

1 **Whole genome sequencing of hepatitis B virus (HBV) using tiled amplicon**
2 **(HEP-TILE) and probe-based enrichment on Illumina and Nanopore**
3 **platforms.**

4
5 Sheila F Lumley^{1,2}, Chris Kent³, Daisy Jennings¹, Haiting Chai¹, George Airey¹, Elizabeth Waddilove⁴,
6 Marion Delphin⁴, Amy Trebes^{5,6}, Anna L McNaughton⁷, Khadija Said Mohammed⁴, Sam Wilkinson³,
7 Yanxia Wu⁵, George MacIntyre-Cockett⁵, Beatrice Kimono⁸, Moses Kwizera⁸, Kevin Ojambo⁸,
8 Tongai G Maponga⁹, Catherine de Lara¹, Jacqueline Martin², James Campbell⁴,
9 Marije Van Schalkwyk¹⁰, Dominique Goedhals^{11,12}, Robert Newton^{8,13}, Eleanor Barnes¹,
10 Nicholas J Loman³, Paolo Piazza⁵, Joshua Quick³, M Azim Ansari*¹, Philippa C Matthews*^{4,14,15}

11
12 *Joint senior and corresponding authors

13
14 **Affiliations**

- 15 1. Nuffield Department of Medicine, University of Oxford, Oxford, UK
16 2. Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS
17 Foundation Trust, John Radcliffe Hospital, Oxford, UK
18 3. Institute of Microbiology and Infection, University of Birmingham, UK
19 4. The Francis Crick Institute, London, UK
20 5. Centre for Human Genomics, Nuffield Department of Medicine, University of Oxford, UK
21 6. Genewiz UK Ltd, Azenta Life Sciences, Stratton Court, Abingdon, UK
22 7. Population Health Sciences, Bristol Medical School, University of Bristol, UK
23 8. MRC/UVRI & LSHTM Uganda Research Unit, Entebbe, Uganda
24 9. Division of Medical Virology, Stellenbosch University / National Health Laboratory Service
25 Tygerberg, Cape Town, South Africa
26 10. Division of Infectious Diseases, Department of Medicine, Stellenbosch University, Cape Town,
27 South Africa
28 11. University of the Free State, Bloemfontein, South Africa
29 12. PathCare, Pretoria, South Africa
30 13. Department of Health Sciences, University of York, UK
31 14. Division of Infection and Immunity, University College London, London, UK
32 15. Department of Infection, University College London Hospitals, London, UK

33
34
35

36 **Abstract**

37

38 Hepatitis B virus (HBV) whole genome sequencing (WGS) is currently limited as the DNA viral
39 loads (VL) of many clinical samples are below the threshold required to generate full genomes
40 using current sequencing methods. We developed two pan-genotypic viral enrichment methods,
41 using probe-based capture and tiled amplicon PCR (HEP-TILE) for HBV WGS. We demonstrate
42 using mock samples that both enrichment methods are pan-genotypic (genotypes A-J). Using
43 clinical samples, we demonstrate that HEP-TILE amplification successfully amplifies full genomes
44 at the lowest HBV VL tested (30 IU/ml), and the PCR products can be sequenced using both
45 Nanopore and Illumina platforms. Probe-based capture with Illumina sequencing required VL
46 >300,000 IU/ml to generate full length HBV genomes. The capture-Illumina and HEP-TILE-
47 Nanopore pipelines had consensus sequencing accuracy of 100% in mock samples with known
48 DNA sequences. Together, these protocols will facilitate the generation of HBV sequence data,
49 enabling a more accurate and representative picture of HBV molecular epidemiology, cast light
50 on persistence and pathogenesis, and enhance understanding of the outcomes of infection and
51 its treatment.

52

53 **Keywords**

54 Hepatitis B virus (HBV), whole genome sequencing, Nanopore, Illumina

55

56

57 Manuscript

58

59 1. Introduction

60

61 Worldwide an estimated 254 million people are living with chronic HBV infection (CHB) [1].
62 Despite the availability of prophylactic vaccines and suppressive antiviral therapies, the annual
63 toll of approximately 1.1 million HBV-related deaths in 2022 [1] highlights the persistent global
64 challenge faced. HBV whole genome sequencing (WGS) provides genetic insights into the
65 pathogen at population and individual levels [2]. At a population level, WGS improves our
66 understanding and monitoring of HBV drug resistance, vaccine efficacy, diagnostic escape and
67 tracking of outbreaks. In individuals, WGS has the potential to contribute to personalised
68 healthcare, e.g. informing modification of antiviral regimens to take into account antiviral
69 resistance associated mutations (RAMs), and risk-stratified hepatocellular carcinoma (HCC)
70 surveillance [3].

71

72 There are several challenges associated with HBV WGS. Target enrichment using probe based
73 capture, HBV specific polymerase chain reaction (PCR) or rolling circle amplification (RCA) [4,5]
74 is required when performing HBV WGS due to frequently low DNA viral loads (VL) [6], the tiny
75 genome (3.2kB), and abundant host nucleic acid background. Sequencing low VL samples,
76 although challenging, is important at both an individual and population level, in order to
77 understand mechanisms of transmission, persistence, immunological control and clearance,
78 breakthrough viraemia on treatment, and for characterising occult HBV infection. HBV has ten
79 distinct genotypes (A–J), with >8% nucleotide divergence, which are geographically distributed
80 and associated with infection outcomes [7]. This diversity leads to challenges in designing pan-
81 genotypic capture probes and PCR primers. The structure of the HBV genome also presents
82 challenges for WGS. In the blood, HBV DNA is present predominantly in a relaxed-circular (rc-
83 DNA) form, a circular, partially double-stranded (ds)DNA configuration (**Figure 1A**) [8]; neither
84 positive nor negative DNA strand is continuous.

85

86 We set out to optimise and validate workflows for HBV WGS. Due to limitations in sensitivity of
87 our RCA protocol [4] (related to difficulty in generating a fully circular template), we initially used
88 probe-based enrichment, a strategy which has previously been validated across diverse
89 pathogens [9–11]. The method is based on the design of biotinylated single-stranded DNA
90 probes, designed to hybridise with target viral sequences within a sequencing library, and
91 selectively capture HBV DNA, thereby enriching the final library for HBV DNA and increasing on-
92 target sequencing yield (**Figure 1B**).

93

94 Second, due to limitations in sensitivity of probe-based enrichment, we developed a tiled amplicon
95 approach for HBV whole genome sequencing (HEP-TILE) (**Figure 1C**) using HBV-specific
96 primers designed using PrimalScheme3, successor to PrimalScheme [12,13]. PrimalScheme3
97 includes key updates to allow the generation of WGS from pathogens with circular genomes, with
98 primers able to deal with greater pathogen diversity. The HBV sequencing wet and dry lab
99 workflows build on the widely used ARTIC-Network nCoV-2019 sequencing protocol [14] and

100 fieldbioinformatics [15], both of which have been used extensively worldwide during the SARS-
101 CoV-2 pandemic.

102

103 Scale up of WGS requires development of open access, easily deployable sequencing methods
104 ideally building on existing infrastructure and expertise. By developing the protocols described
105 here, we aim to facilitate the generation of HBV sequence data globally, to obtain a more accurate
106 and representative picture of HBV genetics and molecular epidemiology.

107

108

109

110 **2. Methods**

111

112

113 **2.1 Samples**

114

115 **2.1.1 Mock samples**

116 We prepared mock HBV samples using plasmids containing ~1.3 copies of the HBV genome
117 representing genotypes A-J in a pUC57 backbone, (genotypes A-F from [16], G-J designed in-
118 house and produced by GeneArt) ([Supplementary table 1](#)). We spiked plasmids into non-
119 infected human serum (Merck, H5667) at a concentration of 10^5 plasmid copies/ml.

120

121 **2.1.2 Clinical samples**

122 We used 63 plasma samples collected between 2011-2023 from adults with CHB in three settings:
123 (1) outpatient clinics in Oxford (UK) [6,17], (2) outpatient clinics in Tygerberg Hospital, Cape Town
124 (Stellenbosch University) and public sector hospital services in Bloemfontein (University of the
125 Free State), South Africa [6], and (3) participants of the Uganda Liver Disease Study (Kalungu
126 District, south west Uganda [18,19]). We selected samples to represent a range of VL and
127 genotypes; methods were developed iteratively and based on available samples. ([Figure 2](#),
128 [Supplementary table 2](#)).

129

130 Written informed consent was provided by participants at enrollment. Approval for this work was
131 provided as follows, all methods were performed in accordance with the relevant protocols:

132

133

134

135

136

137

138

139

140

141 We collected blood samples in EDTA tubes. To separate plasma, we centrifuged whole blood at

142 1800rpm for 10 minutes. We harvested the supernatant and stored it at -80°C .

143

144 Quantification of HBV DNA VL was performed on the clinically validated Abbott M2000 platform
145 by Oxford University Hospitals (OUH) diagnostic microbiology laboratory for UK samples, and
146 with the Cobas Ampliprep/Taqman HBV test for Ugandan and South African samples (performed
147 by the Clinical Diagnostic Laboratories for MRC/UVRI/LSHTM Research Institute Uganda and
148 the NHLS Tygerberg Virology Laboratory respectively). VL ranged from 1.5 to > 8.3 log₁₀ IU/ml
149 (i.e. above the assay limit of quantification) ([Supplementary table 2](#)).

150

151

152 **2.1.3 Host depletion and nucleic acid extraction**

153 For mock and clinical samples initially being enriched by capture, we performed host depletion
154 with micrococcal nuclease (NEB catalogue number M0247S, 250µl plasma with 1.25µl MNase
155 and 23.75µl MNase buffer incubated for 20 mins at 37°C) followed by extraction using the
156 Kingfisher Apex Magmax viral/pathogen nucleic acid isolation kit (two aliquots of MNase-treated
157 plasma eluted into 50µl of kit elution buffer, combined followed by a SPRIselect bead clean up to
158 reduce volume to 40µl).

159

160 For samples from UK and Uganda for HEP-TILE Nanopore sequencing, we extracted DNA using
161 the QIAgen MinElute Virus Spin kit from either 200µL or 400µL input plasma. Double volume
162 (400µl) extraction was used if VL was <3 log₁₀ IU/ml (with double volume protease, ethanol and
163 buffer AL). We used carrier RNA as per manufacturer's instructions, and eluted into 45µl H₂O.
164 The sample volume and extraction method used for each sample is detailed in [Supplementary
165 table 2](#).

166

167

168

169 **2.2 Public data for probe/primer design**

170 We based our probe and primer design on downloads of all available HBV non-recombinant whole
171 genomes from the Hepatitis B Virus Database (HBVdb) [20] on 31st January 2019 for probe
172 design and 15th February 2024 for primer design ([Supplementary table 3](#)). In addition for HEP-
173 TILE, genomes for genotypes I and J were sourced from McNaughton et. al [7].

174

175

176 **2.3 HBV probe based enrichment**

177

178 **2.3.1. Probe design**

179 We used RaxML [21] with a general time reversible model with Gamma model of rate
180 heterogeneity (“-m GTRGAMMA” option in RAXML) to infer a maximum likelihood phylogeny of
181 HBV full genomes. Next, we used RAXML to infer the ancestral sequences and as input we used
182 the midpoint rooted tree and the sequence of our isolates with the GTRGAMMA option. We used
183 the ancestral sequence at the root of the tree to design the first set of probes assuming that this
184 sequence on average has the least amount of divergence relative to all other isolates.

185

186 As the HBV genome is circular, we added 120 bases from the beginning of the ancestral root
187 sequence to the end of the sequence to ensure that capture probes cover the break point which
188 is used to present the genome linearly. We then divided the ancestral root sequence into 120
189 nucleotide (nt) segments with 60 bases overlap which resulted in an initial set of 55 probes
190 (Figure 1B). Genotype G has an insertion of 36 bases relative to other genotypes in the core
191 gene. As the ancestral root sequence contained this insertion, we also designed a probe of 120
192 nt which lacked this insertion. Furthermore, genotype D has a deletion of 33 bases in the pre-S1
193 region relative to all other genotypes. To ensure a probe covers this region for genotype D, we
194 designed a probe of 120 nt that lacked this region.

195
196 Our previous work in HCV probe-based sequence capture demonstrated that probes of 120 bases
197 long can tolerate up to 20% divergence relative to their target sequence before the efficiency of
198 capture drops [11]. To ensure that the designed probes are within a maximum of 20% divergence
199 of each viral sequence, we divided the dataset based on genotype and created a consensus
200 sequence for each genotype. We then aligned the probes to each viral sequence and measured
201 the proportion of mismatches between each probe and the isolate. For each viral sequence, if a
202 continuous region of ≥ 60 bases diverged from the probe sequences by $>20\%$, a new probe was
203 designed for the region using the genotype consensus sequence. As a quality control step, we
204 removed any potential sequences that contained an “N” as we assumed that the sequence may
205 be of low quality. Additionally we counted the number of ambiguous nucleotides and any viral
206 sequence containing ≥ 5 ambiguous nucleotides was also removed. The final probe set contained
207 74 probes.

208
209 The probe sequences and the set of HBV genome sequences that were used for their design are
210 presented in [supplementary table 4](#) and can also be downloaded from the following webpage:
211 https://figshare.com/articles/dataset/HBV_probe_sequences/22127015

212
213 **2.3.2. Library preparation, hybrid capture and sequencing protocol**
214 We prepared libraries for Illumina sequencing using the NEBNext Ultra II FS DNA library
215 preparation kit (protocol for use with inputs $<100\text{ng}$). We used an input of $26\mu\text{L}$ extracted DNA for
216 the fragmentation/end preparation reaction and fragmented for 3 minutes, with a negative
217 (mastermix only) control. We performed indexing PCR using NEBNext multiplex Oligos UDI and
218 13 amplification cycles. We pooled the amplified libraries corresponding to each aliquot in
219 equivolume proportions to generate a final multiplex library. We purified the pool using SPRIselect
220 beads, eluted into a final volume of $20\mu\text{L}$ and subsequently quantified using High Sensitivity
221 dsDNA Qubit assay (Invitrogen) and Agilent 2100 Bioanalyzer high sensitivity DNA protocol.

222
223 We concentrated a $4.4\mu\text{g}$ aliquot of the final multiplexed library using the manufacturer’s AMPure
224 XP Bead DNA concentration protocol, then enriched for HBV using the custom-designed probe
225 panel (IDT Technologies) and xGen Hybridization and Wash Kit (IDT Technologies) following
226 manufacturer’s ‘tube protocol’. We amplified the final enriched library (12 cycles on-bead PCR),
227 repurified (eluting into $12\mu\text{L}$ EB), quantified using High Sensitivity dsDNA Qubit assay (Invitrogen)
228 and Agilent 2100 Bioanalyzer high sensitivity DNA protocol, then sequenced on the Miseq v3 with
229 $2\times 300\text{nt}$ paired-end reads or on a partial lane of Novaseq X $2\times 150\text{nt}$ paired-end reads.

230

231 **2.3.3. Data analysis - capture Illumina pipeline**

232 We trimmed de-multiplexed sequence read-pairs of low-quality bases using QUASR (version
233 7.01) [22], trimmed adapter sequences with CutAdapt (version 4.8) [23] and Skewer (version
234 0.2.2) [24] and subsequently discarded if either read had <50bp sequence remaining. We mapped
235 the cleaned read pairs to human reference genome hg19 using Bowtie2 (version 2.2.4) [25] and
236 excluded from further analyses. All nonhuman read pairs were mapped using BWA (version
237 0.7.10) [26,27] to a set of 44 HBV references covering all known HBV genotypes and
238 subgenotypes to choose an appropriate reference [7]. We chose the HBV reference with the
239 greatest number of HBV reads mapping to it as the genetically closest reference to the sequenced
240 isolate. Next, we re-mapped all non-human read pairs to the closest HBV reference. We then
241 used Picard markduplicates tool (version 1.111) [28] to remove duplicate read pairs (where read
242 pairs starting in the same place and ending in the same place on the genome are assumed to be
243 PCR duplicates). A base was called at the consensus level if allele depth was above x5, with HBV
244 genome coverage calculated by determining the percentage of bases called at the consensus
245 level.

246

247

248 **2.4 HBV tiled amplicon scheme (HEP-TILE)**

249

250 **2.4.1 PrimalScheme 3**

251 We developed a pan-genotypic HBV scheme using an early version of PrimalScheme3
252 (<https://github.com/ChrisgKent/primalscheme3>), the successor to PrimalScheme [13], a web-
253 based primer design tool for developing multiplex primer schemes. A number of changes were
254 made in PrimalScheme3 to enable us to generate an overlapping (tiled) amplicon scheme which
255 covered the circular HBV genome, utilising a number of discrete primers at each position to handle
256 intraspecies diversity.

257

258 Usually an amplicon scheme results in a short sequence at each end of the reference genome
259 which is not covered due to primer placement constraints and downstream primer-trimming. With
260 the conventionally designated 'start site' for the circular HBV genome being within the overlapping
261 surface and reverse transcriptase genes (**Figure 1**), this results in the loss of epidemiologically
262 and clinically relevant information. PrimalScheme3 introduces an option to add an amplicon
263 spanning the start/end of circular genomes, resulting in 100% genome coverage.

264

265 Intra-species diversity has posed an issue for amplicon sequencing, as variation within the primer
266 binding sites reduces primer binding efficacy. The original PrimalScheme identified conserved
267 primer binding sites by heavily penalising those with variation within the input genomes. This
268 approach is effective for closely related genomes, such as outbreak strains, but for diverse inputs
269 a different approach is needed. PrimalScheme3 handles diversity with the use of 'primer clouds'
270 i.e. a discrete set of 3'-anchored primers which cover all mutations above a user-specified
271 frequency, without the use of ambiguous bases. This approach reduces the negative effects of
272 adding additional primers to a multiplexed PCR, such as primer-primer interactions, and
273 mispriming.

274

275

276 **2.4.2 Multiplex primer pool design for HBV (HEP-TILE)**

277 We aligned HBVdb HBV genomes from each genotype (A-H) separately using MAFFT [29], and
278 phylogenetically downsampled each genotype to 0.95 relative tree length using TreeMMer [30].
279 We combined each downsampled genotype's genomes with genotype I and J sequences from [7],
280 aligned with MAFFT [29], and input into PrimalScheme3 command line tool, with the options; --
281 ampliconsize 600, --minbasefreq 0.02, --backtrack and --minoverlap 20.

282

283 Initially, we developed a 500bp, eight amplicon scheme
284 (<https://labs.primalscheme.com/detail/hbv/500/v1.1.0>) [31], however when trialled, one amplicon
285 spanning 1700-2000 nt consistently dropped out in samples with VL <5 log₁₀ IU/ml
286 (**Supplementary figure 1A and 1B**). We hypothesise that is explained by the structure of HBV
287 DNA: in low VL samples only rc-DNA is present, which has discontinuities in both the positive and
288 negative DNA strands in that region leading to amplicon dropout. In contrast, in higher VL
289 samples, covalently closed circular DNA (cccDNA) is also present in the plasma [32,33]. cccDNA
290 has continuous positive and negative strands which act as a template for this amplicon. We
291 identified a suitable location for primers spanning this region, taking into account the structure of
292 rc-DNA, and re-designed a 600bp primer scheme (hbv/600/v2.1.0) (**Supplementary figure 1C-**
293 **E**), resulting in a six amplicon scheme with amplicons ~600-715bp long, generating a theoretical
294 100% HBV genome coverage (**Figure 1C**). This scheme uses 131 primers to cover the HBV
295 diversity present in our reference dataset.

296

297 **2.4.3 HEP-TILE PCR and sequencing protocol**

298 Amplification and library preparation were performed by adapting the nCoV-2019 LoCost v3
299 sequencing protocol [14], with the removal of the reverse transcriptase step (full protocol on
300 protocols.io [34]). Extract from samples with VL >6 log₁₀ IU/ml were diluted 10 fold prior to PCR.
301 Negative (mastermix only) controls were added at the PCR and library prep stages. We used the
302 native barcoding kit SQK-NBD114.96 and R10.4.1 MinION flowcells (Oxford Nanopore
303 Technologies, ONT), multiplexing up to 96 samples per run and sequencing for 72 hours. For
304 Illumina sequencing of tiled amplicons, we fragmented the amplicons and performed library
305 preparation using the NEBNext Ultra II FS DNA kit, multiplexing up to 96 samples per run and
306 sequenced on a partial lane of Novaseq X 150PE.

307

308 **2.4.4 Data analysis - HEP-TILE Nanopore pipeline**

309 We basecalled reads and demultiplexed data using dorado (0.7). We generated a novel pipeline
310 hbv-fieldbioinformatics [35] to produce consensus genomes. Reads are mapped against
311 representative genomes from all genotypes [7], with the genome with the most mapped reads
312 being selected as the primary reference. To handle the circular genome, an intermediate circular
313 primary reference genome is created, where the sequence of the circular amplicon is appended
314 to the 3' end of the reference. Reads are then remapped to the circular reference with minimap2
315 (2.26), primers are trimmed, followed by variant calling by LongShot and Medaka (0.4.5, 1.11.3).
316 Variants and read depth for the appended region are then mapped back to their corresponding
317 position on the original reference. A base was called at the consensus level if allele depth was

318 over x20, with HBV genome coverage calculated by determining the percentage of called
319 consensus bases.

320

321

322 **2.4.5 Data analysis - HEP-TILE Illumina pipeline**

323 Similar to the capture-based Illumina pipeline, we initially used QUASR (version 7.01) [22] to filter
324 out poor de-multiplexed sequencing reads. We then trimmed primer sequences by using
325 CutAdapt (version 4.8) [23] and Skewer (version 0.2.2). Human-like reads were excluded with
326 Bowtie2 (version 2.2.4) [25]. We extended all 44 HBV genome references by adding their first 300
327 bases to the end and mapped the remaining non-human reads to these new HBV reference
328 sequences to identify the genetically closest HBV isolate. In the next two rounds, we mapped the
329 deduplicated reads to 1) the closest HBV isolate and 2) a fine-tuned consensus sequence based
330 on mapping in 1). In the final stage, we cut reads that cross the conventional genome end and
331 combined mappings in the conventional and appended regions. A base was called at the
332 consensus level if allele depth was over x5, with HBV genome coverage calculated by determining
333 the percentage of called consensus bases.

334

335 **2.5 Determining clinically relevant mutations**

336 HBV genotypes, including recombinants, were determined using the online NCBI genotyping tool
337 [36]. Subgenotypes, resistance associated mutations and vaccine escape mutations were called
338 using the online Geno2pheno HBV tool [37].

339

340 **2.6 Software**

341 Graphs were produced using R version 4.2.2 [38], figures were produced with Biorender with a
342 licence to publish [39].

343

344 **2.7 Data availability**

345 Sequences for samples where full length genomes were generated were deposited to the
346 European Nucleotide Archive (ENA) under the study PRJEB79403 for Nanopore data and
347 PRJEB79773 for Illumina.

348

349 **3. Results**

350

351 **3.1 Targeted enrichment with probe based capture**

352

353 **3.1.1 Sequencing mock HBV samples enriched with probe based capture using Illumina**

354 We sequenced 10 mock samples (plasmids with genotypes A-J, 5 log₁₀ copies/ml) in duplicate,
355 with/without probe-based capture on the Illumina Novaseq X, libraries were diluted to 10pM and
356 pooled prior to sequencing. Without capture, a median of 391 (IQR 153 - 1858) HBV reads per
357 million reads sequenced was generated across 20 mock samples. With capture a median of
358 563,956 (IQR 397523 - 896635) HBV reads per million reads sequenced was generated
359 ([Supplementary table 1](#)). Enrichment with probe-based capture led to the generation of full
360 length HBV consensus sequences for all genotypes ([Figure 3A, Supplementary table 1](#)).

361

362 **3.1.2 Sequencing clinical HBV samples enriched with probe based capture using Illumina**

363 We enriched a subset of 23 clinical samples (HBV VL 4.21 to >7.20 log₁₀ IU/ml, as optimisation
364 work established ~4 log₁₀ IU/ml as a conservative estimate for the threshold at which full genomes
365 could be obtained at x5 depth, (Figure 2) using probe based capture and sequenced the samples
366 on Illumina Novaseq X. A median of 22,377 (IQR 3267 - 135,184) HBV reads per million reads
367 sequenced was generated. The relationship between HBV VL and percentage reads on target is
368 shown in (Figure 4A, 4D). Full length (>98%) genomes at x1 coverage were generated in 18/23
369 samples, with partial genomes in 5/23. Full length genomes above a minimum x5 read depth
370 (minimum required by bioinformatic pipeline to call base at the consensus level) were generated
371 in 7/23 samples, all of which had VL >5.5 log₁₀ IU/ml (~300,000 IU/ml). Below VL 5.5 log₁₀ only
372 partial genomes at x5 depth were generated (Supplementary table 2). The negative controls
373 had 0% consensus coverage of the HBV genome.

374

375 **3.2 Targeted enrichment with HEP-TILE tiled amplicon PCR**

376

377 **3.2.1 Sequencing mock HBV samples enriched with HEP-TILE using Nanopore and Illumina**

378 We sequenced the same 10 mock samples in duplicate, enriching using the HEP-TILE tiled
379 amplicon scheme and sequenced on Nanopore MinION and Illumina Novaseq X. With HEP-TILE
380 Nanopore a median of 997,016 (IQR 994,669 - 999,427) HBV reads per million reads sequenced
381 was generated. With HEP-TILE Illumina a median of 977,441 (IQR 963,459 - 982,631) HBV reads
382 per million reads sequenced was generated (Supplementary table 1). Enrichment with HEP-
383 TILE led to the generation of full-length HBV sequences across genotypes on both platforms
384 (Figure 3B, 3C).

385

386 **3.2.2 Sequencing clinical HBV samples enriched with HEP-TILE using Nanopore**

387 We enriched a subset of 50 clinical samples (HBV VL 1.50 to >8.23 log₁₀ IU/ml, Figure 2) using
388 the HEP-TILE protocol and sequenced on the Nanopore MinION. A median of 779,909 (IQR
389 590,906 - 676,536) HBV reads per million reads sequenced was generated. The relationship
390 between HBV VL and percentage reads on target is presented (Figure 4B, 4E).

391

392 Full genome coverage (x1 depth) was generated in 40/50 samples. Full-length HBV consensus
393 genomes at x20 depth (the minimum required to accurately call a mutation), were generated in
394 40/50 samples, down to the lowest VL tested (1.5 log₁₀ IU/ml). Partial genomes were obtained for
395 the remaining 10 samples. These low coverage samples tended to have lower VL (Figure 4),
396 although this did not reach statistical significance (p = 0.16, Mann Whitney U test). We did not
397 observe any relationship between HBV genotype and inability to produce whole genomes
398 (Supplementary figure 2). The negative controls had 0% consensus coverage of the HBV
399 genome.

400

401 **3.2.3 Sequencing clinical HBV samples enriched with HEP-TILE using Illumina**

402 We selected a subset of 10 samples (based on sample availability) for HEP-TILE amplification,
403 followed by fragmentation and sequencing on Illumina (Novaseq X 150PE partial lane) (Figure
404 2). A median of 956,800 (IQR 759,122 - 977,744) HBV reads per million reads sequenced was
405 generated. The relationship between HBV VL and percentage reads on target is shown in (Figure

406 **4C, 4F**). Full genome coverage (x1 depth) was generated in 7/10 samples. Full-length consensus
407 HBV genomes (minimum x5 read depth) were generated in 7/10 samples, partial genomes were
408 obtained for 2/10 samples and the pipeline failed to generate a consensus from 1 sample due to
409 low read coverage and anomalous mapping mismatches. The negative controls had 0%
410 consensus coverage of the HBV genome.

411

412

413

414 **3.3 Methods comparison**

415

416 **3.3.1 Accuracy of consensus genomes**

417 We sequenced all 10 mock samples using all three workflows, with each sample processed in
418 duplicate (capture Illumina, HEP-TILE Illumina, HEP-TILE Nanopore). Comparison of consensus
419 sequences generated from mock samples to the known plasmid sequence, demonstrated 100%
420 accuracy of consensus genomes produced with capture-Illumina and HEP-TILE Nanopore and
421 99.98% with HEP-TILE Illumina (**Table 1**). Four samples sequenced with HEP-TILE Illumina had
422 mismatches or bases designated 'N', however the median number of mismatches and N's was 0.
423 There were no indels for any of the methods.

424

425 **3.3.2 Comparison of capture vs. HEP-TILE ability to generate whole genomes from high** 426 **viral load clinical samples**

427 We sequenced a subset of 10 clinical samples ($> 5 \log_{10}$ IU/ml) using all three workflows so that
428 we could compare the ability of capture vs. HEP-TILE to generate HBV WGS at high VL. At x1
429 depth, capture-Illumina and HEP-TILE Nanopore generated full genomes in 8/10 samples, with
430 HEP-TILE Illumina in 7/10 samples. At depths required to confidently call consensus (x5 for
431 Illumina, x20 for Nanopore), full genomes were generated in 4/10 samples for capture-Illumina,
432 8/10 for HEP-TILE Nanopore and 7/10 for HEP-TILE Illumina (**Supplementary table 2**).

433

434 **3.4. Determining genotypes from HBV WGS data**

435 We identified samples of genotype A-E, in addition to genotype B/C and D/E recombinants (see
436 methods) (**Supplementary table 2**). We were able to call polymorphisms at sites of previously
437 reported resistance associated mutations (RAMs) and vaccine escape mutations (VEMs) [37]
438 (details in **Supplementary text**).

439

440

441

442

443 **4. Discussion**

444

444 **4.1 Summary of outputs**

445

445 We have developed two pan-genotypic viral enrichment methods for HBV WGS: HEP-TILE and
446 probe based capture. Both techniques effectively enrich HBV genomes from all known genotypes
447 (A-J), as evidenced by application to mock samples, and increase the percentage of sequencing

448 reads mapping to the HBV genome, therefore reducing the amount of sequence data needed to
449 obtain full genome coverage. We used HBV WGS to call HBV genotypes, including recombinants,
450 which may be missed with partial-genome sequencing. HEP-TILE capitalises on the rapid global
451 expansion of technical expertise and infrastructure for amplicon based WGS that occurred during
452 the SARS-CoV-2 pandemic. Having a toolbox of diverse sequencing methods allows the most
453 appropriate method to be selected according to local infrastructure, cost and throughput
454 requirements ([Table 2](#)).

455 **4.2 Comparison between methods**

456 HEP-TILE offers the greatest flexibility to sequence across a range of HBV VL, able to sequence
457 full genomes down to $1.50 \log_{10}$ IU/ml (~ 30 IU/ml, lowest tested here), using standard extraction
458 from 200 μ L - 400 μ L plasma. The primers are designed to be pan-genotypic (A-J), unlike other
459 current schemes designed for genotypes A-E [40,41]. The ~ 600 -715bp amplicons used in HEP-
460 TILE improve sensitivity compared to single amplicon schemes [42], enhancing sequencing
461 success from low VL or degraded samples. While HEP-TILE amplicons can be sequenced on
462 either Nanopore or Illumina, the use of the ONT native barcoding kit described here (ligation
463 based) rather than ONT rapid barcoding kit or NEB-Ultra-II-FS kit (both requiring fragmentation)
464 enables the sequencing of whole amplicons. Although shorter amplicon schemes (300-600bp)
465 can facilitate transfer to Illumina platforms without amplicon fragmentation, HBV diversity and
466 genome structure prohibit the generation of a shorter pan-genotypic scheme ([Supplementary
467 figure 1](#)). Generally, shorter amplicons improve sensitivity and platform transferability, however
468 longer 1-2 amplicon schemes or schemes generating concatenated full genomes are useful for
469 haplotype reconstruction [4,40]. The use of two primer pools (instead of individual reactions for
470 each amplicon) simplifies the PCR step and eliminates the need for re-balancing of individual
471 primer pools post-PCR required by other methods [43].

472 We report percent genome coverage at x1 depth in addition to the percentage of the genome with
473 consensus calls to allow comparison with other published methods presenting coverage at x1
474 depth. We use a higher threshold of x20 depth to call bases at a consensus level for Nanopore,
475 and x5 for Illumina to improve certainty of the consensus base call (higher threshold for Nanopore
476 used due to the higher error rate of Nanopore sequencing technology).

477 [Table 2](#) compares the costs of each workflow, estimated per sample based on an initial batch of
478 1000 samples. Cost per sample decreases with multiplexing (e.g. 96 samples for Nanopore, more
479 possible for Illumina). Due to the initial cost of purchasing capture probes (\sim £6300 for this panel),
480 sequencing smaller sample sets with capture becomes prohibitively expensive. Costs for Illumina
481 sequencing vary significantly depending on whether the option of purchasing a partial lane on a
482 high throughput platform is available.

483 **4.3 Real world application of methods**

484 Although full genome HBV sequencing coverage should be achievable for the majority of samples
485 using HEP-TILE, if samples at extreme ends of the VL spectrum are pooled on the same run,
486 uneven read distribution is expected between samples, with loss of coverage for the lowest VL

487 samples. Therefore batching samples by VL and normalising the mass of PCR product being
488 taken forwards into library preparation is recommended. Viral enrichment by PCR is highly
489 sensitive to amplicon contamination from previous experiments therefore caution should be taken
490 to keep work areas, equipment and reagents free of contamination, and multiple negative controls
491 should be included. To avoid *in-silico* 'contamination' double barcoding is required when
492 demultiplexing to eliminate chimeric reads which may otherwise lead to mis-assignment of reads
493 to incorrect barcodes [44]. These wet lab and bioinformatic principles are also applicable to
494 sequencing other targets.

495 Although our capture approach can be used to improve the efficiency of viral sequencing at VL
496 $> \sim 5.50 \log_{10}$ IU/ml, its efficacy for producing full-length consensus genomes diminishes at lower
497 VLs, as observed elsewhere [45]. Additional/alternative host depletion and high volume (e.g. 5ml
498 [46]) extraction is likely to be required to reliably obtain full consensus genomes using this method
499 in samples with VL $< 5.5 \log_{10}$ IU/ml.

500 **4.4 Caveats and limitations**

501 The results presented here are from an iterative set of experiments and protocol development,
502 rather than a pre-planned head-to-head comparison. This limits the direct comparability of
503 workflows due to insufficient sample volumes of the original South Africa sample set. This meant
504 we were unable to sequence all samples using all three enrichment approaches. Furthermore,
505 the extraction method changed over time. Nevertheless, these data showcase the capabilities of
506 each workflow.

507 Correlation between VL and percentage genome coverage was more variable than anticipated.
508 Factors such as storage duration, shipping conditions, and repeated freeze-thaw cycles may all
509 influence sample quality, although DNA is generally robust. Poor sample quality or inaccurate VL
510 reporting (different sample/timepoint tested or lab/reporting error) may explain instances where
511 samples that were initially reported to have high VL yielded lower coverage than expected (e.g.
512 sample 1116 for which VL was reported as $5.41 \log_{10}$ IU/ml but on repeat on our sample was 3
513 \log_{10} IU/ml).

514 The probes and primers were designed to be pan-genotypic, however only take into account the
515 diversity of sequences on HBVdb, which contains relatively few genotype F-H sequences, and
516 under-represents certain global populations [ref Delphin?]. A strength of our study is the use of
517 mock samples, which allowed us to assess performance across all genotypes, including rarer
518 genotypes F-J that were unlikely to be present in our cohorts, however might still be present in
519 other cohorts globally. Analysis of rarer genotypes is often missing in reports of sequencing
520 schemes, however will become increasingly important as more HBV WGS is performed from
521 diverse global regions, and known HBV sequence diversity expands. While we report these
522 methods as pan-genotypic using mock samples generated from plasmids containing HBV
523 genotypes A-J, it is important to note that we only tested one variant per genotype. Some
524 genotypes have multiple subtypes; future studies could test enrichment methods with a wider
525 range of subgenotypes.

526 Sequencing low VL samples has inherent stochasticity, depending on the presence of viral DNA
527 in a sample or extract, and assay chemistry. Future work will test HEP-TILE performance at viral
528 loads $<1.5 \log_{10}$ IU/ml, using high volume extraction to increase the chance of sampling viral DNA
529 (e.g. extraction from 5ml serum/plasma [46]), increasing the number of PCR cycles from 35 to 40
530 and pooling products from PCRs performed in duplicate or triplicate. Future work could compare
531 the ability of capture or HEP-TILE to accurately quantify low abundance minority variants,
532 requiring a panel of control material at varying VLs with known variant percentages, sequenced
533 with multiple technical replicates. Finally, enrichment methods could be trialled on different
534 sample types, for example liver biopsies where high host background is a challenge.

535

536 **4.5 Conclusion**

537 Collectively, these protocols will facilitate the wider generation of HBV sequence data. This data
538 will lead to a more accurate and representative global picture of HBV molecular epidemiology,
539 cast light on persistence and pathogenesis, and enhance understanding of treatment responses
540 and clinical outcomes. These aspirations align with high profile international goals for elimination
541 of HBV as a public threat.

542

543

544

545 **References**

- 546 1. Global hepatitis report 2024: action for access in low- and middle-income countries.
547 World Health Organization; 9 Apr 2024 [cited 7 May 2024]. Available:
548 <https://www.who.int/publications/i/item/9789240091672>
- 549 2. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome
550 sequencing. *Nat Rev Microbiol.* 2017;15: 183–192.
- 551 3. Lok AS. Personalized treatment of hepatitis B. *Clin Mol Hepatol.* 2015;21: 1–6.
- 552 4. McNaughton AL, Roberts HE, Bonsall D, de Cesare M, Mokaya J, Lumley SF, et al.
553 Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV).
554 *Sci Rep.* 2019;9: 7081.
- 555 5. Lumley SF, Jennings D, Waddilove E, Trebes A, Delphin M, Downs LO, et al. Pan-
556 genotypic probe-based enrichment to improve efficiency of Hepatitis B virus sequencing.
557 *bioRxiv.* 2023. p. 2023.02.20.529276. doi:10.1101/2023.02.20.529276
- 558 6. Downs LO, Vawda S, Bester PA, Lythgoe KA, Wang T, McNaughton AL, et al. Bimodal
559 distribution and set point HBV DNA viral loads in chronic infection: retrospective analysis of
560 cohorts from the UK and South Africa. *Wellcome Open Res.* 2020;5: 113.
- 561 7. McNaughton AL, Revill PA, Littlejohn M, Matthews PC, Ansari MA. Analysis of genomic-
562 length HBV sequences to determine genotype and subgenotype reference sequences. *J*
563 *Gen Virol.* 2020;101: 271–283.
- 564 8. McNaughton AL, D'Arienzo V, Ansari MA, Lumley SF, Littlejohn M, Revill P, et al.
565 Insights From Deep Sequencing of the HBV Genome—Unique, Tiny, and Misunderstood.
566 *Gastroenterology.* 2019;156: 384–399.
- 567 9. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome Capture
568 Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *MBio.*
569 2015;6: e01491–15.
- 570 10. Depledge DP, Palser AL, Watson SJ, Lai IY-C, Gray ER, Grant P, et al. Specific
571 Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *PLoS ONE.*
572 2011. p. e27805. doi:10.1371/journal.pone.0027805
- 573 11. Bonsall D, Ansari MA, Ip C, Trebes A, Brown A, Klenerman P, et al. ve-SEQ:
574 Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of
575 HCV and other highly diverse pathogens. *F1000Res.* 2015;4: 1062.
- 576 12. PrimalScheme: primer panels for multiplex PCR. [cited 7 May 2024]. Available:
577 <https://primalscheme.com/>
- 578 13. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al.
579 Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus
580 genomes directly from clinical samples. *Nat Protoc.* 2017;12: 1261–1276.
- 581 14. Quick J. nCoV-2019 sequencing protocol v3 (LoCost) v3. 2020.
582 doi:10.17504/protocols.io.bp2l6n26rgqe/v3

- 583 15. fieldbioinformatics: The ARTIC field bioinformatics pipeline. Github; Available:
584 <https://github.com/artic-network/fieldbioinformatics>
- 585 16. Bannister E, Sozzi V, Mason H, Locarnini S, Hardikar W, Revill PA. Analysis of
586 the in vitro replication phenotype of African hepatitis B virus (HBV) genotypes and
587 subgenotypes present in Australia identifies marked differences in DNA and protein
588 expression. *Virology*. 2020;540: 97–103.
- 589 17. Downs LO, Smith DA, Lumley SF, Patel M, McNaughton AL, Mokaya J, et al.
590 Electronic Health Informatics Data To Describe Clearance Dynamics of Hepatitis B Surface
591 Antigen (HBsAg) and e Antigen (HBeAg) in Chronic Hepatitis B Virus Infection. *MBio*.
592 2019;10. doi:10.1128/mBio.00699-19
- 593 18. O'Hara G, Mokaya J, Hau JP, Downs LO, McNaughton AL, Karabarinde A, et al.
594 Liver function tests and fibrosis scores in a rural population in Africa: a cross-sectional
595 study to estimate the burden of disease and associated risk factors. *BMJ Open*. 2020;10:
596 e032890.
- 597 19. Asiki G, Murphy G, Nakiyingi-Miiró J, Seeley J, Nsubuga RN, Karabarinde A, et
598 al. The general population cohort in rural south-western Uganda: a platform for
599 communicable and non-communicable disease studies. *Int J Epidemiol*. 2013;42: 129–141.
- 600 20. Hayer J, Jadeau F, Deléage G, Kay A, Zoulim F, Combet C. HBVdb: a
601 knowledge database for Hepatitis B Virus. *Nucleic Acids Res*. 2013;41: D566–70.
- 602 21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-
603 analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–1313.
- 604 22. Gaidatzis D, Lerch A, Hahne F, Stadler MB. QuasR: quantification and
605 annotation of short reads in R. *Bioinformatics*. 2015;31: 1130–1132.
- 606 23. Martin M. Cutadapt removes adapter sequences from high-throughput
607 sequencing reads. *EMBnet.journal*. 2011;17: 10–12.
- 608 24. Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer
609 for next-generation sequencing paired-end reads. *BMC Bioinformatics*. 2014;15: 182.
- 610 25. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient
611 alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10: R25.
- 612 26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
613 transform. *Bioinformatics*. 2009;25: 1754–1760.
- 614 27. Li H. bwa: Burrow-Wheeler Aligner for short-read alignment (see minimap2 for
615 long-read alignment). Github; Available: <https://github.com/lh3/bwa>
- 616 28. Picard. [cited 25 Jun 2024]. Available: <http://broadinstitute.github.io/picard>
- 617 29. Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment
618 program. *Bioinformatics*. 2010;26: 1899–1900.
- 619 30. Menardo F, Loiseau C, Brites D, Coscolla M, Gygli SM, Rutaihwá LK, et al.
620 Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC*

- 621 Bioinformatics. 2018;19: 164.
- 622 31. primerschemes/hbv/500/v1.1.0 at main · quick-lab/primerschemes. Github;
623 Available: [https://github.com/quick-](https://github.com/quick-lab/primerschemes/tree/main/primerschemes/hbv/500/v1.1.0)
624 [lab/primerschemes/tree/main/primerschemes/hbv/500/v1.1.0](https://github.com/quick-lab/primerschemes/tree/main/primerschemes/hbv/500/v1.1.0)
- 625 32. Chen Y, Sze J, He M-L. HBV cccDNA in patients' sera as an indicator for HBV
626 reactivation and an early signal of liver damage. *World J Gastroenterol.* 2004;10: 82–85.
- 627 33. Takkenberg RB, Zaaijer HL, Molenkamp R, Menting S, Terpstra V, Weegink CJ,
628 et al. Validation of a sensitive and specific real-time PCR for detection and quantitation of
629 hepatitis B virus covalently closed circular DNA in plasma of chronic hepatitis B patients. *J*
630 *Med Virol.* 2009;81: 988–995.
- 631 34. Lumley S, Kent C, Quick J, Matthews P. HEP-TILE: HBV whole genome
632 sequencing (nanopore protocol) v1. 2024. doi:10.17504/protocols.io.5jyl82bedl2w/v1
- 633 35. Creators nickloman Will Rowe¹ Radoslaw Poplawski Stephen Rudd² Sam W3
634 ChrisKBio³ Andrew Rambaut Trevor Bedford⁴ Karthik⁵ Jared Simpson Josh Quick³
635 mattloose Show affiliations 1. @grailbio 2. @nanoporetech 3. University of Birmingham 4.
636 Fred Hutchinson Cancer Research Center 5. The Scripps Research Institute.
637 ChrisgKent/hbv-fieldbioinformatics: zenodo release. doi:10.5281/zenodo.12626842
- 638 36. Genotyping. [cited 17 Jun 2024]. Available:
639 <https://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>
- 640 37. Geno2pheno hbv. [cited 17 Jun 2024]. Available:
641 <https://hbv.geno2pheno.org/index.php>
- 642 38. Ripley BD. The R project in statistical computing. *MSOR Connect.* 2001;1: 23–
643 25.
- 644 39. Scientific image and illustration software. [cited 26 Jun 2024]. Available:
645 <http://www.biorender.com>
- 646 40. Stenbäck JB, Schmidt D, Noborg U, Gustafsson J, Norberg P, Andersson ME, et
647 al. Accurate and cost-efficient whole genome sequencing of hepatitis B virus using
648 Nanopore. medRxiv. 2024. p. 2024.08.12.24311345. doi:10.1101/2024.08.12.24311345
- 649 41. Tshiabuila D, Choga W, James SE, Maponga T, Preiser W, van Zyl G, et al. An
650 Oxford Nanopore Technology-Based Hepatitis B Virus Sequencing Protocol Suitable For
651 Genomic Surveillance Within Clinical Diagnostic Settings. medRxiv. 2024.
652 doi:10.1101/2024.01.19.24301519
- 653 42. Günther S, Li BC, Miska S, Krüger DH, Meisel H, Will H. A novel method for
654 efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis
655 and reveals deletion mutants in immunosuppressed patients. *J Virol.* 1995;69: 5437–5444.
- 656 43. Ringlander J, Andersson ME, Prakash K, Larsson SB, Lindh M. Deep
657 sequencing of hepatitis B virus using Ion Torrent fusion primer method. *J Virol Methods.*
658 2022;299: 114315.
- 659 44. Xu Y, Lewandowski K, Lumley S, Pullan S, Vipond R, Carroll M, et al. Detection

- 660 of Viral Pathogens With Multiplex Nanopore MinION Sequencing: Be Careful With Cross-
661 Talk. *Front Microbiol.* 2018;9: 2225.
- 662 45. Berg MG, Olivo A, Forberg K, Harris BJ, Yamaguchi J, Shirazi R, et al. Advanced
663 molecular surveillance approaches for characterization of blood borne hepatitis viruses.
664 *PLoS One.* 2020;15: e0236046.
- 665 46. Fu MX, Simmonds P, Andreani J, Baklan H, Webster M, Asadi R, et al.
666 Ultrasensitive PCR system for HBV DNA detection: Risk stratification for occult hepatitis B
667 virus infection in English blood donors. *J Med Virol.* 2023;95: e29144.
- 668
- 669

670 **Funding**

671 This research was funded in whole, or in part, by the Wellcome Trust [220549/Z/20/Z,
672 206298/B/17/Z, 220171/Z/20/Z]. For the purpose of Open Access, the author has applied a CC
673 BY public copyright licence to any Author Accepted Manuscript version arising from this
674 Submission.

675 SFL is funded by a Wellcome Doctoral Training Fellowship (grant number 220549/Z/20/Z).
676 JQ/NL/CK are funded by the Wellcome Trust through the ARTIC Network Collaborative Award
677 (Grant number 206298/B/17/Z). MAA is supported by a Sir Henry Dale Fellowship jointly funded
678 by the Royal Society and Wellcome Trust (220171/Z/20/Z). PCM has funding from the UCLH
679 NIHR Biomedical Research Centre, and core funding from the Francis Crick Institute (ref
680 CC2223).

681

682 **Author contributions**

683 Conceptualization: SFL, JQ, MAA, PCM

684 Methodology: SFL, CK, DJ, HC, AT

685 Software: CK, HC, SW, JQ, MAA

686 Formal Analysis: CK, HC, SFL

687 Investigation: SFL, CK, DJ, HC, GA, EW, MD, AT, SW, YW, G M-C, MK, KO, JC

688 Resources: BK, TGM, CdL, JM, MVS, DG, RN, EB, PCM

689 Data curation: CK, HC, JQ, MAA

690 Writing (original draft): SL, CK

691 Writing (review and editing): all authors

692 Visualisation: SL, CK

693 Supervision: JQ, MAA, PCM

694 Funding acquisition: SFL, JQ, MAA, PCM

695

696 **Data availability statement**

697 Data for this study is available within the manuscript, supplementary table and online links.
698 Sequences for samples where full length genomes were generated were deposited to the
699 European Nucleotide Archive (ENA) under the study PRJEB79403 for Nanopore data and
700 PRJEB79773 for Illumina.

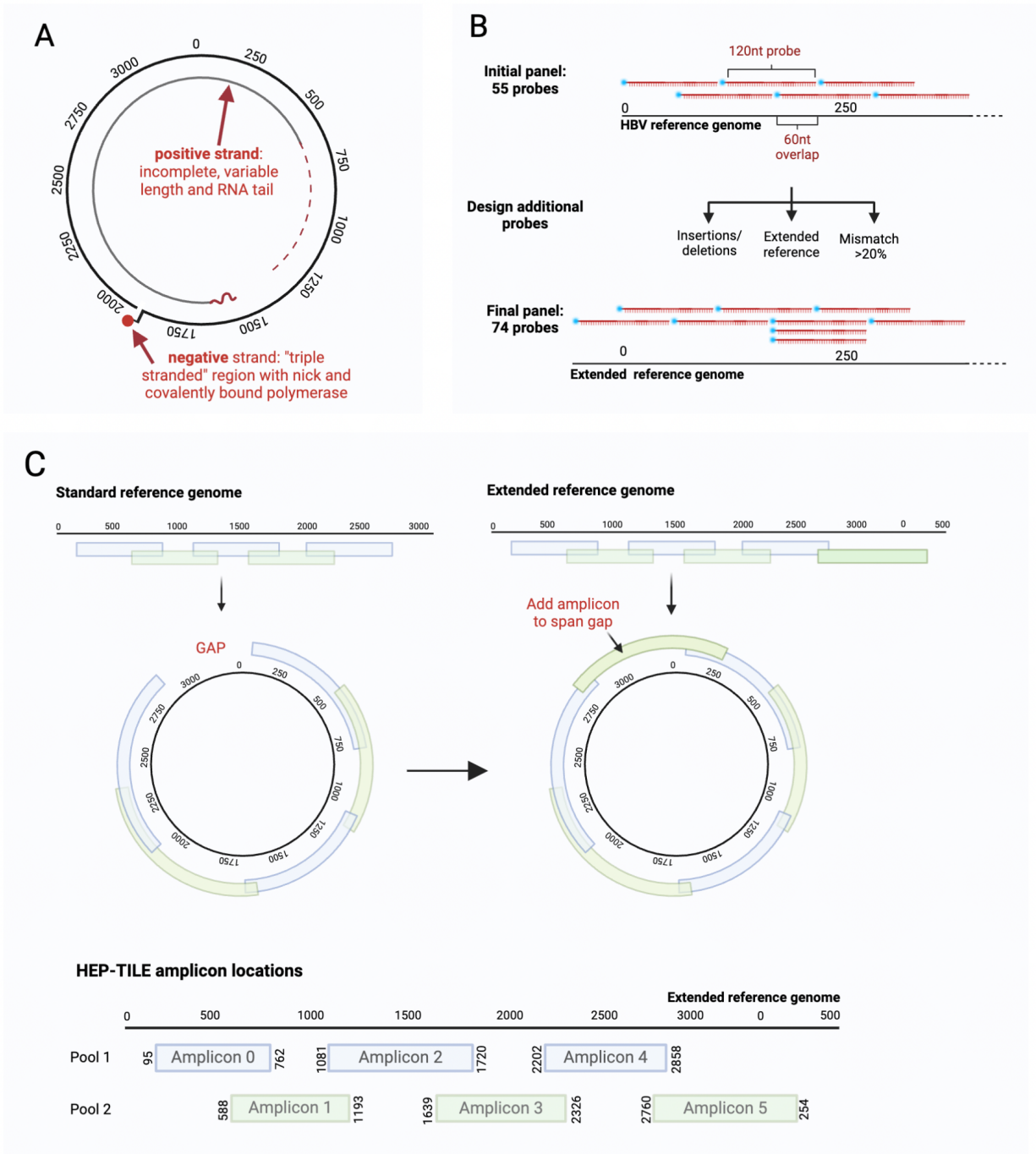
701

702 **Competing interests**

703 Chris Kent has received an honorarium from Illumina for a talk on PrimalScheme3

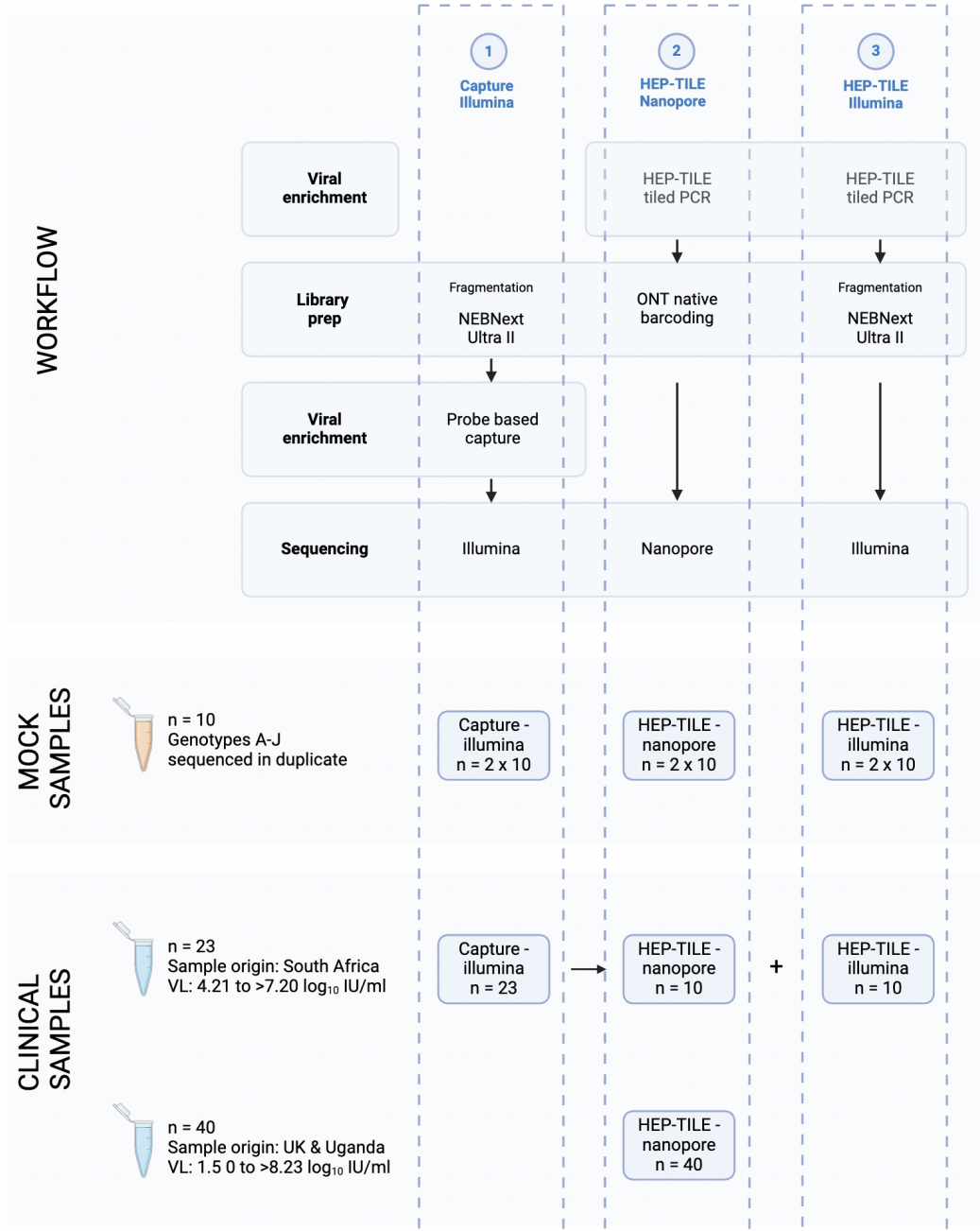
704

705 **Figures**
706



707
708 **Figure 1: Design of HBV targeted enrichment methods.** (A) HBV genome structure: relaxed
709 circular DNA (rc-DNA). The complete negative strand is discontinuous, with a 'nick' and covalently
710 bound polymerase. The positive strand is incomplete with a variable length and RNA tail. (B)
711 Probe design schematic for HBV probe based enrichment. An initial set of probes are designed
712 based on non-recombinant whole genomes downloaded from HBVdb. Additional probes are

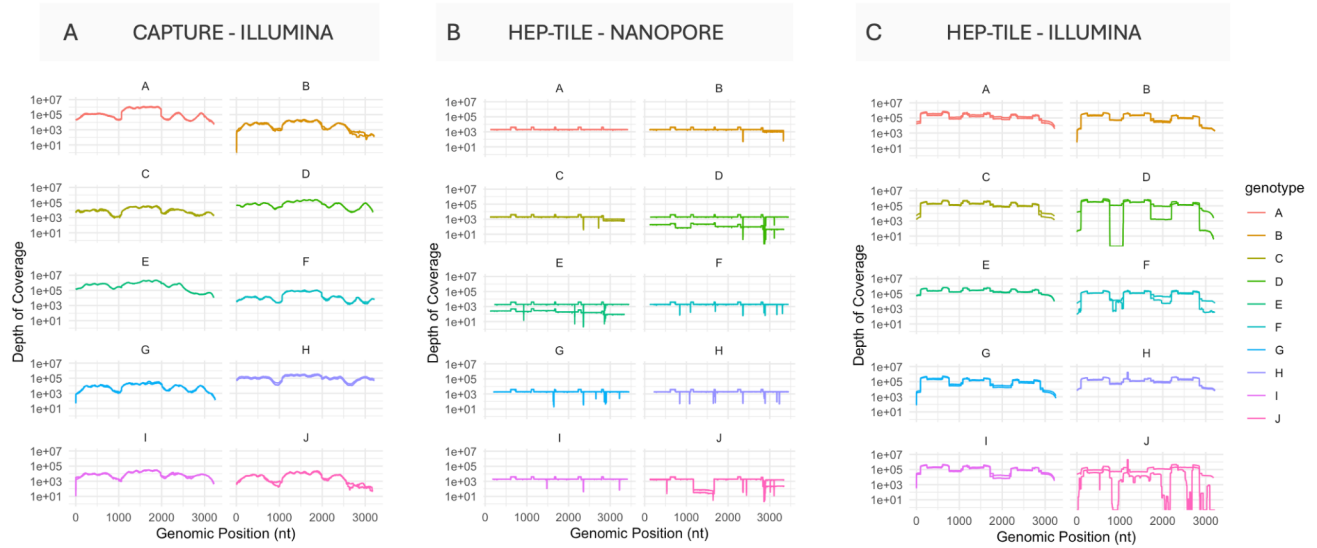
713 added in regions of high diversity, using an extended reference sequence to account for the
714 circular genome, and to account for insertions/deletions, then the final probe panel confirmed. (C)
715 Multiplex tiling PCR for the circular HBV genome. Generation of tiled schemes from linear
716 reference genomes leaves a gap in coverage for a circular genome. An additional amplicon
717 spanning the start/end of the linear reference ensures a theoretical maximum 100% genome
718 coverage can be obtained. Primer positions and amplicon length for HBV tiled amplicon scheme.
719 Numbering of nt positions based on X02763 reference genome. Created in BioRender. Lumley,
720 S. (2024) BioRender.com/r40b439
721
722
723



724
725
726
727
728
729
730
731
732
733

Figure 2: Overview of sequencing workflows. Numbers of samples sequenced at each stage indicated. Due to sample availability, project stage, and local infrastructure, we used different subsets of samples to validate each method. In the clinical sample set from South Africa, if residual extract was available after capture, two further aliquots were taken forward for HEP-TILE enrichment and sequencing on Nanopore and Illumina. Created in BioRender. Lumley, S. (2024) BioRender.com/k41p399

734



735

736

737 **Figure 3: HBV genomes for all HBV genotypes (A-J) in mock clinical samples enriched**
738 **using HBV probes and primers (A) Enriched with probe based capture, sequenced on Illumina;**
739 **(B) Enriched with HEP-TILE, sequenced on Nanopore; (C) Enriched with HEP-TILE, sequenced**
740 **on Illumina.**

741

742

743

744

745

746

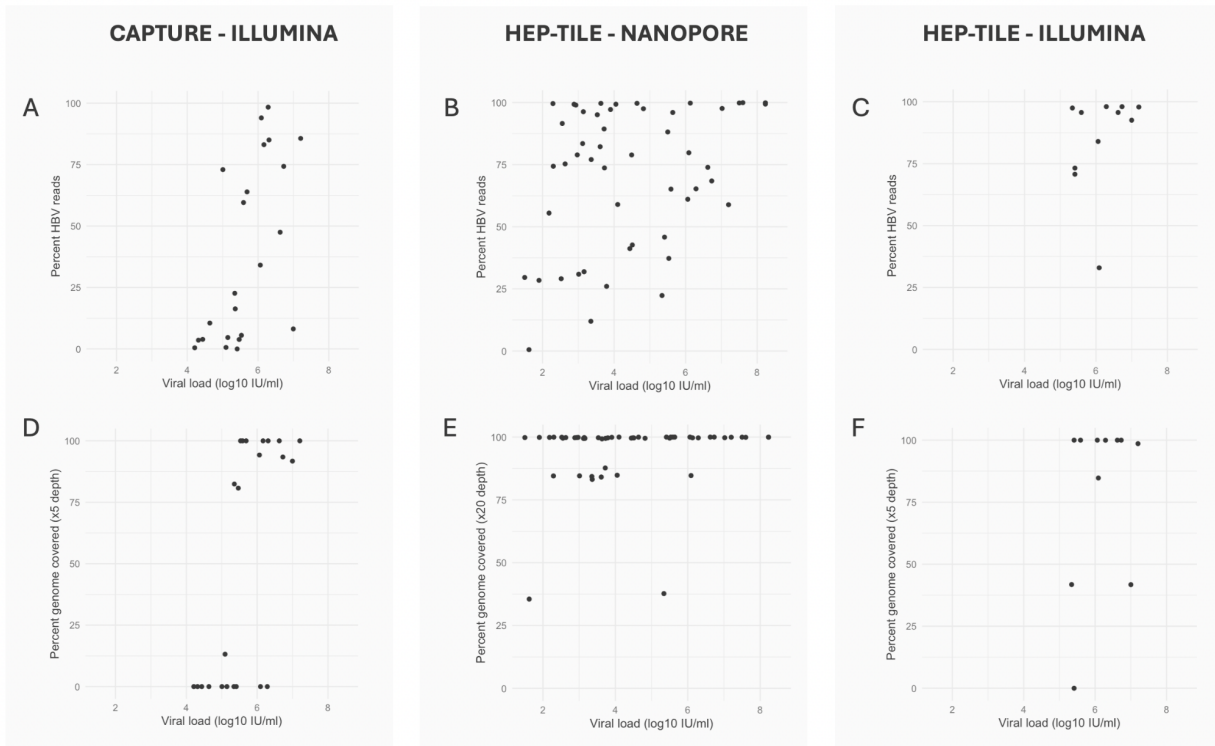
747

748

749

750

751



752

753

754

755 **Figure 4: HBV WGS using capture and amplicon based target enrichment.**

756 *Top row:* Relationship between viral load and percentage of reads on-target (ie. HBV) for (A)
757 probe-based capture sequenced on Illumina, (B) HEP-TILE sequenced on Nanopore and (C)
758 HEP-TILE sequenced on Illumina.

759 *Bottom row:* Relationship between viral load and percentage of genome with high confidence
760 consensus for (D) probe-based capture sequenced on Illumina (x5 minimum depth), (E) HEP-
761 TILE sequenced on Nanopore (x20 minimum depth) and (F) HEP-TILE sequenced on Illumina
762 (x5 minimum depth). Panels E and F show an outlier with viral load 5.41 log₁₀ IU/ml, with
763 disproportionately low percentage genome coverage. Repeat VL was 3 log₁₀ IU/ml.

764

765

766

767

Genotype	Capture - Illumina		HEP-TILE - Nanopore		HEP-TILE - Illumina	
	No. mismatches	No. N's	No. mismatches	No. N's	No. mismatches	No. N's
A	0	0	0	0	0	0
	0	0	0	0	0	0
B	0	0	0	0	0	0
	0	0	0	0	0	0
C	0	0	0	0	0	0
	0	0	0	0	0	0
D	0	0	0	0	2	311*
	0	0	0	0	0	0
E	0	0	0	0	1	0
	0	0	0	0	0	0
F	0	0	0	0	77	0
	0	0	0	0	0	0
G	0	0	0	0	0	0
	0	0	0	0	0	0
H	0	0	0	0	0	0
	0	0	0	0	0	0
I	0	0	0	0	0	0
	0	0	0	0	0	0
J	0	0	0	0	1	0
	0	0	0	0	0	0
Median	0	0	0	0	0	0
Percent accuracy	100%		100%		99.98%	

768

769

770

771

Table 1: Comparison of sequencing errors when sequencing mock samples using three workflows.

Each mock sample was sequenced in duplicate, one replicate per row.

* 311 base dropout of a continuous regio

772

	Capture - Illumina	HEP-TILE - Nanopore	HEP-TILE - Illumina
Time required for enrichment and library prep	1.5 days (capture overnight)	1.5 days (PCR overnight)	1.5 days (PCR overnight)
Sensitivity (HBV DNA VL) for generating HBV WGS	> 5.5 log ₁₀ IU/ml	> 1.5 log ₁₀ IU/ml	> 1.5 log ₁₀ IU/ml
Consensus sequence accuracy*	100%	100%	99.98%
Cost per sample** (enrichment & library prep reagents)	£29	£17	£25
Cost per sample*** (sequencing)	£25 E.g. multiplexed on partial lane on Novaseq X, 1 million reads per sample	£5.60 E.g. 96 samples multiplexed per flow cell, bulk purchasing 12 flow cells, 100,000 reads per sample	£2.50 E.g. multiplexed on partial lane on Novaseq X, 100,000 reads per sample
Throughput	Platform dependent e.g. Miseq vs. Novaseq X	Platform dependent e.g. Minion vs. Promethion	Platform dependent e.g. Miseq vs. Novaseq X
Portability	No	Yes with MinION	No

773

774

Table 2: Summary of three HBV WGS pipelines, comparing time, cost and outputs.

775

All costs are based on cost of reagents in the UK in 2024 (GBP) and do not include staff or equipment costs.

776

*based on sequencing of mock samples with known sequence

777

** reagent costs for enrichment and library prep reagents calculated in GBP for first 1000 samples. Cost excludes extraction and QC (Qubit/Bioanalyser reagents).

778

779

*** NB price extremely variable - depends on platform used and whether performing in-house or outsourcing commercially. Excludes hardware (i.e. Laptop/MinION for Nanopore, and Novaseq/Miseq for Illumina)

780

781

782

783

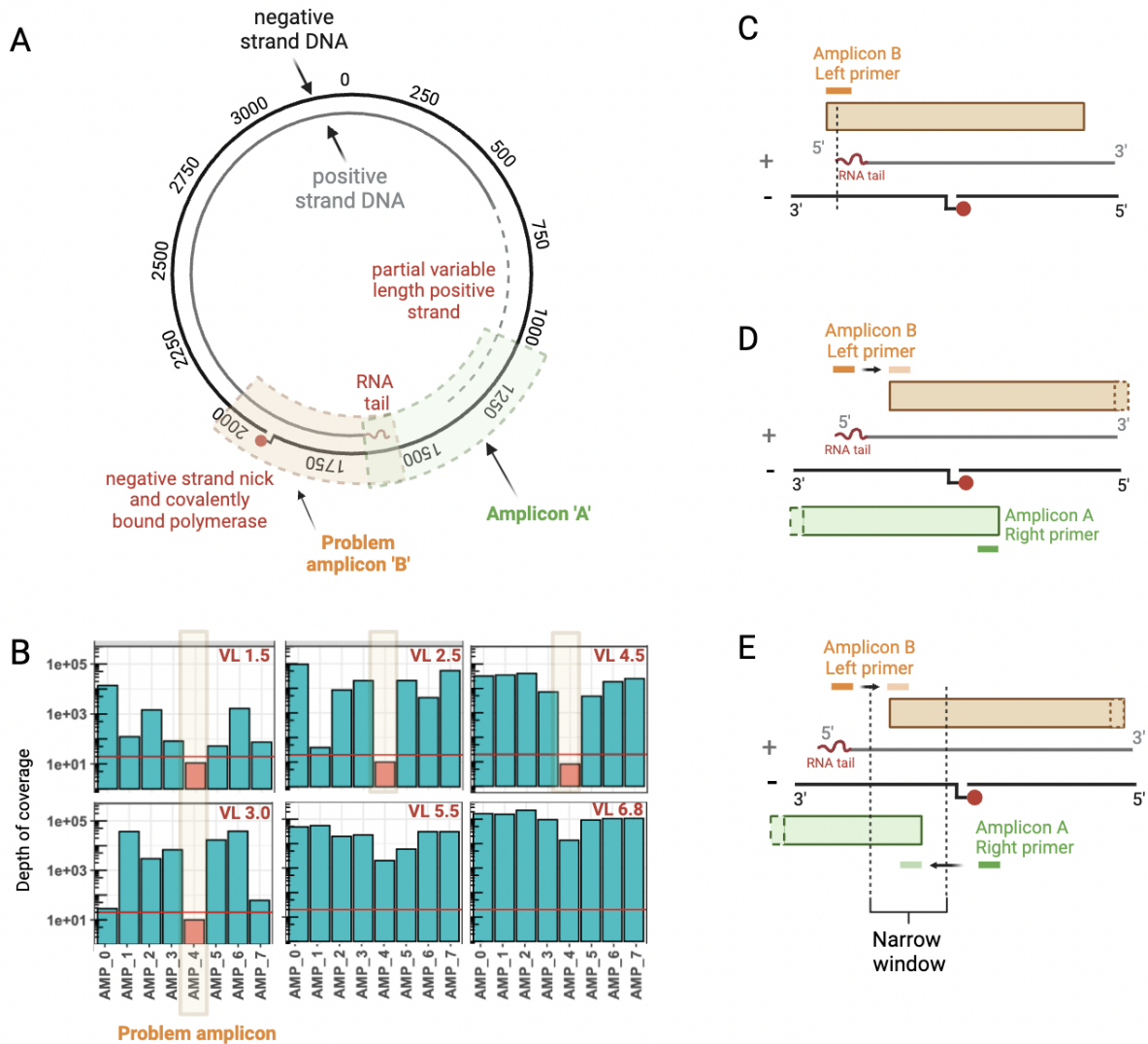
784

785

786

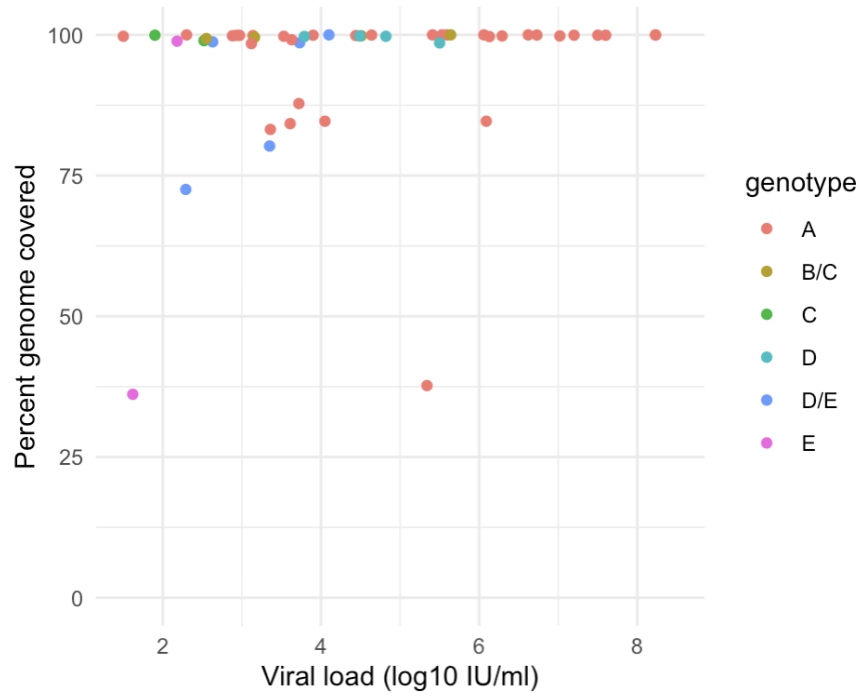
787

788 **Supplementary material**



789
 790 **Supplementary figure 1: Modifications to primer positions to take into account HBV genome**
 791 **structure.** (A) Schematic of HBV relaxed circular DNA (rc-DNA), the HBV partially dsDNA genome has 5'
 792 modifications to both the positive DNA strand (RNA tail) and negative DNA strand (covalently bound
 793 polymerase). The negative strand is complete but not fully circular due to the presence of a 'nick' and
 794 covalently bound polymerase. The partial positive strand is variable in length. (B) Depth of coverage for
 795 amplicon scheme hbv/500/v1.1.0 - illustrating drop-out of the 'problem amplicon' (initially Amplicon 4) in
 796 samples below a VL of 5 log₁₀ IU/ml (VL shown in red). (C) Amplicon B drops out in low-medium viral load
 797 samples, the hypothesis is that neither the positive nor negative strands of the HBV relaxed circular genome
 798 are complete in that region. The left hand primer only partially overlaps the 5' tail of the positive strand, the
 799 negative strand contains a nick. (D) Attempt to solve issue by moving the amplicon B primer upstream,
 800 however this has the knock on effect of the Amplicon A right hand primer moving upstream, which now
 801 means amplicon A falls across a non-continuous DNA region. (E) Solution requires the left hand primer for
 802 amplicon B and right hand primer for amplicon A to be placed in a narrow 200bp window, to ensure
 803 Amplicon B can be generated from the positive DNA strand, and amplicon A from the negative DNA strand.
 804 Created in BioRender. Lumley, S. (2024) BioRender.com/l10u171

805



806
807
808
809
810
811
812
813
814

Supplementary figure 2: Relationship between HBV genotype and performance of HEP-TILE Nanopore workflow. Percentage genome coverage shown at x20 depth.

815
816
817

Genotype	Reference genome	No enrichment - Illumina sequencing					Capture Illumina					HEP-TILE nanopore					HEP-TILE Illumina							
		Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage	Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x5)	Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x20)	Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x5)
A	KU736919	2564112	10405	4058	0.41	100	7703618	6846222	901683	90.17	100	100	95199	94946	997342	99.73	100	100	4779606	4517282	945116	94.51	100	100
	not published	2135330	7887	3694	0.37	100	9257976	8358193	902810	90.28	100	100	100155	100099	999441	99.94	100	100	7560102	7056345	933366	93.34	100	100
B		2545038	330	130	0.01	0	511478	149377	292050	29.2	100	100	79709	79284	994668	99.47	100	100	5013574	4923003	981935	98.19	100	100
		2030784	314	155	0.02	0	551530	172912	313513	31.35	100	100	53460	53283	996689	99.67	100	100	4757802	4633475	973869	97.39	100	100
C	MT111596	0	0	0	0.00	0	727296	309195	425130	42.51	100	100	67502	67463	999422	99.94	100	100	3920376	3773438	962519	96.25	100	100
		2819740	449	159	0.02	0	667198	270331	405174	40.52	100	100	50962	50937	999509	99.95	100	100	4546038	4386456	964937	96.49	100	100
D	NC_003977	2594796	3433	1323	0.13	100	2359894	1997417	846401	84.64	100	100	777	777	1000000	100.00	100	100	7559004	7386456	977173	97.72	100	100
		1787174	2604	1457	0.15	100	2365408	1999929	845490	84.55	100	100	151565	150757	994669	99.47	100	100	4595728	4581007	996797	99.68	90	90
E	KU736895	2456130	42317	17229	1.72	100	17116198	16524062	965405	96.54	100	100	2310	1363	590043	59.00	100	100	7866442	7537133	958137	95.81	100	100
		2050518	36693	17895	1.79	100	17556920	16941576	964951	96.5	100	100	174563	174504	999662	99.97	100	100	7175694	6915732	963772	96.38	100	100
F	DQ623095	2446438	1418	580	0.06	98.48	1069550	776761	726250	72.63	100	100	64275	64029	996173	99.62	100	100	4011426	3770749	940002	94	100	100
		1355114	750	553	0.06	83.05	981698	671782	684306	68.43	100	100	68247	68081	997568	99.76	100	100	4544664	4392756	966574	96.66	100	100
G	AB056513	2390922	357	149	0.01	0	605598	226835	374570	37.46	100	100	60886	60813	998801	99.88	100	100	4868602	4790237	963702	96.37	100	100
		2030368	334	160	0.02	0	708842	290037	409170	40.92	100	100	78879	78761	998504	99.85	100	100	4111952	4046170	984728	98.47	100	100
H	FJ356716	2382670	4789	2010	0.20	100	3247752	2906583	894952	89.5	100	100	83243	82928	996216	99.62	100	100	5613544	5488416	977710	97.77	100	100
		1250400	2260	1807	0.18	100	4593134	4154349	904469	90.45	100	100	95882	95846	999625	99.96	100	100	5122012	5002049	980132	98.01	100	100
I	AB562463	2476292	530	214	0.02	0	612856	270695	441694	44.17	100	100	255721	254408	994865	99.49	100	100	3159806	3103797	982275	98.23	100	100
		1342546	308	229	0.02	0	538454	238861	443605	44.36	100	100	289336	284791	984292	98.43	100	100	4547926	4498098	989264	98.93	100	100
J	AB486012	2729238	357	131	0.01	0	470730	147289	312895	31.29	100	100	83559	81982	981127	98.11	100	100	3057900	3038911	993136	99.31	100	100
		1779390	248	139	0.01	0	474726	129406	272591	27.26	100	100	553110	549429	993345	99.33	100	100	3148202	3092929	982131	98.21	100	100
Median		2259000	640	391	0.04	41.525	854497	490489	563956	56.40	100	100	81476	80633	997015.76	99.70	100	100	4676765	4549145	977441	97.75	100	100

818
819
820
821

Supplementary table 1: Sequencing metrics for mock HBV samples for capture-Illumina, HEP-TILE Nanopore and HEP-TILE Illumina workflows vs. no enrichment

828

Genotype	Capture (2019 HBVdb download)		HEP-TILE (2024 HBVdb download)	
	Genomes in HBVdb	Genomes used	Genomes in HBVdb	Downsampled Genomes
A	506	506	1066	559
B	1218	1218	2018	1143
C	1447	1447	3095	1764
D	823	823	1523	920
E	254	254	397	254
F	197	197	304	156
G	28	28	50	18
H	26	26	28	17

829

830

**Supplementary table 3 - Number of sequences downloaded from publicly available
sequence database HBVdb[20]**

831

832

833

834

835

836 **Supplementary table 4 - HBV probe sequences**

837
838 >probe_1
839 CTC AACACAGTTCCACCAAGCTCTGCAAGATCCGAGAGTAAGGGCCTGTATTTTCTGCTGGTGGCTCCAGTT
840 CAGAAACAGAGAACCCTGTTCCGACTATTGCCTCTCTCACATCATC
841
842 >probe_2
843 TGGTGGCTCCAGTTCAGAAACAGAGAACCCTGTTCCGACTATTGCCTCTCTCACATCATCAATCTTCTCGAAGAC
844 TGGGGACCCTGCAATGAACATGGAGAACATCACATCAGGACTCCT
845
846 >probe_3
847 AATCTTCTCGAAGACTGGGGACCCTGCAATGAACATGGAGAACATCACATCAGGACTCCTAGGACCCCTGCTCG
848 TGTTACAGGCGGGGTGTTTCTTGTTGACAAAAATCCTCACAATACC
849
850 >probe_4
851 AGGACCCCTGCTCGTGTTACAGGCGGGGTGTTTCTTGTTGACAAAAATCCTCACAATACCGCAGAGTCTAGACT
852 CGTGGTGGACTTCTCTCAATTTTCTAGGGGGACCACCCGGGTGTCC
853
854 >probe_5
855 GCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGACCACCCGGGTGTCCTGGCCAAAATTCGC
856 AGTCCCCAACCTCCAATCACTCACCAACCTCCTGTCTCCAACCTG
857
858 >probe_6
859 TGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCCTGTCTCCAACCTGTCCTGGCTATCGCTG
860 GATGTGTCTGCGGCGTTTTATCATCTTCTTCTCATCCTGCTGCT
861
862 >probe_7
863 TCCTGGCTATCGCTGGATGTGTCTGCGGCGTTTTATCATCTTCTTCTCATCCTGCTGCTATGCCTCATCTTCTTG
864 TTGGTTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCTCT
865
866 >probe_8
867 ATGCCTCATCTTCTTGTGGTTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCTCTAATTCCAGGATCCACA
868 ACCACCAGCACGGGACCCTGCAAAACCTGCACGACTCCTGCTCA
869
870 >probe_9
871 AATTCCAGGATCCACAACCACCAGCACGGGACCCTGCAAAACCTGCACGACTCCTGCTCAAGGAACCTCTATGT
872 TTCCCTCATGTTGCTGTACAAAACCTTCGGACGGAAATTGCACCTG
873
874 >probe_10
875 AGGAACCTCTATGTTTCCCTCATGTTGCTGTACAAAACCTTCGGACGGAAATTGCACCTGTATCCCATCCCATC
876 ATCTTGGGCTTTTCGGAAAATACCTATGGGAGTGGGCCTCAGCCCG
877
878 >probe_11
879 TATCCCATCCCATCATCTTGGGCTTTTCGGAAAATACCTATGGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAG
880 TTTACTAGTGCAATTTGTTTCACTGGTGCCTAGGGCTTTCCCCAC
881
882 >probe_12
883 TTTCTCCTGGCTCAGTTTACTAGTGCAATTTGTTCACTGGTGCCTAGGGCTTTCCCCACTGTCTGGCTTTTCACT
884 TATATGGATGATGTGGTATTGGGGGCCAAATCTGTGCAGCATCTT
885
886 >probe_13
887 TGTCTGGCTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAATCTGTGCAGCATCTTGTGCTTATACC
888 GCTGTTACCAATTTTCTGTTGTCTGTGGGTATACATTTAAACACT
889

890 >probe_14
891 GAGTCCCTTTATACCGCTGTTACCAATTTTCTGTTGTCTGTGGGTATACATTTAAACACTAACAAAACAAAAAGAT
892 GGGGTTATCCCTAAAATTCATGGGTTATGTAATTGGAAGTTGG
893
894 >probe_15
895 AACAAAACAAAAAGATGGGGTTATCCCTAAATTCATGGGTTATGTAATTGGAAGTTGGGGGACATTGCCACAA
896 GAACATATTGTACAAAAAATCAAAGAATGTTTTCGAAAACCTCCT
897
898 >probe_16
899 GGGACATTGCCACAAGAACATATTGTACAAAAAATCAAAGAATGTTTTCGAAAACCTCCTGTTAATAGGCCTATTG
900 ATTGGAAAGTCTGTCAACGAATTGTGGGTCTTTTGGGCTTTGCT
901
902 >probe_17
903 GTTAATAGGCCTATTGATTGAAAAGTCTGTCAACGAATTGTGGGTCTTTTGGGCTTTGCTGCCCTTTTACCCAAT
904 GTGGTTATCCTGCTCTAATGCCTTTGTATGCATGTATTAAGCT
905
906 >probe_18
907 GCCCCTTTTACCCAATGTGGTTATCCTGCTCTAATGCCTTTGTATGCATGTATTAAGCTAAGCAGGCTTTTGGCTT
908 TCTCGCCAACCTACAAGGCCTTTCTCTGTAAACAATACATGAAC
909
910 >probe_19
911 AAGCAGGCTTTTGGCTTTCTCGCCAACCTACAAGGCCTTTCTCTGTAAACAATACATGAACCTTTACCCCGTTGCTC
912 GGCAACGGCCAGGCTGTGCCAAGTGTGGCTGACGCAACCCCTT
913
914 >probe_20
915 CTTTACCCCGTTGCTCGGCAACGGCCAGGCCTGTGCCAAGTGTGGCTGACGCAACCCCACTGGCTGGGGCT
916 TGGCCATTGGCCATCAGCGCATGCGTGGAACCTTTGTGGCTCCTCTG
917
918 >probe_21
919 ACTGGCTGGGGCTTGGCCATTGGCCATCAGCGCATGCGTGGAACCTTTGTGGCTCCTCTGCCGATCCATACTGC
920 GGAACCTCTAGCAGCTTGTTCGCTCGCAGCCGGTCTGGAGCGAAA
921
922 >probe_22
923 CCGATCCATACTGCGGAACCTCCTAGCAGCTTGTTCGCTCGCAGCCGGTCTGGAGCGAACTTATCGGGACTGA
924 CAACTCTGTTGTCTCTCTCGGAAGTACACCTCCTTTCCATGGCTG
925
926 >probe_23
927 CTTATCGGGACTGACAACTCTGTTGTCCTCTCTCGGAAGTACACCTCCTTTCCATGGCTGCTAGGCTGTGCTGCC
928 AACTGGATCCTGCGCGGGACGTCCTTTGTTTACGTCCCCTCGGCG
929
930 >probe_24
931 CTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTACGTCCCCTCGGCGCTGAATCCCGCGG
932 ACGACCCCTCTCGGGGTCGCTTGGGGCTCTATCGCCCCCTTCTCCGT
933
934 >probe_25
935 CTGAATCCCGCGGACGACCCCTCTCGGGGTCGCTTGGGGCTCTATCGCCCCCTTCTCCGTCTGCCGTTCCGGC
936 CGACGACGGGGCGCACCTCTTTACGCGGACTCCCCGTCTGTGCCT
937
938 >probe_26
939 CTGCCGTTCCGGCCGACGACGGGGCGCACCTCTTTACGCGGACTCCCCGTCTGTGCCTTCTCATCTGCCGG
940 ACCGTGTGCACTTCGCTTACCTCTGCACGTCGCATGGAGACCACCG
941
942 >probe_27

943 TCTCATCTGCCGGACCGTGTGCACTTCGCTTACCTCTGCACGTGCGATGGAGACCACCGTGAACGCCCTCAG
944 AGCTTGCCAAAGGTCTTACATAAGAGGACTCTTGGACTTTCAGCAA
945
946 >probe_28
947 TGAACGCCCTCAGAGCTTGCCAAAGGTCTTACATAAGAGGACTCTTGGACTTTCAGCAAGGTCAACGACCGGG
948 ATTGAGGAATACATCAAAGACTGTGTATTTAAGGACTGGGAGGAGT
949
950 >probe_29
951 GGTCAACGACCGGGATTGAGGAATACATCAAAGACTGTGTATTTAAGGACTGGGAGGAGTTGGGGGAGGAGAT
952 TAGGTTAAAGGTCTTTGTATTAGGAGGCTGTAGGCATAAATTGGTCT
953
954 >probe_30
955 TGGGGGAGGAGATTAGGTTAAAGGTCTTTGTATTAGGAGGCTGTAGGCATAAATTGGTCTGTGCACCAGCACCA
956 TGCAACTTTTTACCTCTGCCTAATCATCTCTTGTTCATGTCCTAC
957
958 >probe_31
959 GTGCACCAGCACCATGCAACTTTTTACCTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAG
960 CTGTGCCTTGGGTGGCTTTGGGGCATGGATAGAACAACCTTTGCC
961
962 >probe_32
963 TGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGATAGAACAACCTTTGCCATATGGCCTTTTTGG
964 CTTAGACATTGACCCTTATAAAGAATTTGGAGCTTCTGTGGAGTT
965
966 >probe_33
967 ATATGGCCTTTTTGGCTTAGACATTGACCCTTATAAAGAATTTGGAGCTTCTGTGGAGTTACTCTCGTTTTTGCT
968 TCTGACTTCTTCCCGTCTGTTTCGGGACCTACTCGACACCGCTT
969
970 >probe_34
971 ACTCTCGTTTTTGCTTCTGACTTCTTCCCGTCTGTTTCGGGACCTACTCGACACCGCTTACGCCCTGTACCGGGA
972 TGCTTAGAGTCACCTGAACATTGCACACCTAACCATACAGCACT
973
974 >probe_35
975 AGCCCTGTACCGGGATGCCTTAGAGTCACCTGAACATTGCACACCTAACCATACAGCACTCAGGCAAGCTATTC
976 TGTGCTGGGGTGAATGACTCTGGCTACCTGGGTGGGCAATA
977
978 >probe_36
979 CAGGCAAGCTATTCTGTGCTGGGGTGAGTTAATGACTCTGGCTACCTGGGTGGGCAATAATTTGGAAGATCCAG
980 CAGCCAGGATTTAGTAGTCAATTATGTCAATACTAATATGGGCCT
981
982 >probe_37
983 TTTGGAAGATCCAGCAGCCAGGGATTTAGTAGTCAATTATGTCAATACTAATATGGGCCTAAAAATCAGACA
984 TTGTGGTTTACATTTCTGCCTTACTTTTGAAGAGAACTGT
985
986 >probe_38
987 AAAAATCAGACA
988 TCTTTTGGAGTGTGGATTGCACTCCTCCTGCTTATAGACCACC
989
990 >probe_39
991 TCTTGAGTATTTGGTGTCTTTTGGAGTGTGGATTGCACTCCTCCTGCTTATAGACCACCAAATGCCCTATCCTA
992 TCAACACTTCCGGAACTACTGTTGTTAGACGACGGGACCGAGG
993
994 >probe_40
995 AAATGCCCTATCCTATCAACACTTCCGGAACTACTGTTGTTAGACGACGGGACCGAGGCAGGTCCCCTAGAA
996 GAAGAACTCCCTCGCCTCGCAGACGAAGATCTCAATCGCCGCGTCG

997
998 >probe_41
999 CAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGCAGACGAAGATCTCAATCGCCGCGTGCAGAAAGATCTCAAT
1000 CTCCAGCTTCCCAATGTTAGTATTCCCTTGACTCATAAGGTGGAA
1001
1002 >probe_42
1003 CAGAAGATCTCAATCTCCAGCTTCCCAATGTTAGTATTCCCTTGACTCATAAGGTGGGAACTTTACGGGGCTTT
1004 ATTCTTCTACTGTACCTGTCTTTAATCCTGACTGGCAAACCTCCTT
1005
1006 >probe_43
1007 ACTTTACGGGGCTTTATTCTTCTACTGTACCTGTCTTTAATCCTGACTGGCAAACCTCCTTCTTTCTGATATTCAT
1008 TTGCATCAAGACATGATAAAATAAATGTGAACAATTTGTAGGCC
1009
1010 >probe_44
1011 CTTTTCTGATATTCATTTGCATCAAGACATGATAAAATAAATGTGAACAATTTGTAGGCCCTCTCACAGAAAATGA
1012 AAAGAGAAGATTAATAAATTATGCCAGCTAGATTTTATCCTA
1013
1014 >probe_45
1015 CTCTCACAGAAAATGAAAAGAGAAGATTAATAAATTATGCCAGCTAGATTTTATCCTAACGTTACCAATATTTG
1016 CCTTTGGATAAAGGTATTAACCTTATTATCCAGAGAATGTAG
1017
1018 >probe_46
1019 ACGTTACCAAATATTTGCCTTTGGATAAAGGTATTAACCTTATTATCCAGAGAATGTAGTTAATCATTACTTCAAA
1020 ACCAGACATTATTTACATACTCTATGGAAGGCGGGAATTCTAT
1021
1022 >probe_47
1023 TTAATCATTACTTCAAAACCAGACATTATTTACATACTCTATGGAAGGCGGGAATTCTATATAAGAGAGAAACCAC
1024 ACGTAGCGCCTCATTTTGTGGGTCAACATATTCTTGGGAACAAG
1025
1026 >probe_48
1027 ATAAGAGAGAAACCACACGTAGCGCCTCATTTTGTGGGTCAACATATTCTTGGGAACAAGAGCTACAGCATGGG
1028 AGCACCTCGGTCAACGCCTCGAAGGGGCATGGGAAAGAATCTTTCT
1029
1030 >probe_49
1031 AGCTACAGCATGGGAGCACCTCGGTCAACGCCTCGAAGGGGCATGGGAAAGAATCTTTCTGTGCCCAATCCTCT
1032 GGGATTCTTTCCAGACCACCAGTTGGATCCAGCATTCCAGAGCAAAT
1033
1034 >probe_50
1035 GTGCCCAATCCTCTGGGATTCTTTCCAGACCACCAGTTGGATCCAGCATTCCAGAGCAAATCCAGCAGTCCAGAT
1036 TGGGACTTCAACACAAACAAGGACAATTGGCCAGAGGCAAACAAG
1037
1038 >probe_51
1039 TCCAGCAGTCCAGATTGGGACTTCAACACAAACAAGGACAATTGGCCAGAGGCAAACAAGGTAGGAGTGGGAG
1040 GCTTCGGTCCAGGGTTCACACCCCCACACGGAGGCTTCTGGGGTGG
1041
1042 >probe_52
1043 GTAGGAGTGGGAGGCTTCGGTCCAGGGTTCACACCCCCACACGGAGGCTTCTGGGGTGGAGCCCTCAGGCA
1044 CAGGGCATACTAACAACCTTGCCAGCAGATCCGCCTCCTGCCTCCACC
1045
1046 >probe_53
1047 AGCCCTCAGGCACAGGGCATACTAACAACCTTGCCAGCAGATCCGCCTCCTGCCTCCACCAATCGGCGGTGAG
1048 GAAGGAAGCCAACCCAGTCTCTCCACCTTAAGAGACACTCATCCA
1049
1050 >probe_54

1051 AATCGGCGGTCAGGAAGGAAGCCAACCCAGTCTCTCCACCTCTAAGAGACTCATCCACAGGCCATGCAGT
1052 GGAACTCAACACAGTTCCACCAAGCTCTGCAAGATCCGAGAGTAAGG
1053
1054 >probe_55
1055 CAGGCCATGCAGTGGAACCTCAACACAGTTCCACCAAGCTCTGCAAGATCCGAGAGTAAGGGCCTGTATTTTCC
1056 TGCTGGTGGCTCCAGTTCAGAAACAGAGAACCCTGTTCCGACTATT
1057
1058 >Deletion_region1903_1939_removed probe
1059 TCTTGTTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTAT
1060 AAAGAATTTGGAGCTTCTGTGGAGTTACTCTCGTTTTTGCCTTCT
1061
1062 >Deletion_Genotype_D_region2897_2929_removed probe
1063 ACGTAGCGCATCATTTTGCGGGTCAACATATTCTTGGGAACAAGAGCTACAGCATGGGAGCGAATCTTTCTGTTC
1064 CCAACCCCTCTGGGATTCTTTCCCGATCATCAGTTGGACCCTGCAT
1065
1066 >H_cons_seq_s2035_e2154_ConS2035
1067 CGAACATTGCACCCCAACCACACTGCTCTCAGGCAAGCTATTTTGTGCTGGGGTGAGTTGATGACCTTGGCTT
1068 CCTGGGTGGGCAATAATTTAGAGGATCCTGCAGCAAGAGATCTAGT
1069
1070 >G_cons_seq_s2845_e2964_ConS2845
1071 TTTGTGGGTCACCATATACTTGGGAACAAGATCTACAGCATGGGGCTTTCTTGGACGGTCCCTCTCGAGTGGGG
1072 AAAGAACCTTTCCACCAGCAATCCTCTAGGATTCCTTCCCGATCAC
1073
1074 >G_cons_seq_s1651_e1770_ConS1651
1075 TATATAAGAGGACTCTTGGACTGTTTGTATATGTCACAACCGGGGTGGAGAAATACTTCAAGGACTGTGTTTTTG
1076 CTGAGTGGGAAGAATTAGGCAATGAGTCCAGGTTAATGACCTTTG
1077
1078 >E_cons_seq_s1975_e2094_ConS1975
1079 TTCAGTAAGAGATCTTCTAGATACCGCCTCAGCTCTGTATCGGGATGCCTTAGAATCTCCTGAGCATTGTTACC
1080 TCACCACACTGCACTCAGGCAAGCCATTCTTTGCTGGGGGGAACCT
1081
1082 >A_cons_seq_s391_e510_Target_seq391
1083 TTTTATCATATTCCTCTTCATCCTGCTGCTATGCCTTAGCTCTGTATCGGTTCTTCTGAAGCCTTAGAGTCTCCAG
1084 AGCATTGCTCCCCTCACCTCAGGATCCACAACAACCAAGTACGGG
1085
1086 >A_cons_seq_s2875_e2994_ConS2875
1087 CCTCGCAAAGGCATGGGGACGAATCTTCTGTTCCCAACCCTCTGGGATTCTTTCCCGATCATCAGTTGGACCCT
1088 GCATTCCGAGCCAACCTCAAACAATCCAGATTGGGACTTCAACCCCT
1089
1090 >Genotype_D_deletion_at_33_55
1091 CACTCATCCTCAGGCCATGCAGTGGAACCTCCACAACCTTCCACCAAACCTCTGCAAGATCCCCCTGCTGGTGGCT
1092 CCAGTTCAGGAACAGTAAACCCTGTTCCGACTACTGTCTCTCCCAT
1093
1094 >D_cons_seq_s2956_e3075_ConS2956
1095 AACAAGGACACCTGGCCAGACGCCAACAAGGTAGGAGCTGGAGCATTGGGGCTGGGATTACCCCCACCGCACC
1096 GAGGCCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATACTACAAACC
1097
1098 >B_cons_seq_s2095_e2214_ConS2095
1099 AATGAATCTAGCCACCTGGGTGGGAAGTAATTTGGAAGATCCAGCATCCAGGGAATTAGTAGTCAGCTATGTCAA
1100 CGTTAATATGGGCCTAAAAATCAGACAACCTATTGTGGTTTTCACAT
1101
1102 >B_cons_seq_s1711_e1830_Target_seq1711
1103 AAGACTGTTTTGTTTAAAGACTGGGAGGAGTTGGGTTTACTGATTGGGGAGTTGGGGGATGTTCTAAGGGGATGG
1104 AAGGCTAAATTGGTCTGTTCCACCAGCACCATGCAACTTTTTTCACCT

1105
1106 >C_cons_seq_s2990_e3109_ConS2990
1107 ACAAGGATCACTGGCCAGAGGCAAATCAGGTAGGAGCGGGAGCATTGCGGCCAGGGTTCACCCCACACACG
1108 GCGGTCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATATTGACAACAG
1109
1110 >C_cons_seq_s3080_e3199_ConS3080
1111 GCCCTCAGGCTCAGGGCATATTGACAACAGTGCCAGCAGCACCTCCTCCTGCCTCCACCAATCGGCAGTCAGG
1112 AAGACAGCCTACTCCCATCTCTCCACCTCTAAGAGACAGTCATCCTC
1113
1114 >C_cons_seq_s2629_e2748_ConS2629
1115 AATTAATTATGCCTGCTAGGTTCTATCCTAACCTTACCAAATATTTGCCCTTGACAAAGGCATTAACCTTATTAT
1116 CCTGAACATGCAGTTAATCATTACTTCAAACTAGGCATTATT
1117
1118 >C_cons_seq_s1651_e1770_ConS1651
1119 TACATAAGAGGACTCTTGGACTCTCAGCAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTTTGTTTA
1120 AAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAATGATCTTTG
1121
1122 >C_cons_seq_s3080_e3199_Target_seq3080
1123 GCCCTCAGGCTCAGGGGACATTGACACCTCCGGCAGCCGCCCTCCTCCTGCCTCCAGAAAACCTCCTTCAGG
1124 TTCTCCACCTACTCTAATTTCTTCTCCTCTAAGAGACAGTCATCCCA
1125
1126 >C_cons_seq_s3140_e3259_ConS3140
1127 AACCTCCTTCAGGTTCTCCACCTACTCTAATTTCTTCTCCTCTAAGAGACAGTCATCCCAACGCCATACACTGCCA
1128 CTCCACGCTATTCCACCAAGTTCTGCGAGATCCCTTAGTTGCGG
1129
1130 >C_cons_seq_s3200_e3319_ConS3200
1131 AGGCCATGCAGTGGAACCTCCACAACATTCCACCAAGCTCTGCTAGATCCCAGAGTGAGGGGCCTATACTTTCTCCT
1132 GCTGGTGGCTCCAGTTCCGGAACAGTAAACCCTGTTCCGACTACTG
1133
1134

1135 **Supplementary text**

1136

1137 **Determining HBV drug resistance and immune escape from sequences using HBV WGS**

1138 We were able to call polymorphisms at sites of previously reported resistance associated
1139 mutations (RAMs) and vaccine escape mutations (VEMs) as defined by geno2pheno[37].
1140 Polymorphisms were detected in RT and HBsAg regions. The polymorphism rt169M was
1141 identified in viral sequences from two individuals (906116N/907395N) and 202R in one (HEP-
1142 1763). The rt169 and rt202 are sites associated with HBV drug resistance however the mutations
1143 169M and 202R have unknown significance. In the HBsAg region, a number of mutations
1144 associated with HBsAg escape from vaccines and detection were identified (144E, 100C, 128V,
1145 133I and 134V, each in one individual). All samples with polymorphisms had only been sequenced
1146 with a single method, HEP-TILE Nanopore.