights the need for increased vigilance regarding the risk of contamination. Our study provides necessary data for clinical and research laboratory teams to make informed decisions about the most suitable method. This would advance our understanding of hepatitis B and other viruses, ultimately benefiting both current and future patients.

Introduction

The advent of high-throughput sequencing platforms in the mid-2000s, coined 'next-generation sequencing' (NGS), has made dramatic advances in the characterisation of host and pathogen genomes in health and disease[1]. Metagenomics, a commonly used NGS method to sequence all genomic material in a sample, has been instrumental in detecting broad-spectrum microbes, surveillance of emerging infections, and identifying novel pathogens[2,3]. Compared to Sanger sequencing where sub-genomic amplicons spanning regions of interest are sequenced, the ability to sequence complete microbial genomes using NGS in an unbiased manner has proven to be a powerful tool[2,3]. The wider adoption of NGS in clinical microbiology and public health has been hindered by complex methodological workflows, complications in the detection and reporting of contaminating sequences, and a reduced sensitivity compared to real-time and nested PCR assays[2,4]. In particular, the inadequate sensitivity of untargeted metagenomics has limited the detection of hepatitis B virus (HBV), where many HBV-infected clinical samples contain low copy numbers of the viral genome[5,6].

HBV remains a global public health issue despite the availability of a prophylactic vaccine and antiviral treatments; the virus chronically infects around 246 million individuals worldwide, accounting for 1.1 million HBV-related deaths in 2022[7]. In the blood, HBV exists in relaxed-circular DNA (rcDNA) form: a partially double-stranded genome of approximately 3.2 kb, consisting of overlapping reading frames encoding four genes: polymerase, surface, precore/core, and X. HBV can be classified into 10 genotypes (A to J) based upon sequence divergence ≥8%, with inter-genotypic differences in geographic distribution, transmission mode, and clinical outcomes[8]. Sequencing full-length HBV genomes can help infer accurate (sub)genotypes and recombinants, better understand mechanisms of

transmission and persistence, identify treatment-resistant and immune-escape mutations, and develop new therapeutics[9]. Furthermore, whole-genome sequencing (WGS) of low viral load (VL) samples enables the characterisation of occult HBV infection and the assessment of viral clearance and reactivation, and may be progressively relevant to the field of functional cure.

Methodologies for sequencing viral genomes need to be time- and cost-efficient, specific, and sensitive[10]. A comprehensive evaluation of different methodologies is vital to inform their wider uptake in clinical, public health and research laboratories. Previous studies have been limited to individual workflows and laboratories, with few coordinated comparisons. When protocols were compared, studies often used synthetic materials with lower sample diversity due to limited clinical sample availability[11]. A comparable analysis of the effectiveness of NGS methods for the characterisation of hepatitis C virus was conducted almost a decade ago[12]. Another comparison only investigated samples with relatively high VLs and did not provide information on the detection limits of different NGS methods[11].

Amid calls for the HBV field to work progressively toward developing NGS methods[8], this study brought together six European laboratories with nine established virus sequencing protocols, encompassing four commonly used NGS methods. First, the untargeted metagenomics method sequenced all genomic material in the samples. The second method involved the addition of biotinylated nucleic acid probes to enrich for large numbers of known pathogens and selectively capture target sequences before Illumina sequencing. The final methods involved PCR amplification of HBV fragments, followed by sequencing using the short-read Illumina or the long-read Oxford Nanopore Sequencing (ONT)

platforms. The latter three methods have previously been shown to be sensitive for sequencing HBV genomes[3,6,10,13].

The four NGS methods were evaluated using a panel of 23 HBV DNA-positive plasma samples with low VLs. We compared their utility in detecting and generating HBV sequences and co-detecting other virus species.

Materials and Methods

Samples. Twenty-three large-volume plasma samples from individuals infected with HBV genotypes A, B, C, D, and E were obtained from blood donors from NHS Blood and Transplant (NHSBT). To reflect the heterogeneity in the extraction methods used by NGS protocols and account for the non-negligible stochastic effect of measuring low VLs, geometric mean VLs were calculated across eight separate real-time PCR measurements from eight extractions (five of which were obtained from a previous investigation[14] and three further extractions from 5 ml of plasma). The samples had geometric mean VLs ranging from 0.2 to 6207 IU/ml. Furthermore, a control plasma sample, negative for HBV markers and other screened blood-borne infections, was obtained from NHSBT. All donors provided signed consent to use their samples for research purposes. This study was approved by the Blood Supply Clinical Audit, Risk and Effectiveness Committee of NHSBT (numbers 22031 and 23002). All 24 samples, including the negative control, were sent to each laboratory on dry ice, with anonymised sample identifiers and no other sample information.

Sequencing methods. The panel of 24 blinded samples was used to evaluate the performance of nine established NGS protocols at six expert European laboratories. Methods included two untargeted metagenomic protocols using Illumina (MTG-A and MTG-

B), three pan-viral probe-based enrichment protocols using Illumina (TAC-A, TAC-B, and TAC-C), two HBV-specific PCR amplification protocols using Illumina (PCR-A and PCR-B [PCR-Illumina]) and two HBV-specific PCR amplification protocols using ONT (PCR-1 and PCR-2 [PCR-ONT]). Nucleic acid extractions were performed at individual laboratories, except for three protocols for which extractions were conducted separately in the coordinating laboratory in Oxford (PCR-A, PCR-B, PCR-2). The input plasma volumes for extractions ranged from 0.2 to 20 ml (Table 1). Extracted nucleic acids underwent library preparation according to local protocols, where input volumes ranged from 5 to 50 µl. Effective test volumes ranged from 4 to 800 µl. Sequencing platforms and average sequencing depth per sample differed between methods (Table 1). Full details of protocols are found in the supplementary material. The pre-capture libraries for TAC-B were sequenced for MTG-B; thus, the samples were prepared similarly for both protocols.

Factors relating to the technical practicality and costs associated with the use of NGS assays greatly influence their adoption for clinical diagnostics. Laboratories were asked to provide the total time and cost of each protocol step from extraction to sequencing on a standardised template document, considering the typical number of samples that would normally be included per run for each protocol. Times for each stage were separated into per-person labour time and waiting times (which included machine running times), and costs were standardised using a 1:0.90:0.75 USD:EUR:GBP exchange rate, as of May 18, 2025.

Bioinformatic processing. Post-read filtering and genome assembly, either mapping to the closest available reference or *de novo*, were performed using established protocol-specific pipelines at each centre (Table 1). After deduplication of reads or removal of PCR

duplicates and processing of sequences to begin at the EcoR1 restriction site (the conventional origin of the genome[15]), CSV files for each assembled genome/sample were sent from each centre, containing the number of each base called at each nucleotide site. These files were received at the coordinating laboratory, where further bioinformatics analyses were performed using a common set of tools previously described[12]. A 95% majority base consensus sequence was calculated at each genome position possessing one or more base reads, incorporating the calling of ambiguous bases when necessary. Constructed assembled sequences for each sample were examined and realigned to ensure consistent numbering, compare nucleotide similarities, and assess genome coverage and accuracy. Within-sample consensus sequences were assembled for each sample from the different NGS protocols by combining more than one similar assembled sequence with genome coverage of around 100% using the program Sequence Join in the SSE package[16], followed by visual inspection of sites with ambiguous bases. Specific sequences from each method were compared with the within-sample consensus, and the numbers of nucleotide differences were recorded using the program Sequence Dist[16]. Missing bases were not included in comparisons of sequence identity. Whilst a threshold of one base read per nucleotide site allowed for representation of all called bases in all protocols regardless of background contamination or sequencing errors, a more stringent minimum threshold of 10 base reads at each nucleotide site was also applied; the re-assembled sequences were re-aligned to ascertain potential differences when utilising a higher threshold. Whilst this stringent threshold was applied, such a high threshold is not advocated for most targeted enrichment or untargeted metagenomics purposes of low VL samples.

Nested PCR and Sanger sequencing. Sequences of a portion of the surface gene obtained previously from extensive nested PCR and Sanger sequencing were included for comparison to the NGS sequences[17] (see Figure S1 for phylogenetic tree of all HBV-positive samples obtained via Sanger sequencing of nested PCR amplicons). Using DNA extracted from 5 ml of plasma[14], nested PCR and sequencing of two further regions were performed: one spanning the surface and polymerase genes, and another spanning the X and precore/core genes (see Table S1 for primer details and nucleotide positions). Within-sample consensus NGS sequences were genotyped as previously described[14]. When whole-genome assemblies could not be constructed (i.e. no two complete NGS sequences were similar), genotypes were defined from either the most complete NGS sequence or Sanger sequence, the latter if there were no similar NGS sequences.

Definitions. The following terms used in this manuscript have been defined to aid readability.

Protocol: unique individual NGS workflows performed by individual laboratories.

Method: a collection of similar protocols based on sequencing technology and platform. Methods included untargeted metagenomics followed by Illumina sequencing (MTG), targeted metagenomics using probe capture followed by Illumina sequencing (TAC), PCR amplification of HBV before Illumina sequencing (PCR-Illumina), and PCR amplification of HBV before ONT sequencing (PCR-ONT).

Genotype: distinct genetic variations of HBV based on nucleotide differences of ≥8%.

Subgenotype: subgroups within specific HBV genotypes differing by 4-8% in viral genome sequence and distinct from the broader HBV genotypes.

Strain: a sequence that differed by >5% from other sequences from the same sample, indicating a different (sub)genotype.

Reference sequence: a known digital HBV DNA sequence used as a standard for genome assembly.

Assembled sequence: individual sequences constructed from individual protocols, which underwent read filtering and assembly via protocol-specific bioinformatic pipelines, followed by calculation of majority base consensus sequence at the coordinating laboratory and realignment of assembled sequences where necessary.

Within-sample consensus sequence: a more robust complete genome sequence for each sample, derived from the most commonly observed nucleotide at each genome position across available assembled sequences from all protocols for the same sample.

Contamination/contaminating sequence: a genetically different sequence and strain from the within-sample consensus sequence or other assembled sequences, where applicable (>5% nucleotide differences).

True-positive: similar to other sequences from the same sample (<5% nucleotide differences).

False-positive sequence: detection of an HBV DNA sequence in the negative control.

HBV-specific bases read/reads: the total number of base reads at all nucleotide positions, after deduplication of sequencing reads and PCR duplicates, where applicable.

Genome coverage: the proportion of the viral genome with consensus bases called.

Shannon entropy: a measure of the diversity of base counts in a sequence, which may result from technical sequencing errors or natural variability within infecting viral populations and could also be influenced by VL. Shannon entropy value ranges from 0 (no variability) to 1 (equal frequencies of all bases).

Detection of other virus species by PCR. Since HBV was not the only possible virus sequenced from the five TAC and MTG protocols from three laboratories, the deduplicated read counts and genome coverage percentages were obtained for any additional viruses detected from the sample panel, following the full bioinformatics pipelines for each centre (see supplementary material for details of full pipelines). When a certain virus was detected by more than one NGS method in the same sample, confirmatory PCR was performed in all samples (see supplementary material for details of the PCR protocols).

Statistical analyses. Spearman's correlation test deduced the significance of the association between VLs and multiple parameters, where α was 0.05. Data analyses and visualisation were performed with GraphPad Prism (v10.4.2, LLC).

Results

HBV detection and accuracy of assembled sequences. HBV amplicons were obtained for up to 23/23 samples by Sanger sequencing, with good concordance between sequences generated by Sanger and NGS sequencing. Within-sample consensus NGS sequences were constructed for 14/15 samples with VL above 50 IU/ml (Table 2). Sequences assembled using different NGS protocols were generally identical or nearly identical to the within-sample consensus. A few individual sequences were >5% divergent from the within-sample consensus (shown in pink), indicating a different contaminating HBV strain. Assembled sequences were compared for eight samples with lower VLs, for which within-sample consensus sequences were unavailable. Different HBV strains were detected by at least one protocol (>5% divergent from the others) for all eight samples. Three samples, VLs 1.1, 1.8, and 8.5 IU/ml, yielded discrepant sequences in all NGS protocols.

All four PCR-based protocols detected false-positive HBV sequences in the negative control (genome coverage, 48%-100%), which were identified as different strains. The four protocols also detected numerous contaminating sequences in other samples, including those with low VL (<10 IU/ml). These four protocols were performed in three different laboratories, except for extraction, which the coordinating centre conducted for three of the four protocols. Some contaminants and false positives matched sequences of other *ex vivo* viruses (Table 2), but most were non-identical. For PCR-based methods other than PCR-1, the proportion of contaminants or the lack of detection at the lowest VLs deterred the ability to ascertain true-positive detection confidently. TAC methods had very few contaminating sequences, and no contaminating sequences were assembled from MTG methods. Utilising a 10-base instead of a 1-base threshold for calling consensus bases did not affect the detection of contamination in the negative control. However, the higher threshold eliminated the detection of contaminants in other samples for PCR-1, reduced the detection of

contaminants for PCR-2 (including reduced genome coverage), reduced the genome coverage for lower VL samples for PCR-B, and eliminated the detection of contaminants in TAC-C (Table S2).

Overall, true-positive HBV sequences of >5% genome coverage were detected in 5-6/23 samples using MTG, 10-17/23 samples using TAC, 9-13/23 samples using PCR-Illumina, and 16-19/23 samples using PCR-ONT protocols. MTG-B detected 7/13 samples at lower coverage and read numbers, which were detected by TAC-B, including the six samples with the highest coverage detected by TAC-B; samples for both protocols were prepared in the same run before the capture step.

HBV read depths and genome coverage. The highest number of HBV-specific bases was read in PCR methods (up to 10^9), followed by TAC (up to 10^7) and the lowest in MTG (up to 10^4). A strong correlation was observed between nucleotide counts and VL using the PCR-ONT and TAC methods, whereas this correlation was moderate when using the PCR-Illumina and MTG methods (Figure 1). Base counts were similar between samples for the PCR-Illumina methods. For PCR-ONT, bases obtained for true-positive samples with VL <50 IU/ml ranged from $1x10^4$ to $5x10^7$. For TAC and MTG methods, base counts were much lower for samples with lower VL. The approximate threshold to reliably detect the sample panel with true-positive sequences was >50 IU/ml or >10^6 bases for PCR-ONT methods, >50 IU/ml or >10^4 bases for TAC methods, >100 IU/ml or >10^7 bases for PCR-Illumina methods, and >3000 IU/ml or >10^3 bases for MTG methods. Utilising a 10-base instead of a 1-base threshold for calling consensus bases failed to detect HBV in the majority of samples for MTG methods, and removed the need for a VL threshold to detect

most of the sample panel with true-positive sequences for TAC and PCR-1 methods (Figure S2).

Following multiple alignment of read assemblies, the completeness of assembled sequences was analysed. With a minimum coverage threshold of 1 for calling consensus bases, complete or near-complete genome sequences were assembled for all samples with VL >10 IU/ml for PCR-1 protocol, >200 IU/ml for PCR-B, >1000 IU/ml for TAC-C, and in none of the samples for MTG methods (Figure 2). Sub-analyses using a minimum threshold of 10 bases found that coverage thresholds of 80% and 40% could be applied to exclude contamination in the PCR-Illumina and PCR-ONT methods, respectively (Figure S3).

Inspection of the mean genome coverage revealed that the methods yielded uneven coverage across the genome (Figure 3). Read depth was most uniform for PCR-ONT methods, while PCR-Illumina methods had more regions of reduced coverage. Coverage of the HBV genome was even sparser for the TAC and MTG methods. When analysing individual samples, there was considerable heterogeneity in the distribution of HBV genome coverage between protocols, without clear patterns (Figure S4).

Heterogeneity. The diversity of base counts at each site was quantified through calculations of Shannon entropy. Variability in Shannon values was observed between the NGS methods for the same samples, which may reflect the variations in the quantities of bases read by each method. MTG methods showed the least intra-sample HBV sequence heterogeneity, with a maximum mean Shannon entropy of 0.01, followed by the TAC methods (up to 0.04) and the PCR-Illumina methods (up to 0.09). However, heterogeneity was extremely high (up to 0.32) for PCR-ONT methods (Figure 4); sub-analyses using a

10-base threshold for calling consensus bases did not affect these values (Figure S5). The presence of ambiguous bases in sequences from PCR-ONT methods did not affect the perceived 95% accuracy of consensus sequences. Diversity and VL were poorly correlated for all methods, but diversity was lower when VL was higher for PCR-2 after applying a 10-base threshold.

Co-detection. Protocol-specific bioinformatic pipelines were employed to determine the presence, read counts, and genome coverage of other virus species for the TAC and MTG methods, which simultaneously screened for multiple viruses from generated sequences (Table 3). Altogether, anellovirus sequences were detected in 22/24 samples by the five NGS protocols. Anellovirus was detected by at least two protocols in 19/24 samples, and 2/24 samples by all five protocols. >50% genome coverage for anellovirus was observed in two samples. Confirmatory PCR for alpha, beta, and gammatorqueviruses revealed the presence of DNA sequences of all three anellovirus genera in all 24 samples (20/24 αTTV-positive αTTV; 22/24 βTTV-positive; 17/24 γTTV-positive), all with relatively high VLs (Ct values ranged from 4.7 to 26.1). Sanger sequencing of amplicons generated by nested PCR from randomly selected anellovirus-positive samples confirmed the presence of different strains from each genus.

Various human herpesviruses were detected in 17/24 samples across the five MTG and TAC protocols at <50% genome coverage, and were detected by more than one protocol for three samples (EBV in 4/5 protocols in two laboratories for sample HBV8; HHV6 in 2/5 protocols in two laboratories for HBV9; KSHV in 3/5 protocols in three laboratories for HBV12). Real-time PCR confirmed the presence of these herpesviruses in these three samples, but no other sample was HHV PCR-positive, including the 13 samples where

TAC-C reported HHV-7 DNA sequences. Polyomavirus DNA sequences were detected in 11/24 samples using NGS protocols; however, they were only detected by more than one protocol in one sample (HPyV6, detected by two protocols in two laboratories in HBV6). HyPV6 was not detected by real-time PCR. Instead, MCPyV was detected by real-time PCR in low copy numbers in HBV6 (detected by one NGS protocol) and HBV21 (not detected by any NGS protocol). Overall, TAC methods detected a greater number of viruses at higher read counts and coverage than MTG methods.

Time and costs. The per-person labour and waiting times varied between protocols (Figure 4a). The turnaround time was shortest for MTG and most PCR-based protocols. TAC methods took longer, mostly due to overnight hybridisation steps. Costs were lowest for PCR-ONT methods, higher for PCR-Illumina methods, followed by MTG methods, whilst TAC methods had the greatest cost (Figure 4b).

Discussion

To our knowledge, this is the first multicenter comparison of currently available NGS methods in low HBV DNA VL clinical samples. Distinctive differences were found between the methods' capabilities in generating accurate HBV sequences at the lowest VLs, detecting other virus species, and their labour and monetary burden.

NGS methods involving pre-amplification steps reliably detected HBV at VL ~10 IU/ml (for PCR-1 and TAC-C) or ~100 IU/mL (for other pre-amplification protocols), where detection was often non-specific at the lowest VLs for PCR-based methods but was most costly for TAC methods. Previous investigations of PCR-1 and PCR-2 also showed successful amplification of full HBV genomes at VL <30 IU/ml[6,13], and alternative probe-capture

methods observed a similar 50 copies/ml limit of detection for HBV[18] and 90-100% coverage for a range of DNA and RNA viruses at ≥1000 copies/ml[19,20]. The substantially reduced depth of coverage for MTG methods we observed has been well-described for a range of viruses[10–12]. However, the method retains its value in other applications due to its ability to detect unanticipated or potentially entirely novel pathogens without a requirement to pre-select targets in capture and PCR-assisted NGS methods[2].

Although detection sensitivity was similar, genomes generated from PCR-based methods were near-complete compared to lower genome coverage from TAC methods for our low VL samples. As an inexpensive method, PCR-ONT may be a beneficial screening tool for the HBV and microbiology fields, especially in low- and middle-income countries[21] when ONT allows preliminary analysis before a run is completed, which would be advantageous for high VL samples. Our findings contrast with a similar study on hepatitis C virus, which concluded that PCR-based NGS was relatively laborious[12]; improvements in technologies and workflows could explain differences found. However, PCR-based NGS would be more challenging for more divergent and/or larger viruses than HBV, as primer design for whole-genome amplification can be difficult, and amplicon drop-out necessitating primer re-design is not uncommon. Variables that may have affected methods' detection capabilities in this study include extraction from larger volumes of plasma (as previously found with PCR-1[13] and TAC-A/B protocols[17]), having greater sample representation and fewer samples for library preparation, and sequencing more gigabases per sample. For PCR-based methods, sensitivity is affected less by the sequencing platform and relies heavily on the specific PCR approach, particularly the conservation of primer-binding sites. A nested PCR strategy with shorter amplicons would improve sensitivity if WGS is not required.

The frequent detection of HBV contaminants regardless of the presence or absence of the virus demonstrates a substantial limitation to the diagnostic reliability of PCR-based methods. Without other sequence data for comparison, minimum VL or base count thresholds need to be applied to detect and characterise sequences confidently. Alternatively, utilising a minimum threshold of 10 for calling consensus bases was useful for certain protocols to exclude background contamination (yet HBV was still detected in the negative control), and other protocols required additional genome coverage thresholds. The detection of contaminating sequences has been observed in other studies comparing NGS protocols[11,19], reinforcing this challenge for NGS sequencing in clinical diagnostic settings. For detecting and quantifying samples with very low VL, such as those infected with occult HBV[14], conventional real-time or nested PCR would still be necessary. As with the detection of viruses around the limit of detection of conventional PCR assays, the stochastic effect around the limits of detection of NGS protocols affects the detection of samples with low VL, albeit at a higher detection limit than conventional PCR.

The high diversity of base counts and error rates for ONT base-calling algorithms hinder the detection of minority variants. Although these errors are unlikely to impact interpretation for most clinical and public health purposes, we advise against performing variant calling directly from mapped ONT reads without considerable bioinformatics expertise to exclude false substitutions and improve variant calling[22]. However, the introduction of R10 ONT kits and improved base-calling capabilities after our ONT protocols were performed has since changed the ONT field and may lower the diversity of base counts found in our study[23,24].

NGS methods involving PCR amplification are inherently limited to the characterisation of one known target pathogen and the length of that pathogen's genome; metagenomic

methods showed clear advantages. Co-detections found high prevalences of anelloviruses, human herpesviruses, and polyomaviruses in the samples. Whilst the clinical relevance of these viruses for the healthy population and blood donation is questionable, the viruses may potentially cause harm if transmitted to immunocompromised individuals.

The 100% (24/24) prevalence of anellovirus DNA detection by PCR was consistent with previous studies, which have shown that most immunocompetent individuals carry the virus in their blood and at high VLs in patients with HBV infection[25]. The low detection rate by all TAC protocols may have originated from the extreme genetic heterogeneity of the anellovirus species and genera that were not adequately represented among the probes used for target capture[26]. However, this does not explain the similarly low rates of detection by MTG methods which were not a sensitivity problem as the anelloviruses had high VLs by PCR; it is possible that their circular sequences were too divergent to map to what might be a very limited number of reference sequences used by the bioinformatic software to assign sequence reads. Our detection rates of polyomaviruses and HHVs were comparable to those reported in population studies[27,28]. The high detection rate of HHVs and polyomaviruses with TAC-C may be explained by DNAemia from disintegrating persistently infected cells[29], or by free fragmented DNA not detected by amplicon-specific real-time PCR.

Our study only included samples with lower VL; sequencing samples with higher VL would have increased the proportion of samples with high read depths and complete genome coverage. We did not include protocols using commercially available HBV-exclusive capture panels. We did not investigate host transcriptomic information, which could be combined with pathogen metagenomics to improve accuracy[30]. Although we centralised much of the bioinformatics to facilitate comparison, the initial pipeline-specific steps not centrally

performed could have affected sensitivity, such as the choice of taxonomic classifier[10]. In a follow-up study, we plan to evaluate the performance of different bioinformatic pipelines to ascertain optimal tools for HBV characterisation. It is noted that pipelines have specific criteria for calling majority consensus bases and for detection (Table S3), outside of the criteria already used for read quality in this study.

In conclusion, this study provides a comprehensive comparison of four NGS sequencing methodologies for detecting and generating HBV genomes in low VL samples. HBV-specific PCR-amplification methods using ONT platforms were the most cost-efficient in detecting the most HBV-positive samples with the highest genome coverage. However, experience, laboratory skills, and precautionary steps are necessary to minimise contamination in complex NGS workflows[31], such as using separate PCR tubes, exercising extreme care when pipetting and handling tubes, and including appropriate negative plasma controls. Together, these protocols could facilitate the sensitive and rapid generation of global HBV whole genome sequence data. Such advancements in viral pathogen diagnostics are vital as we move toward a future where NGS is becoming more widely adopted.

References

1 Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics 2016 17:6* 2016; **17**: 333–51.

2 Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet* 2019; **20**: 341–55.

3 Deng X, Achari A, Federman S, *et al.* Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat Microbiol* 2020; **5**: 443–54.

4 Thorburn F, Bennett S, Modha S, Murdoch D, Gunson R, Murcia PR. The use of next generation sequencing in the diagnosis and typing of respiratory infections. *Journal of Clinical Virology* 2015; **69**: 96.

5 Matthews PC, Downs LO, Vawda S, *et al.* Bimodal distribution and set point HBV DNA viral loads in chronic infection: Retrospective analysis of cohorts from the UK and South Africa. *Wellcome Open Res* 2020; **5**. DOI:10.12688/WELLCOMEOPENRES.15941.2.

6 Lumley SF, Kent C, Jennings D, *et al.* Whole genome sequencing of hepatitis B virus using tiled amplicon (HEPTILE) and probe based enrichment on Illumina and Nanopore platforms. *Sci Rep* 2025; **15**: 5795.

7 World Health Organization. World health statistics 2024: monitoring health for the SDGs, Sustainable Development Goals. Geneva, 2024.

8 McNaughton AL, D'Arienzo V, Ansari MA, *et al.* Insights From Deep Sequencing of the HBV Genome—Unique, Tiny, and Misunderstood. *Gastroenterology* 2019; **156**: 384–99.

9 Rajoriya N, Combet C, Zoulim F, Janssen HLA. How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? *J Hepatol* 2017; **67**: 1281–97.

10 Buddle S, Forrest L, Akinsuyi N, *et al.* Evaluating metagenomics and targeted approaches for diagnosis and surveillance of viruses. *Genome Medicine 2024 16:1* 2024; **16**: 1–20.

11 Lopez-Labrador FX, Huber M, Sidorov IA, *et al.* Multicenter benchmarking of short and long read wet lab protocols for clinical viral metagenomics. *Journal of Clinical Virology* 2024; **173**. DOI:10.1016/J.JCV.2024.105695,.

12 Thomson E, Ip CLC, Badhan A, *et al.* Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes. *J Clin Microbiol* 2016; **54**: 2470.

13 Stenbäck JB, Schmidt D, Noborg U, *et al.* Accurate and cost-efficient whole genome sequencing of hepatitis B virus using Nanopore. *medRxiv* 2024; : 2024.08.12.24311345.

14 Fu MX, Simmonds P, Andreani J, *et al.* Ultrasensitive PCR system for HBV DNA detection: Risk stratification for occult hepatitis B virus infection in English blood donors. *J Med Virol* 2023; **95**. DOI:10.1002/JMV.29144.

15 Ono Y, Onda H, Sasada R, Igarashi K, Sugino Y, Nishioka K. The complete nucleotide sequences of the cloned hepatitis B virus DNA; subtype adr and adw. *Nucleic Acids Res* 1983; **11**: 1747.

16 Simmonds P. SSE: A nucleotide and amino acid sequence analysis platform. *BMC Res Notes* 2012; **5**. DOI:10.1186/1756-0500-5-50,.

17 Fu MX, Larralde O, Mayne R, *et al.* Use of polyethylene glycol precipitation and ultracentrifugation to enhance the sensitivity of hepatitis B virus DNA detection. *Journal of Clinical Virology* 2025; : 105802.

18 Kapoor V, Briese T, Ranjan A, *et al.* Validation of the VirCapSeq-VERT system for differential diagnosis, detection, and surveillance of viral infections. *J Clin Microbiol* 2023; **62**: e00612-23.
19 Mourik K, Sidorov I, Carbo EC, *et al.* Comparison of the performance of two targeted metagenomic virus capture probe-based methods using reference control materials and clinical samples. *J Clin Microbiol* 2024; **62**. DOI:10.1128/JCM.00345-24,.
20 Lewandowska DW, Zagordi O, Geissberger FD, *et al.* Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. *Microbiome* 2017; **5**: 94.
21 Lemoine M, Eholié S, Lacombe K. Reducing the neglected burden of viral hepatitis in Africa: Strategies for a global approach. *J Hepatol* 2015; **62**: 469–76.
22 Ye F, Zhu J, Zhang X, *et al.* Characteristics and filtering of low-frequency artificial short deletion variations based on nanopore sequencing. *Gigascience* 2025; **14**: 1–15.
23 Zhang T, Li H, Ma S, *et al.* The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl Environ Microbiol* 2023; **89**. DOI:10.1128/AEM.00605-23/SUPPL_FILE/AEM.00605-23-S0008.EPS.
24 Liu-Wei W, van der Toorn W, Bohn P, Hölzer M, Smyth RP, von Kleist M. Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. *BMC Genomics* 2024; **25**: 1–15.
25 Spandole S, Cimponeriu D, Berca LM, Mihăescu G. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch Virol* 2015; **160**: 893–908.
26 Pyöriä L, Pratas D, Toppinen M, Hedman K, Sajantila A, Perdomo MF. Unmasking the tissue-resident eukaryotic DNA virome in humans. *Nucleic Acids Res* 2023; **51**: 3223–39.
27 Kamminga S, van der Meijden E, de Brouwer C, Feltkamp M, Zaaijer H. Prevalence of DNA of fourteen human polyomaviruses determined in blood donors. *Transfusion (Paris)* 2019; **59**: 3689–97.
28 Edelman DC. Human herpesvirus 8 - A novel human pathogen. *Virol J* 2005; **2**: 1–32.
29 Delbue S, Tremolada S, Elia F, *et al.* Lymphotropic Polyomavirus is detected in peripheral blood from immunocompromised and healthy subjects. *Journal of Clinical Virology* 2010; **47**: 156–60.
30 Kalantar KL, Neyton L, Abdelghany M, *et al.* Integrated host-microbe plasma metagenomics for sepsis diagnosis in a prospective cohort of critically ill adults. *Nat Microbiol* 2022; **7**: 1805–16.
31 López-Labrador FX, Brown JR, Fischer N, *et al.* Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure. *Journal of Clinical Virology* 2021; **134**: 104691.

Data Availability Statement: The data for this study are included in the manuscript. Within-sample consensus full-length genome sequences have been submitted to GenBank, accession numbers PV786119 to PV786132.

Figure 1. Relationship between viral loads and HBV-specific bases read for each group of methods on a common log scale, when majority consensus bases were called at a threshold of 1 base per nucleotide site. Pink indicates contaminating sequences, including where HBV sequences were obtained for the negative HBV10 sample on the y-axis. Samples not detected are found on the x-axis. Spearman's correlation test deduced the significance of the association between viral loads and base counts (R^2 value and *p*-values listed in boxes). HBV: hepatitis B virus; MTG: untargeted metagenomics; TAC: target capture; ONT: Oxford Nanopore Technologies.

Figure 2. Relationship between viral loads and breadth coverage of HBV assembled sequences for each group of methods on a common axes scale, when majority consensus bases were called at a threshold of 1 base per nucleotide site. Pink indicates contaminating sequences, including where HBV sequences were obtained for the negative HBV10 sample on the y-axis. Samples not detected are found on the x-axis. Spearman's correlation test deduced the significance of the association between viral loads and genome coverage (R^2 value and *p*-values listed in boxes). HBV: hepatitis B virus; MTG: untargeted metagenomics; TAC: target capture; ONT: Oxford Nanopore Technologies.

Figure 3. Normalised mean read depth of the HBV genome for each of the sequencing methods. The normalised mean read depth was calculated as the number of bases at each genome position as a proportion of the total reads of the sequence, then multiplied by the total number of sites, where the expected mean value was one base per site. The eleven highest viral load samples were included when detected, and only true-positive sequences with more than 1000 total HBV-specific bases read were included; the number of samples from each protocol that were included is indicated in each plot’s key. Genome positions

were based on the D00330 reference sequence. A genome diagram of HBV, drawn to the x-axis scale, shows gene positions in shaded boxes and regulatory regions in unshaded boxes. TAC: target capture; MTG: untargeted metagenomics.

Figure 4. Relationship between viral loads and mean Shannon entropy values for polymorphic sites for each group of methods on a common axes scale, when majority consensus bases were called at a threshold of 1 base per nucleotide site. Pink indicates contaminating sequences, including where HBV sequences were obtained for the negative HBV10 sample on the y-axis. Samples not detected are found on the x-axis. Spearman's correlation test deduced the significance of the association between viral loads and entropy values (R^2 value and *p*-values listed in boxes). HBV: hepatitis B virus; MTG: untargeted metagenomics; TAC: target capture; ONT: Oxford Nanopore Technologies.

Figure 5. Comparison of a) time taken, and b) cost based on the total time and costs for a protocol-specific representative number of samples included in a typical run. For Figure 5a, undotted bars to the left of the plot show per-person labour time per sample, whereas dotted bars to the right show the waiting time per sample, which may include machine running times. For clinical purposes, PCR-1 is normally run overnight on the MinIon flow cell (average of 16 hours), but was run for 72 hours for this study. Costs were standardised using a 1:0.8959:0.7531 USD:EUR:GBP exchange rate, as of May 18, 2025, and total costs were rounded to the nearest ten. USD: United States Dollar; EUR: Euro; GBP: Great British Pound; QC: quality control; TAC: target capture; MTG: untargeted metagenomics.

Table 1. Protocol details of the NGS protocols analysed, including the sequencing workflows and analysis pipelines (until the point of genome assembly) for each method. Effective test volume was calculated as [library prep input volume]/[extraction output volume]*[extraction input volume], where TAC-A, TAC-B and MTG-B also accounted for reductions in effective test volume through cDNA synthesis, and PCR volumes were also accounted for in PCR-based protocols. Numerical superscripts next to method identifiers indicate where the same laboratory performed the protocols.

	MTG-A	**MTG-B[1]**	**TAC-A[1]**	**TAC-B[1]**	**TAC-C**	**PCR-A**	**PCR-B[2]**	**PCR-1**	**PCR-2[2]**
Technology	Illumina	Illumina	Illumina	Illumina	Illumina	Illumina	Illumina	ONT	ONT
Pre-extraction concentration	No	PEG Virus Precipitation	No	PEG Virus Precipitation	No	No	No	No	No
Nucleic acid extraction	QIA-symphony DSP Virus/Pathogen Mini Kit	Roche Large Volume Kit	Roche Large Volume Kit	Roche Large Volume Kit	QIAamp DNA Mini Kit	Roche Large Volume Kit*	Roche Large Volume Kit*	Roche Large Volume Kit	Roche Large Volume Kit*
Input/output volume	200 µL / 60 µL	20,000 µL / 50 µL	5,000 µL / 50 µL	20,000 µL / 50 µL	400 µL / 60 µL	5,000 µL / 50 µL	5,000 µL / 50 µL	5,000 µL / 50 µL	5,000 µL / 50 µL
Pre-prep PCR amplification (volume of extract added, total reaction volume)	No	No	No	No	No	Yes Round 1: 5 µL, 25 µL Round 2 (x4): 1 µL, 25 µL	Yes 2x (7 µL, 25 µL), followed by clean-up into 20 µL	Yes 2x (5 µL, 25 µL)	Yes, 2x (7 µL, 25 µL), followed by clean-up into 20 µL
Library prep input volume	26 µL	5 µL	5 µL	5 µL	50 µL	5 µL (0.2 ng/µL)	8.3 µL	10 µL	8.3 µL
Effective test volume	87 µL	800 µL	350 µL	800 µL	333 µL	~4 µL	581 µL	200 µL	581 µL
rRNA depletion	No	No	No	No	No	No	No	No	No
Human DNA depletion	No	No	No	No	No	No	No	No	No
cDNA synthesis	No	Yes	Yes	Yes	No	No	No	No	No
Library prep kit	NEBNext Ultra II FS (DNA)	Twist (cDNA)	Twist (half DNA & half cDNA)	Twist (cDNA)	Roche Hyperprep (DNA)	Nextera XT (DNA)	NEBNext Ultra II FS (DNA)	SQK-RBK114.96 (DNA)	SQK-NBD114.96(DNA)

Probe capture	No	No	Yes	Yes	Yes	No	No	No	No
Random amplification cycles	N/A	16	16	16	30 (2x 15)	12	0	30	0
Average gigabases (Gb) or megabases (Mb) sequenced per sample	5.03 Gb	208.2 Mb	49.6 Mb	105.1 Mb	2.5 Gb	317.3 Mb	47.4 Mb	2.2 Mb	312.7 Mb
Platform	NextSeq 2000	NovaSeq X	NovaSeq X	NovaSeq X	NovaSeq X	MiSeq V2	NovaSeq X Plus	MinION	MinION
Sequencing strategy	Paired end 150	Paired end 150	Paired end 150	Paired end 150	Paired end 150	Paired end 250	Paired end 150	N50: 1.06 kb	N50: 600 bp
Read filtering	nf-core taxprofiler (MetaMix was used for detecting additional viruses)	Kraken2, FastQC, Trimmomatic, BWA-mem2, custom script for aggregating reads by reference organism, samtools, viralconsensus	Castanet	Castanet	TracesPipe	FASTQC, Trimmomatic, Smalt	QUASR, CutAdapt, Skewer, Bowtie2	See supplementary material.	dorado, hbv-fieldbioinformatics
Assembly method	Reference-based mapping	Reference-based mapping	Reference-based mapping	Reference-based mapping	Bowtie 2 and metaSPAdes	IVA, de novo using custom scripts	Reference guided to closest reference	Minimap2	Reference guided to closest reference

*These extractions were performed separately in the central coordinating centre, and eluates (or spin columns before the elution step) were sent to the laboratories instead of plasma.

Table 2. Overview of the detection of HBV in sample sets, when majority consensus bases were called at a threshold of 1 base per nucleotide site. The percentage genome completeness is given in each cell. The left-hand side of the table with blue headers shows the sequences obtained from the NGS methods, whilst the right-hand side (calamine colour) shows the sequences obtained by Sanger sequencing (length of S amplicon = 395 nucleotides; SP = 973 nucleotides; XC = 698 nucleotides). Within-sample consensus sequences were assembled for each sample from the individual sequences of the NGS methods that had around 100% genome coverage and were genetically similar, which were assigned to HBV genotypes. The percentage genetic relatedness of the sequences from each method was compared to the within-sample consensus sequence, where assembled sequences similar to the within-sample consensus were shaded on a grayscale, and different strains (differing in >5% nucleotide sequence from the within-sample consensus) were shaded in a pink scale. The absence of sequences was shaded in black. Where within-sample consensus sequences could not be assembled (i.e. no two complete sequences were similar), genotypes were defined from Sanger sequences, and the NGS sequences were instead compared to one another. Samples are ranked from highest to lowest geometric mean viral load. For TAC-B: library preparation failed for sample HBV13; samples HBV4, HBV9, and HBV19 had reads of poor quality for the original pipeline to process, so the mean base quality was dropped below 1.0. #: For PCR-1, a negative water control was tested instead of the negative plasma control. Alphabetical superscripts (a, b, c) represent identical sequences. Numerical superscripts next to method identifiers indicate where the same laboratory performed the protocols.

		PCR (Nanopore)		PCR (Illumina)		Target capture			Metagenomics				Sanger sequencing		
Sample	**Viral load (IU/mL)**	**PCR-1**	**PCR-2[1]**	**PCR-A**	**PCR-B[1]**	**TAC-A[2]**	**TAC-B[2]**	**TAC-C**	**MTG-A**	**MTG-B[2]**	**Within-sample consensus**	**Genotype assigned**	**S**	**SP**	**XC**
HBV15	**6207.0**	100%	100%	91%	100%	11%	100%	90%	25%	72%	**Y**	**A2**	12%	30%	22%
HBV8	**3154.0**	100%	100%	98%	100%	91%[c]	97%	99%	50%	14%	**Y**	**D1**	12%		
HBV2	**3108.0**	100%	91%		100%	90%[c]	81%	100%		25%	**Y**	**A2**	12%		22%
HBV7	**1058.0**	100%	100%	98%	99%	17%	83%	93%	11%	25%	**Y**	**C1**	12%	29%	22%
HBV19	**859.8**	100%	49%		100%	22%	4%	88%	10%		**Y**	**D3**	12%	31%	22%
HBV16	**599.2**	100%	85%	59%	100%	9%		75%			**Y**	**A1**	12%		
HBV1	**551.6**	100%	86%	76%	100%	6%	38%	88%	4%		**Y**	**D2**	12%		
HBV14	**522.3**	100%	100%	92%	100%	9%	95%	80%		49%	**Y**	**E**	12%	30%	
HBV23	**337.1**	100%	86%	99%	100%	64%	71%	93%		16%	**Y**	**D4**	12%	31%	22%
HBV17	**299.9**	100%	100%	96%	92%		14%	50%			**Y**	**E**	12%	30%	
HBV9	**138.1**	100%	100%	33%	99%	9%	8%	35%	2%		**Y**	**A1**	12%	30%	
HBV5	**128.4**	100%	35%		89%			66%			**N**	**E**	12%	30%	
HBV22	**73.6**	100%	83%		86%			54%			**Y**	**D2**	12%	31%	
HBV4	**60.5**	100%	85%	59%	100%	35%	7%	100%	12%		**Y**	**B4**	12%	30%	22%
HBV13	**60.4**	100%	85%		98%			91%	5%		**Y**	**D1**	12%	31%	22%
HBV21	**41.0**	100%	68%[a]		74%		7%	43%			**N**	**A**	12%	30%	
HBV6	**11.3**	100%	33%		90%		3%	28%		3%	**N**	**D3**	12%	31%	
HBV18	**8.5**	60%	67%		88%						**N**	**A**	12%		
HBV20	**3.3**	100%	82%[a]		81%			26%	<1%		**N**	**D**	12%		
HBV12	**1.8**	15%	67%		97%			25%			**N**	**A**	12%		
HBV24	**1.1**	99%	33%[b]		88%						**N**	**E**	12%		
HBV11	**0.3**	100%	68%		85%						**N**	**D**	12%		
HBV3	**0.2**	71%	68%		91%			5%			**N**	**A3**	5%		
Negative (HBV10)	**-**	100%[#]	67%[b]	48%	88%						**N**	**-**			

Within-sample consensus	Shading key					
Y (% sequence identity to within-sample consensus)						
	No sequence	≥85%	≥90%	≥95%	≥98%	≥99.5%
N (similarity of sequences)			
	No sequence	Different sequence to others	Similar sequence

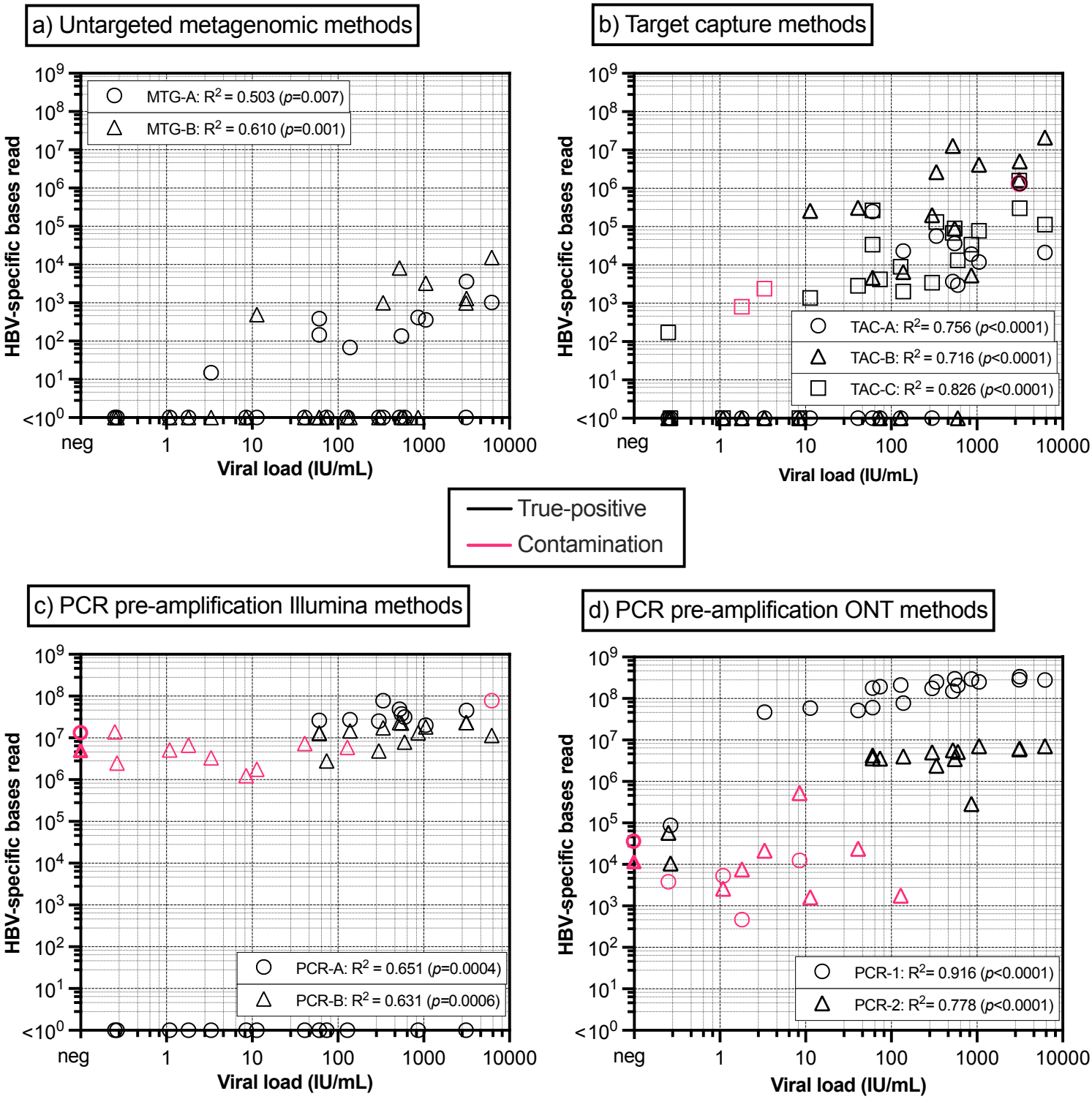
Table 3. Detection of other viruses by the targeted capture and untargeted metagenomics methods, showing deduplicated read counts (genome coverage %). Results reported for TAC-C are for viruses where at least 10 reads were obtained. More than one TTV species may have been identified for MTG-A, but the most abundant was included. The furthest right column shows the PCR results, where PCR was performed for all samples when two or more methods detected a virus in the same sample. These PCRs were specific for alpha, beta and gamma-torquevirus, Merkel cell polyomavirus, and herpesviruses. '-' denotes where the method did not detect the virus, '=' denotes where the PCR was not performed. A numerical superscript indicates where the same laboratory performed the protocols. HHV: human herpesvirus; MCPyV: Merkel cell polyomavirus; AV: anellovirus; TTV: torquetenovirus; EBV: Epstein-Barr virus; KSHV: Kaposi's sarcoma-associated herpesvirus; JCPyV: John Cunningham virus.

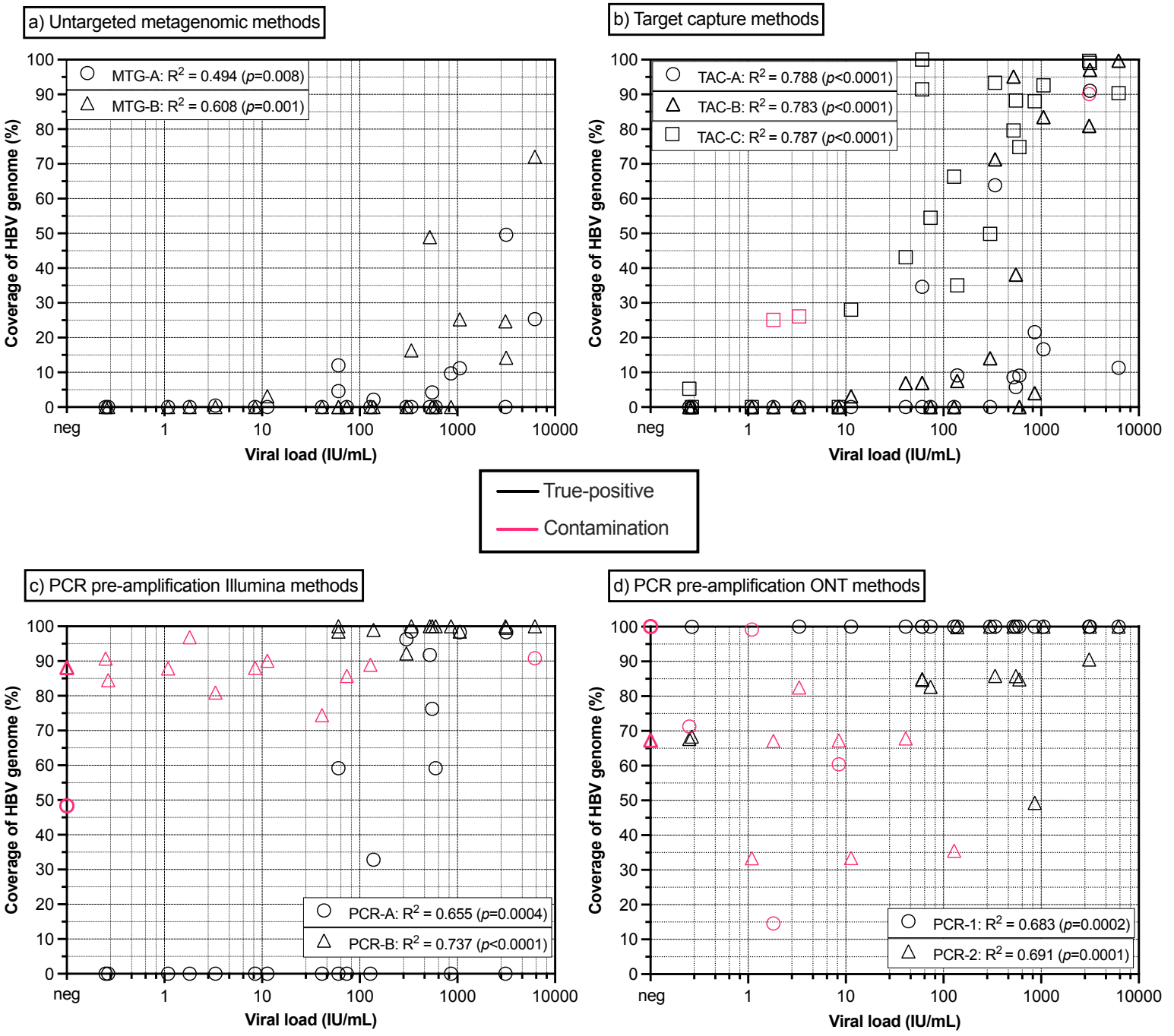
Sample	Virus	MTG-A	MTG-B[1]	TAC-A[1]	TAC-B[1]	TAC-C	PCR
HBV1	AV	43 (54%)	-	6 (14%)	3 (3%)	37 (22%)	**Ct 13.1 (αTTV)**
	HHV7	-	-	-	-	885 (10%)	**Negative**
HBV2	MCPyV	-	-	-	-	14 (20%)	**Negative**
	AV	-	-	-	-	35 (21%)	**Ct 9.6 (αTTV), Ct 8.0 (βTTV), Ct 21.1 (γTTV)**
	HHV7	-	-	-	-	649 (10%)	**Negative**
HBV3	AV	9 (20%)	2 (1%)	7 (20%)	19 (12%)	167 (49%)	**Ct 6.3 (αTTV), Ct 5.8 (βTTV), Ct 8.5 (γTTV)**
	EBV	-	-	-	-	321 (6%)	**Negative**
	HHV7	-	-	-	-	1699 (9%)	**Negative**
HBV4	AV	26 (49%)	-	11 (20%)	14 (16%)	117 (45%)	**Ct 10.9 (αTTV), Ct 4.7 (βTTV), Ct 13.0 (γTTV)**
	Simian adenovirus 25	7 (0.74%)	-	-	-	-	**=**

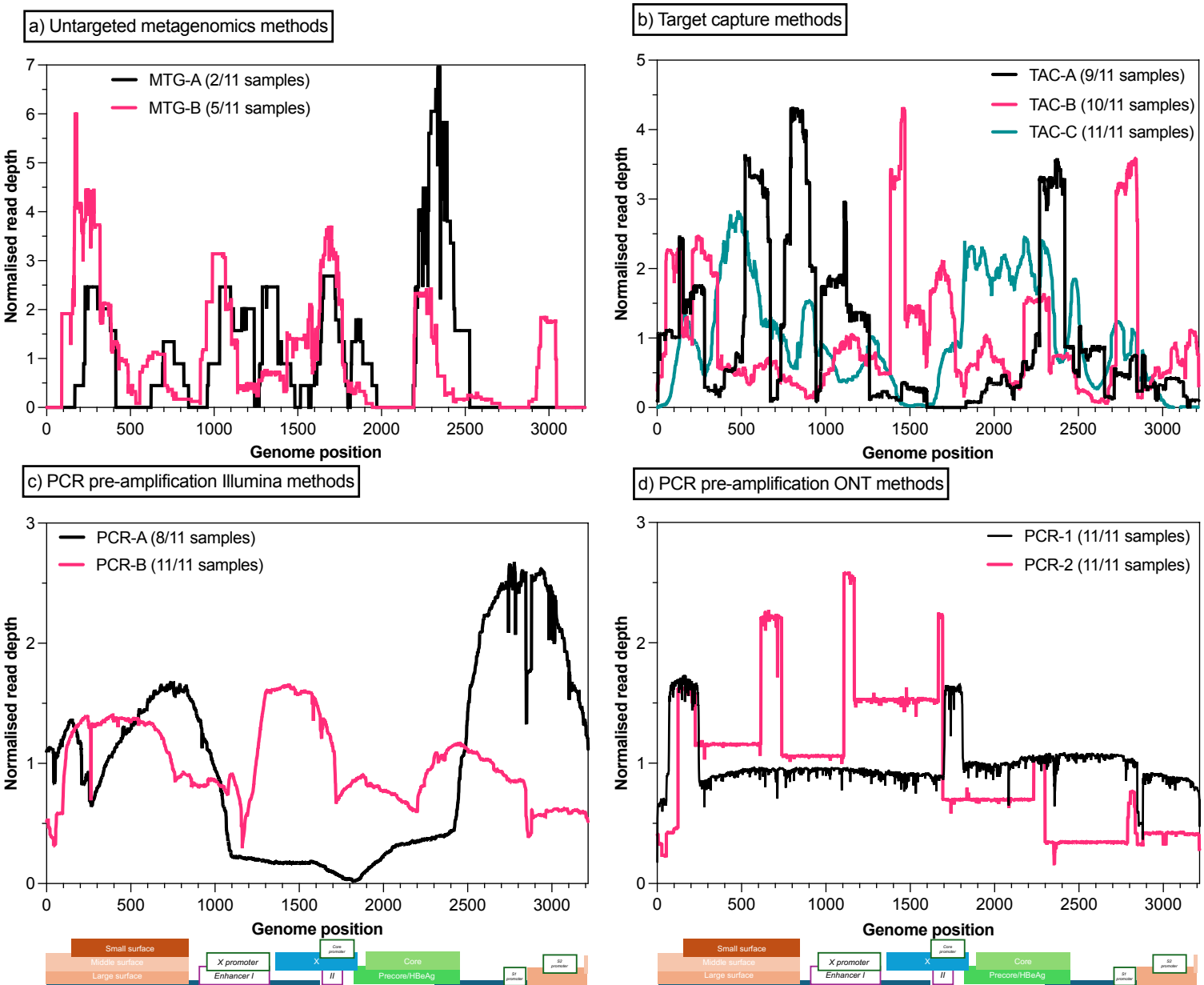
HBV5	AV	-	5 (8%)	-	20 (16%)	15 (15%)	**Ct 22.2 (αTTV), Ct 11.0 (βTTV)**
	JCPyV	-	-	-	-	63 (52%)	**=**
HBV6	AV	6 (21%)	44 (40%)	7 (12%)	310 (75%)	84 (36%)	**Ct 7.2 (βTTV)**
	MCPyV	-	-	-	-	126 (65%)	**152 copies/mL**
	HPyV6	-	-	-	4 (4%)	26 (30%)	**Negative**
	HHV7	-	-	-	-	565 (10%)	**Negative**
	JCPyV	-	-	-	-	70 (56%)	**=**
HBV7	AV	-	2 (2%)	13 (6%)	11 (24%)	14 (17%)	**Ct 7.2 (γTTV)**
	HPV-2	-	-	-	-	10 (8%)	**=**
	HHV7	-	-	-	-	1112 (9%)	**Negative**
	JCPyV	-	-	-	-	34 (41%)	**=**
HBV8	EBV	12 (0.2%)	-	42 (14%)	3 (2%)	2020 (39%)	**Ct 35.9**
	AV	-	1 (1%)	1 (5%)	-	-	**Ct 12.5 (αTTV), Ct 8.3 (βTTV), Ct 21.4 (γTTV)**
	MCPyV	-	-	-	-	35 (31%)	**Negative**
HBV9	AV	-	-	3 (4%)	1 (1%)	-	**Ct 15.0 (αTTV), Ct 8.9 (βTTV), Ct 19.2 (γTTV)**
	HHV6	-	-	-	1 (1%)	1626 (15%)	**Ct 38.2**
HBV10	AV	-	3 (1%)	3 (4%)	1 (9%)	-	**Ct 12.6 (αTTV), Ct 10.7 (βTTV)**
	HHV7	-	-	-	-	730 (7%)	**Negative**
HBV11	AV	-	-	1 (6%)	2 (13%)	22 (10%)	**Ct 6.1 (αTTV), Ct 5.9 (βTTV), Ct 22.5 (γTTV)**
	EBV	-	-	-	-	774 (7%)	**Negative**
	HHV7	-	-	-	-	805 (7%)	**Negative**
HBV12	KSHV	18 (0.63%)	-	1 (5%)	-	7353 (93%)	**Ct 36.6**
	EBV	-	-	-	1 (1%)	-	**Negative**
	HHV7	-	-	-	-	1324 (9%)	**Negative**
	AV	-	2 (1%)	-	-	-	**Ct 26.1 (αTTV), Ct 13.3 (βTTV)**
HBV13	MCPyV	-	-	-	-	16 (15%)	**Negative**
	TTV	-	3 (2%)	-	-	20 (24%)	**Ct 11.9 (αTTV), Ct 5.1 (βTTV), Ct 13.0 (γTTV)**

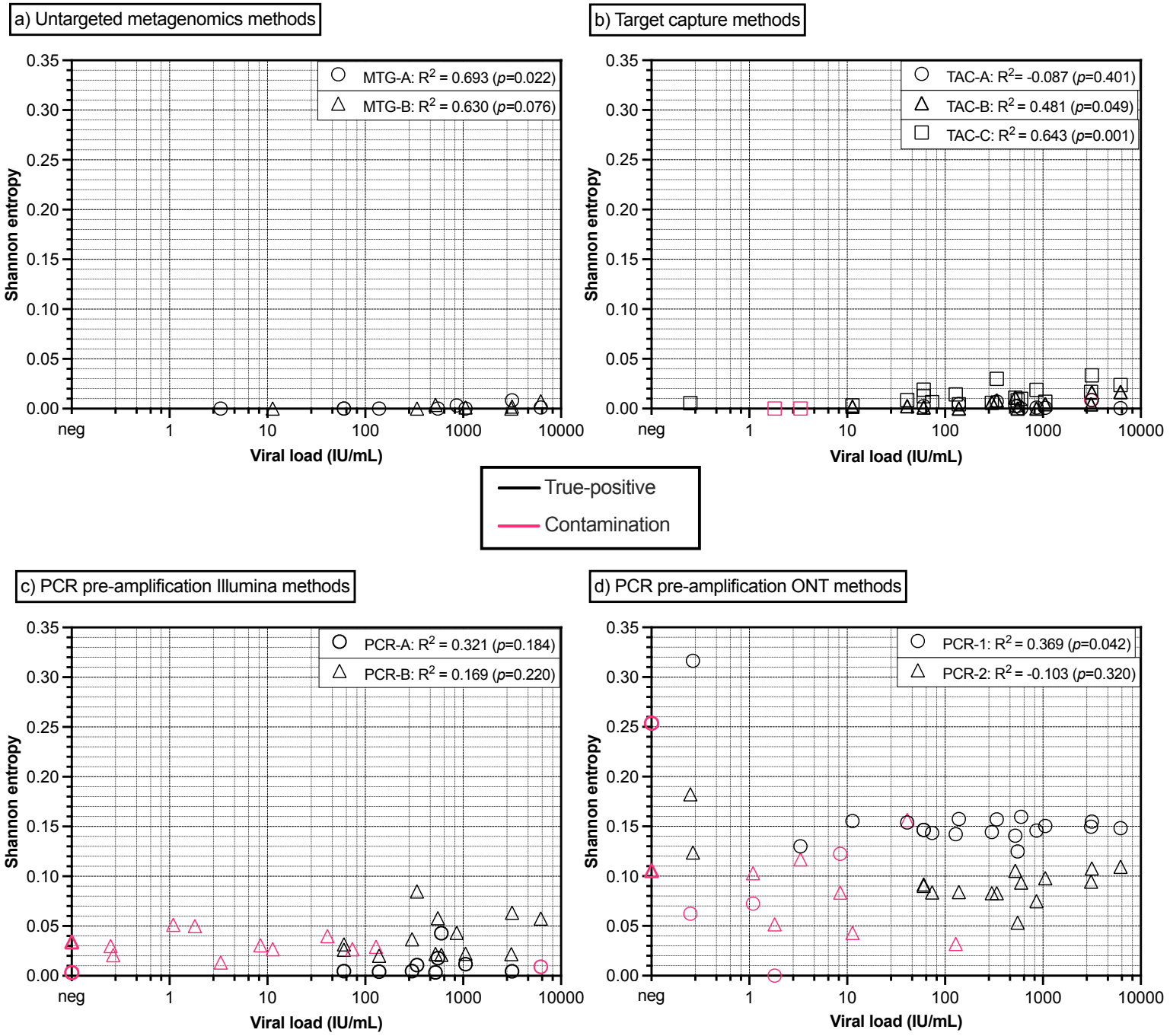
	HHV7	-	-	-	-	825 (11%)	**Negative**
HBV14	AV	-	8 (11%)	-	88 (33%)	-	**Ct 7.0 (αTTV), Ct 10.1 (βTTV)**
	JCPyV	-	-	-	-	10 (10%)	**=**
HBV15	JCPyV	-	-	-	-	69 (59%)	**=**
	AV	7 (9.6%)	7 (1%)	-	16 (35%)	11 (8%)	**Ct 12.4 (αTTV), Ct 11.4 (βTTV), Ct 15.2 (γTTV)**
	EBV	-	-	-	1 (1%)	-	**Negative**
HBV16	AV	-	1 (1%)	-	1 (12%)	-	**Ct 13.8 (αTTV), Ct 9.1 (βTTV)**
HBV17	HHV7	-	-	-	-	413 (9%)	**Negative**
	AV	-	-	-	-	-	**Ct 13.3 (βTTV), Ct 15.3 (γTTV)**
HBV18	AV	-	-	1 (3%)	1 (1%)	-	**Ct 12.9 (αTTV), Ct 11.4 (βTTV), Ct 21.0 (γTTV)**
HBV19	AV	-	-	-	-	-	**Ct 19.1 (αTTV), Ct 16.2 (βTTV), Ct 9.5 (γTTV)**
HBV20	AV	27 (41%)	-	8 (13%)	2 (8%)	25 (18%)	**Ct 12.2 (αTTV), Ct 11.5 (βTTV), Ct 6.9 (γTTV)**
HBV21	TTV	-	-	1 (4%)	3 (9%)	-	**Ct 11.9 (βTTV), Ct 9.4 (γTTV)**
	HPV	-	-	-	-	17 (15%)	**=**
	HHV7	-	-	-	-	833 (9%)	**Negative**
	MCPyV	-	-	-	-	-	**219 copies/mL**
HBV22	AV	-	-	-	8 (11%)	-	**Ct 11.4 (αTTV), Ct 12.1 (βTTV), Ct 21.4 (γTTV)**
	HHV6	-	-	-	-	1070 (7%)	**Negative**
HBV23	AV	-	2 (1%)	11 (35%)	22 (26%)	56 (30%)	**Ct 5.8 (αTTV), Ct 6.4 (βTTV), Ct 19.8 (γTTV)**
	EBV	-	-	-	-	212 (6%)	**Negative**
	HHV6	-	-	-	-	1535 (7%)	**Negative**
	HHV7	-	-	-	-	761 (5%)	**Negative**
	MCPyV	-	-	-	-	10 (8%)	**Negative**

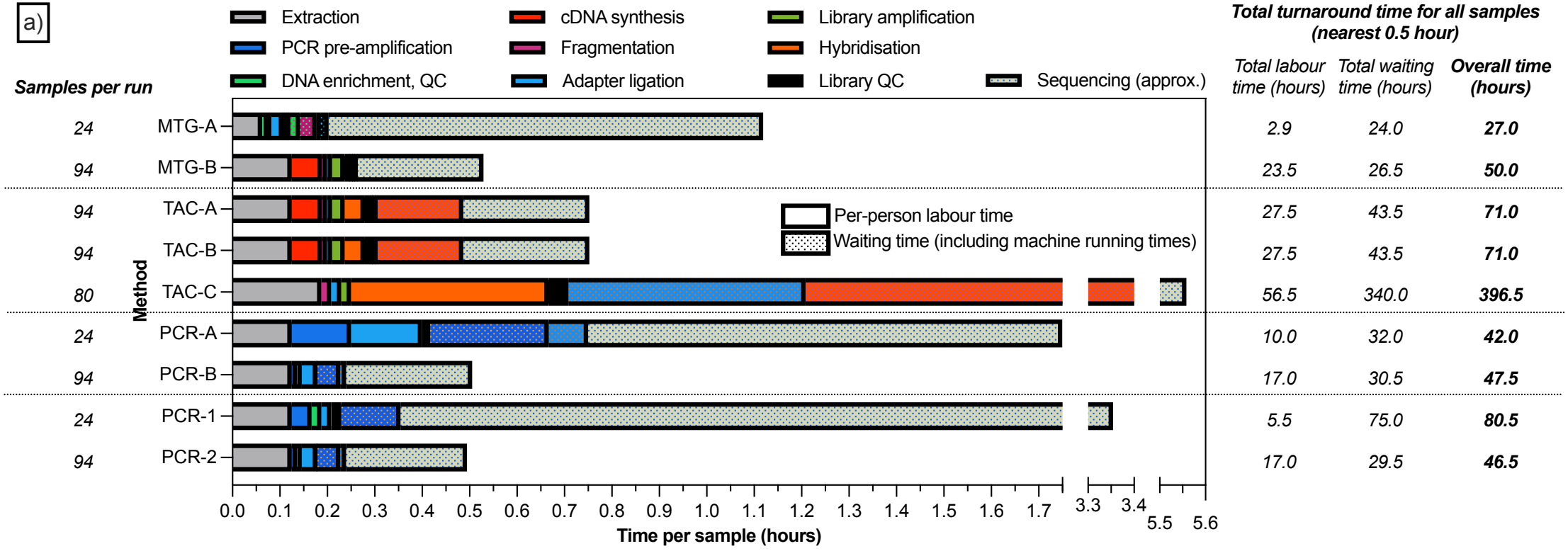
HBV24	AV	-	2 (1%)	22 (28%)	4 (6%)	31 (21%)	**Ct 7.8 (αTTV), Ct 7.0 (βTTV), Ct 13.5 (γTTV)**
	HHV6	-	-	-	-	616 (11%)	**Negative**
	HPyV7	-	-	-	-	13 (12%)	**=**
	HPyV6	-	-	-	-	27 (24%)	**Negative**
	KSHV	-	-	-	-	226 (5%)	**Negative**
	HHV7	-	-	-	-	326 (5%)	**Negative**

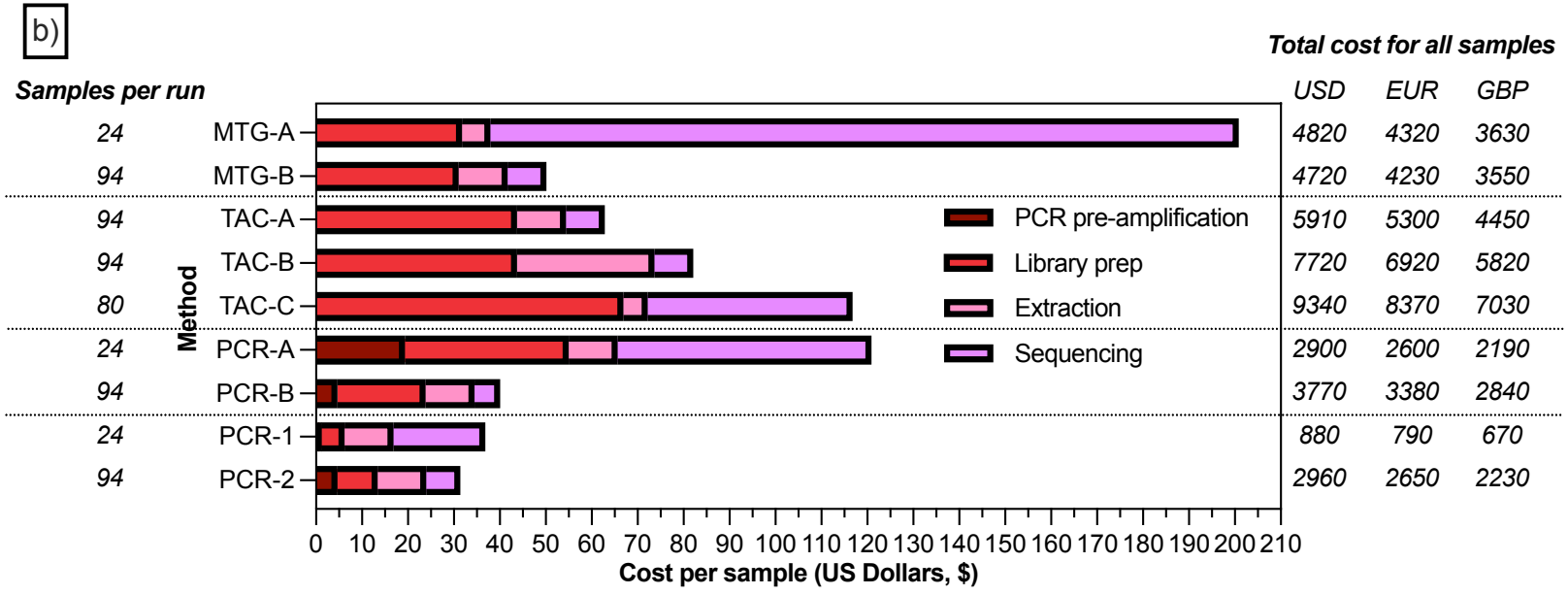

a) Untargeted metagenomic methods
MTG-A: R2 = 0.503 (p=0.007)
MTG-B: R2 = 0.610 (p=0.001)
b) Target capture methods
TAC-A: R2= 0.756 (p<0.0001)
TAC-B: R2 = 0.716 (p<0.0001)
TAC-C: R2 = 0.826 (p<0.0001)
True-positive
Contamination
c) PCR pre-amplification Illumina methods
PCR-A: R2 = 0.651 (p=0.0004)
PCR-B: R2 = 0.631 (p=0.0006)
d) PCR pre-amplification ONT methods
PCR-1: R2= 0.916 (p<0.0001)
PCR-2: R2= 0.778 (p<0.0001)
HBV-specific bases read
Viral load (IU/mL)
neg
1
10
100
1000
10000
<10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9


a) Untargeted metagenomic methods
MTG-A: R2 = 0.494 (p=0.008)
MTG-B: R2 = 0.608 (p=0.001)
b) Target capture methods
TAC-A: R2 = 0.788 (p<0.0001)
TAC-B: R2 = 0.783 (p<0.0001)
TAC-C: R2 = 0.787 (p<0.0001)
True-positive
Contamination
c) PCR pre-amplification Illumina methods
PCR-A: R2 = 0.655 (p=0.0004)
PCR-B: R2 = 0.737 (p<0.0001)
d) PCR pre-amplification ONT methods
PCR-1: R2 = 0.683 (p=0.0002)
PCR-2: R2 = 0.691 (p=0.0001)
Coverage of HBV genome (%)
Viral load (IU/mL)
neg
1
10
100
1000
10000


a) Untargeted metagenomics methods
MTG-A (2/11 samples)
MTG-B (5/11 samples)
Normalised read depth
Genome position
b) Target capture methods
TAC-A (9/11 samples)
TAC-B (10/11 samples)
TAC-C (11/11 samples)
Normalised read depth
Genome position
c) PCR pre-amplification Illumina methods
PCR-A (8/11 samples)
PCR-B (11/11 samples)
Normalised read depth
Genome position
d) PCR pre-amplification ONT methods
PCR-1 (11/11 samples)
PCR-2 (11/11 samples)
Normalised read depth
Genome position
Small surface
Middle surface
Large surface
Polymerase
X promoter
Enhancer I
X
Core promoter
II
Core
Precore/HBeAg
S1 promoter
S2 promoter
Polymerase


a) Untargeted metagenomics methods
MTG-A: R2 = 0.693 (p=0.022)
MTG-B: R2 = 0.630 (p=0.076)
b) Target capture methods
TAC-A: R2= -0.087 (p=0.401)
TAC-B: R2 = 0.481 (p=0.049)
TAC-C: R2 = 0.643 (p=0.001)
True-positive
Contamination
c) PCR pre-amplification Illumina methods
PCR-A: R2 = 0.321 (p=0.184)
PCR-B: R2 = 0.169 (p=0.220)
d) PCR pre-amplification ONT methods
PCR-1: R2 = 0.369 (p=0.042)
PCR-2: R2 = -0.103 (p=0.320)
Shannon entropy
Viral load (IU/mL)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
neg
1
10
100
1000
10000


a)
Extraction
cDNA synthesis
Library amplification
PCR pre-amplification
Fragmentation
Hybridisation
DNA enrichment, QC
Adapter ligation
Library QC
Sequencing (approx.)
Samples per run
Total turnaround time for all samples (nearest 0.5 hour)
Total labour time (hours)
Total waiting time (hours)
Overall time (hours)
24 MTG-A 2.9 24.0 27.0
94 MTG-B 23.5 26.5 50.0
94 TAC-A 27.5 43.5 71.0
94 TAC-B 27.5 43.5 71.0
80 TAC-C 56.5 340.0 396.5
24 PCR-A 10.0 32.0 42.0
94 PCR-B 17.0 30.5 47.5
24 PCR-1 5.5 75.0 80.5
94 PCR-2 17.0 29.5 46.5
Method
Per-person labour time
Waiting time (including machine running times)
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 3.3 3.4 5.5 5.6
Time per sample (hours)


b)
Total cost for all samples
Samples per run
USD EUR GBP
24 MTG-A 4820 4320 3630
94 MTG-B 4720 4230 3550
94 TAC-A 5910 5300 4450
94 TAC-B 7720 6920 5820
80 TAC-C 9340 8370 7030
24 PCR-A 2900 2600 2190
94 PCR-B 3770 3380 2840
24 PCR-1 880 790 670
94 PCR-2 2960 2650 2230
Method
PCR pre-amplification
Library prep
Extraction
Sequencing
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
Cost per sample (US Dollars, $)