SUPPLEMENTARY MATERIALS & METHODS

DNA and RNA isolation from liver biopsy

Liver tissues, previously cryopreserved in 10% DMSO and 90% fetal bovine serum freezing media until DNA/RNA extraction, were firstly disrupted on ice in 600 µl of homogenization RLT buffer (provided in the extraction kit) by using the Tissue Ruptor homogenizer (Qiagen) at full speed for 30 seconds. After the homogenization step, the AllPrep DNA/RNA Mini kit procedure was followed in order to extract simultaneously both DNA and RNA from the same liver biopsy. In the last step, DNA and RNA were eluted each in 50 µL of elution buffer and quantified with the fluorometer (Qubit 2.0 Life Technologies, Invitrogen Division, Darmstadt, Germany) to verify the efficiency of DNA and RNA extraction. The quality of the extracted DNA and RNA was further verified by measuring 260/280 and 260/230 ratios by NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

It-HBV-DNA quantification

ItHBV-DNA was quantified by a modification of the commercial assay COBAS® TaqMan HBV test, v2.0 (Roche), opportunely adapted to quantify HBV-DNA extracted from liver biopsy, as previously described[1]

cccDNA quantification

cccDNA levels were quantified by applying an in-house real-time PCR assay running on Light-Cycler 2.0 (Roche) instrument. A pre-treatment with 10 U of Plasmid-Safe™ATP-Dependent DNase (Epicentre, USA) was used (37°C for 1 hour and 70°C for 30 min), in order to digest single-strand and linear double-strand HBV-DNA, according to the latest update of the international working group on cccDNA standardization (Allweiss et al., 2017 International HBV Meeting, O-45). Digestion was performed with 1 µL of DNase, 5 µL of 10x buffer, 2 µL of ATP 25mM, 37 µL of nuclease free water and 5 µL of intrahepatic total DNA. Digested samples were further purified by silica-membrane

Svicher V, *et al. Gut* 2021;0:1–12. doi: 10.1136/gutjnl-2020-323300

columns (QIAquick PCR purification kit, Qiagen) and re-suspended in 100 µL of elution buffer according to the manufacturer's instructions. The obtained 100 µL of digested DNA was dried and re-suspended in 10 µL of nuclease-free water.

For cccDNA quantification, ~~a~~ 20 µL of reaction mix was prepared with 7.7 µl of water PCR grade, 4 µl of MasterMix (5X), 1.3 µl of the probe [3µM], 1 µl of forward primer [10µM], 1 µl of reverse primer [10µM] and 5 µL of DNA, pretreated as previously described. The primers and the probe used for cccDNA quantification targets the HBV-DNA gap region comprised between DR1 and DR2 sequences: Forward, 5'-ACCTCTCTTTACGCGG-3'; Reverse, 5'-ACAGCTTGGAGGCTTGAA-3' and Probe, 5'-6FAM-CTCCCCGTCTGTGCCTTCTCATC-BHQ1-3'.

For cccDNA quantification, the thermal profile for Real-Time PCR was as follows: 1 denaturation cycle of 10 min at 95°C, 40 cycles of denaturation for 30s at 94°C, annealing for 1 min at 58°C and a final incubation of 10 min at 98°C. Serial dilutions of a plasmid containing an HBV monomer, pAM6 (ATCC® 45020D™ ATCC, Manassas, VA), were used to build the standard curve for cccDNA quantification (from 10^6 to 10^{-1} copies).

pgRNA quantification

In detail, 20 µL ddPCR reaction mix contained 5 µl of One-Step RT-ddPCR Advanced Kit for probes (Bio-Rad, Pleasanton, California, USA), 2 µL of Reverse Transcriptase, 1 µL of 300 mM DTT, 1 µL of 20x primers/probe mix, 6 µL of nuclease-free water and 5 µL of extracted RNA. Specific ad-hoc designed primers and a fluorescence hybridization probe targeting a conserved region of HBV Core were used: Forward, 5'-CTTTCGGAGTGTGGATTCG-3'; Reverse, 5'-CACCTTATGAGTCCAAGGAA-3', Probe, FAM-5'-TCCCTCGCCTCGCAGACGAA-FQ-3'. Reaction droplets were generated according to manufacturer's instructions by using the QX200TM Droplet Generator (Bio-Rad, Hercules, California, USA). pgRNA was amplified using T100TM

Thermal Cycler (Bio-Rad, Hercules, California, USA) under the following amplification profile: a reverse transcription step of 1 hour at 50°C, 1 denaturation cycle of 10 min at 95°C, 40 cycles of denaturation for 30 s at 94°C, annealing for 1 min at 56°C and a final incubation of 10 min at 98°C. After amplification, positive droplets were quantified by the QX200TM Droplet Reader (Bio-Rad, Hercules, California, USA) and analyzed by using the QuantaSoft^{TM} software (Bio-Rad, Hercules, California, USA).

Whole Exome Sequencing (WES)

WES was carried out by applying a well consolidated and next-generation sequencing approach, based on Illumina technology. In detail, 50ng of DNA, extracted form liver biopsies was used to build the sequencing library following the Nextera DNA Exome Kit procedure (Illumina), according to manufacturer's instructions. This procedure contains a crucial step for capturing the exome-amplified fragments by using streptavidin beads, binding biotinylated oligonucleotide probes, used for exome hybridization. This allows an in-depth analysis of the coding fraction of human genome and its flanking intronic/intergenic regions[2]. As a quality control measure, each WES library has been validated quantitatively by using the fluorometer Qubit (Life Technologies) while WES library pools were analyzed by both Qubit and Agilent 2200 TapeStation system. Finally, the obtained libraries were processed through paired end 150-bp read-length sequencing by the Nextseq 550 platform (Illumina Inc., SanDiego, CA).

Bioinformatic approach for the identification of HBV integrations into the human genome

To determine HBV integration sites, WES reads were analyzed using Virus-Clip[3]. Briefly, after a quality control step, performed by Trimmomatic[4], sequencing reads were aligned to virus specific reference genome (A=KY382410, B=KY470962, C=KY363287, D=KP322604, E=KU736913) by BWA-MEM[5] in order to extract all reads that contained HBV fragments (Fig. 1). For each patient, the alignment was genotypic specific since the reference HBV genotype infecting the patient was

used for alignment. Subsequently, a second software (SAMTools), able to identify human/viral breakpoints, was applied in order to infer all chimeric HBV/human sequences, representing the HBV integrations into human genome[6] (Fig. 1). Additionally, a further alignment against human reference genome (GRCh37/Hg19) was performed by BLAST for a further validation of the recognized integration sites. Lastly, we submitted the identified human portion involved in HBV integration events to the software ANNOVAR to obtain a refined mapping of HBV integrations at chromosome and gene level (Fig. 1)[7]. The functionality of genes involved in HBV integration was retrieved by querying three different online available databases: Gene Cards, Protein Atlas and KEGGS (Fig. 1)[8–10].

Quantification of HBV integrations by a droplet digital PCR approach (ddPCR)

ddPCR protocols were constructed by using each specific human-HBV chimeric sequence, revealed by whole exome sequencing, as reference sequence. In details, each ddPCR protocol was based on the use of two primers: the former matching the HBV region and the latter human region present in the human-HBV chimeric sequence. The probe was built in order to specifically recognize the region comprising HBV-human junctions. Details of primers and probes used for each ddPCR assay were reported in the table below. For each ddPCR assay, the target-specificity of all primers and probes was verified by Primer Blast software (NCBI), thus excluding any potential amplification of unspecific human regions.

Table. Primers and probes used for each ddPCR assay used to quantify HBV integrations

Pt No	Gene	Primer 1 (5'→3')	Primer 2 (5'→3')	Probe (5'→3')
39	ELAC2	*GGCATAAATTGGTCTGCGCC*	CAGGGGAGCTGAGTGTTGAG	GCACACAGGCCTCAGTCACGTGC
40	IFITM1	GAGAGGAGATGGTGAGGGGA	*GGAGACTCTAAGGCTTCCCG*	CCAGGATCCCCCAGCTGTTTTGCCT
40	IFITM1	*TGTCAACGACCGACCTTGAG*	GGAGGTGGTCCCTGATCTCA	AGACTGGGAGGAGTTGGGGGAGGAG
44	NUP85	GCCAAGTTGCCCCTTTATCT	*CCCGTAAAGTTTCCCACCTT*	GGGCCTCCAGAGAATTGACG
62	ANKRD52	*TGTCGGGTCTGGAAGTAATGA*	GCTCAATAGCCACAGCATCC	GCACATCGCCTGCTACCTGGGCC
62	ANKRD52	*CCGACCTTGAGGCCTACTTC*	GGCTACATGTGCTGGCTTCT	AGGACTGGGAGGAGTTGGGGGAGG
75	AGBL5	*ACTAGGAGGCTGTAGGCATAA*	GTTTGCCAGCCTCCTCCAT	GGTCTGCGCACCAGCACCATGCA
83	NR3C1	*ACGACCGACCTTGAGGCATA*	ATGATAGCTCTGTTCCAGACTCAA	GGAGGAGTTGGGGGAGGAGA
84	CYP2U1	*GACCGACCTTGAGGCATACT'*	AAACCGGAGGATTTCTACCC	TCCTTGGTCATCCAGAAATCGA

In the table the primer matching the HBV region is reported in italics while the primer matching the human region is reported as underlined

The linearity, detection sensitivity and specificity of each specific ddPCR assay were firstly verified by using serial dilutions of a synthetic DNA gene, corresponding to each specific chimeric HBV-human DNA sequence (length: 150 bp), identified by whole exome sequencing as positive controls, and by using proper negative controls.

For each ddPCR assay, ~~a~~ 20 µL of ddPCR reaction mix was prepared with 10 µl of 2xddPCR Supermix for probes (no dUTP) (Bio-Rad, Pleasanton, California, USA), 1µL of 20x primers/probe mix, 4 µL of nuclease-free water and 5 µL of extracted DNA. Reaction droplets were generated according to manufacturer's instructions by using the QX200TM Droplet Generator (Bio-Rad, Hercules, California, USA). The chimeric HBV-human regions were amplified by using T100TM Thermal Cycler (Bio-Rad, Hercules, California, USA) under the following amplification profile: 1 denaturation cycle of 10 min at 95°C, 40 cycles of denaturation for 30 s at 94°C, annealing for 1 min at 57°C and a final incubation of 10 min at 98°C. After amplification, positive droplets were quantified by the QX200TM Droplet Reader (Bio-Rad, Hercules, California, USA) and analyzed by using the QuantaSoftTM software (Bio-Rad, Hercules, California, USA). The number of positive droplets, representing the amplified chimeric HBV-human regions, were lastly normalized to the hepatocyte number, according to the quantification obtained by Albumin-based ddPCR copy number

assay (Alb, Human, Bio-Rad, Pleasanton, California, USA). This allowed to obtain the number of each specific HBV integration/1000 hepatocytes. Quantification of each HBV-DNA integration was performed at least in duplicate for each sample. Three negative controls and two positive controls (the above mentioned synthetic genes) were also included in each ddPCR reaction to verify the efficiency of amplification and to exclude samples' contamination.

The quantification of HBV DNA integrations by ddPCR was performed in samples where we had a sufficient amount of residual DNA to undertake the analysis. We prioritised those integration events within the exons and in close proximity to signal sequences for mRNA splicing as this is where the protein coding regions are located.

Whole transcriptome analysis

RNA library preparation was carried out using NEB Next Ultra RNA Library Prep kit (New England BioLabsInc), and ribosomal RNAs were removed from total RNA using Ribo-Zero Gold rRNA Removal kit, Illumina). RNA libraries were sequenced with an Illumina Hiseq 2000 (Illumina Inc., CA). Approximately 45–60 million paired-end 150 bp reads were obtained per sample. After a quality control step, performed by Trimmomatic, the expression of transcripts was quantified by Salmon[11], a software estimating the relative abundance of each annotated transcript for each sample in units of transcripts-per-million (TPM) an accurate and widely-used parameter to express transcript levels according to the following formula: $\mathrm{TMP} = \mathrm{A}\,\frac{1}{\sum(A)}\,10^6$, where A is $\frac{\text{total reads mapping to a specific gene} * 10^3}{gene\ length\ in\ bp}$ [12]. Statistically significant differences (adjusted P value) in the expression levels of all transcripts between groups were evaluated by DESeq[13].

SUPPLEMENTARY FIGURE LEGEND

Figure S1: Description of the patient cohort and of the performed virological analyses

For all 84 HBeAg-negative CHB patients, a liver biopsy and a matched serum sample were collected. Serum sample was used to quantify HBcrAg, HBsAg and for HBV genotyping. Liver biopsy was used for cccDNA, itHBV-DNA and pgRNA quantification and for whole exome sequencing (N=40) and whole transcriptome analysis (N=15).

REFERENCES

[1] Salpini R, Piermatteo L, Gill U, Battisti A, Stazi F, Guenci T, et al. Quantification of intrahepatic total HBV DNA in liver biopsies of HBV-infected patients by a modified version of COBAS®Ampliprep/COBAS®TaqMan HBV test v2.0. Med Microbiol Immunol 2017;206:295–9. https://doi.org/10.1007/s00430-017-0504-3.

[2] Guo Y, Long J, He J, Li C-I, Cai Q, Shu X-O, et al. Exome sequencing generates high quality data in non-target regions. BMC Genomics 2012;13:194. https://doi.org/10.1186/1471-2164-13-194.

[3] Ho DWH, Sze KMF, Ng IOL. Virus-Clip: a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability. Oncotarget 2015;6:20959–63. https://doi.org/10.18632/oncotarget.4187.

[4] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114. https://doi.org/10.1093/BIOINFORMATICS/BTU170.

[5] Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 2010;26:589–95. https://doi.org/10.1093/bioinformatics/btp698.

[6] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.

[7] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164–e164. https://doi.org/10.1093/nar/gkq603.

[8] Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinforma 2016;54:1.30.1-1.30.33. https://doi.org/10.1002/cpbi.5.

[9] Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. Protein Sci 2018;27:233–44. https://doi.org/10.1002/pro.3307.

[10] Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016;44:D457-62. https://doi.org/10.1093/nar/gkv1070.

[11] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14:417–9. https://doi.org/10.1038/nmeth.4197.

[12] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621–8. https://doi.org/10.1038/nmeth.1226.

[13] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.

Supplementary Table 1. Diagnostic performance of serological markers in identifying a limited intrahepatic reservoir

Serological markers predictive of cccDNA<1.5 log copies/1000cells	PPV	NPV	Diagnostic accuracy
Serum HBV-DNA<20,000 IU/ml	40%	86.8%	61.4%
Serum HBV-DNA<20,000 IU/ml and HBsAg<1,000 IU/ml	90%	81%	82.1%
Serum HBV-DNA <20,000 IU/ml and HBcrAg<3.0 log U/ml	51.7%	90.2%	74.3%
Serum HBV-DNA <20,000 IU/ml, HBcrAg<3.0 log U/ml and HBsAg<1,000 IU/ml	90%	83.3%	84.3%

Abbreviations: PPV, positive predictive value; NPV, negative predictive value

Supplementary Table 2: Characteristics of patients used for whole exome sequencing

Characteristics	Overall population
No. Cases	40
Age, (years) median (IQR)	35 (29- 42)
Male, n (%)	32 (80.0%)
Ethnicity, n (%):	
European	11 (27.5%)
Asian	20 (50.0%)
African	9 (22.5%)
Serum HBV-DNA, log IU/mL (median, IQR)[b]	3.9 (3.4– 5.3)
Serum ALT, log IU/dl (median, IQR) [b]	42 (28-72)
HBsAg, log IU/ml (median, IQR)	3.6 (3.1-4.0)
HBV genotype, n (%):	
A	2 (5%)
B	2 (5%)
C	9 (22.5%)
D	18 (45.0%)
E	9 (22.5%)
Ishak Fibrosis Stage, n (%):	
0-2	31 (77.5%)
3-6	9 (22.5%)