



OPEN ACCESS

Original research

Whole exome HBV DNA integration is independent of the intrahepatic HBV reservoir in HBeAg-negative chronic hepatitis B

Valentina Svicher,¹ Romina Salpini,¹ Lorenzo Piermatteo,¹ Luca Carioti,¹ Arianna Battisti,^{1,2} Luna Colagrossi,^{1,3} Rossana Scutari,¹ Matteo Surdo,⁴ Valeria Cacciafesta,⁴ Andrea Nuccitelli,⁴ Navjyot Hansi,² Francesca Ceccherini Silberstein,¹ Carlo Federico Perno,⁵ Upkar S Gill,² Patrick T F Kennedy

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2020-323300>).

For numbered affiliations see end of article.

Correspondence to

Dr Patrick T F Kennedy, Barts Liver Centre, Immunobiology, Blizard Institute, Barts and The London School of Medicine & Dentistry, Queen Mary University of London, Newark Street, London, E1 2AT, UK; p.kennedy@qmul.ac.uk

VS and RS contributed equally.

Received 11 October 2020

Revised 2 November 2020

Accepted 16 November 2020

Published Online First

21 December 2020

ABSTRACT

Objective The involvement of HBV DNA integration in promoting hepatocarcinogenesis and the extent to which the intrahepatic HBV reservoir modulates liver disease progression remains poorly understood. We examined the intrahepatic HBV reservoir, the occurrence of HBV DNA integration and its impact on the hepatocyte transcriptome in hepatitis B 'e' antigen (HBeAg)-negative chronic hepatitis B (CHB).

Design Liver tissue from 84 HBeAg-negative patients with CHB with low (n=12), moderate (n=25) and high (n=47) serum HBV DNA was analysed. Covalently closed circular DNA (cccDNA), pregenomic RNA (pgRNA) were evaluated by quantitative PCR, whole exome and transcriptome sequencing was performed by Illumina, and the burden of HBV DNA integrations was evaluated by digital droplet PCR.

Results Patients with low and moderate serum HBV DNA displayed comparable intrahepatic cccDNA and pgRNA, significantly lower than in patients with high HBV DNA, while hepatitis B core-related antigen correlated strongly with the intrahepatic HBV reservoir, reflecting cccDNA quantity. Whole exome integration was detected in a significant number of patients (55.6%, 14.3% and 25% in high, moderate and low viraemic patients, respectively), at a frequency ranging from 0.5 to 157 integrations/1000 hepatocytes. Hepatitis B surface antigen >5000 IU/mL predicted integration within the exome and these integrations localised in genes involved in hepatocarcinogenesis, regulation of lipid/drug metabolism and antiviral/inflammatory responses. Transcript levels of specific genes, including the proto-oncogene HRAS, were higher in patients with HBV DNA integration, supporting an underlying oncogenic risk in patients with low-level to moderate-level viraemia.

Conclusions HBV DNA integration occurs across all HBeAg-negative patients with CHB, including those with a limited HBV reservoir; localising in genes involved in carcinogenesis and altering the hepatocyte transcriptome.

INTRODUCTION

Chronic hepatitis B (CHB) virus infection is a dynamic disease reflecting the balance between the

Significance of this study

What is already known on this subject?

- Hepatitis B 'e' antigen (HBeAg)-negative phase of HBV infection is associated with a wide disease spectrum, ranging from quiescent low viraemic disease to chronic HBeAg-negative hepatitis, with a high risk of evolution to cirrhosis and hepatocellular carcinoma.
- Hepatitis B surface antigen (HBsAg) and hepatitis B core-related antigen (HBcrAg) are increasingly used as surrogates of the intrahepatic HBV reservoir.
- HBV DNA integration is believed to play a key role in hepatocarcinogenesis.

What are the new findings?

- HBV DNA integration in the whole exome is detected in a significant proportion of HBeAg-negative patients, including low viraemic patients with a limited HBV reservoir.
- HBV DNA integrations occur in regions crucial for human gene expression and involve genes regulating cell proliferation, carcinogenesis in addition to antiviral immunity and hepatocyte metabolism.
- High levels of HBsAg (>5000 IU/mL) can predict HBV DNA integration events in HBeAg-negative patients.

How might it impact on clinical practice in the foreseeable future?

- The demonstration of HBV DNA integration in patients with chronic hepatitis B perceived as low risk mandates a re-evaluation of treatment candidacy.
- Future studies will have to dissect the role of HBV DNA integration in hepatocarcinogenesis and how this can be mitigated against.
- Novel HBV therapies should target both integrated HBV DNA in addition to covalently closed circular DNA to enhance treatment outcomes and reduce hepatocellular carcinoma risk.



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Svicher V, Salpini R, Piermatteo L, et al. Gut 2021;70:2337–2348.

virus itself and the host immune response. In CHB, the hepatitis B 'e' antigen (HBeAg)-negative phase is associated with a spectrum of clinical outcomes spanning from HBeAg-negative chronic infection (serum HBV DNA <2000IU/mL and normal alanine aminotransferase (ALT)) to HBeAg-negative chronic hepatitis characterised by a higher risk of disease progression, the development of cirrhosis and hepatocellular carcinoma (HCC).¹

There remains a paucity of data on the extent and productivity of the intrahepatic HBV reservoir (in terms of covalently closed circular DNA (cccDNA), intrahepatic total HBV DNA (itHBV DNA) and pregenomic RNA (pgRNA)) in the HBeAg-negative phase of CHB. Unravelling this issue is critical since the intrahepatic HBV reservoir plays a pivotal role in modulating liver disease progression.^{2,3} Elucidating the accuracy of novel non-invasive HBV biomarkers in predicting the intrahepatic HBV reservoir also remains an unmet clinical need.^{4,5} HBV replication can promote HBV DNA integration into the genome of the human hepatocyte, considered a key mechanistic step underlying HBV-mediated carcinogenesis, even in the absence of necroinflammation.^{6–8} Indeed, HBV DNA integration can compromise cell genome stability and modify the expression of genes regulating cell cycle/proliferation, predisposing the hepatocytes to pre-neoplastic transformation.⁷

HBV DNA integration has been detected in the early phases of CHB virus infection.⁹ Importantly, a recent study from our group demonstrated high rates of HBV DNA integration along with clonal hepatocyte expansion in patients with HBeAg-positive chronic infection, highlighting a potential carcinogenetic risk in these patients, often considered to be in a benign disease phase, devoid of disease progression.⁹ A lower rate of HBV DNA integration has been observed in HBeAg-negative patients with CHB.¹⁰ Although controversial, it has been hypothesised that, during the immune clearance phase and HBeAg seroconversion, the development of a strong immune response may favour the selection of those hepatocytes in which HBV DNA integrations confer a survival advantage, thus potentially contributing to liver cancer development.^{8,11,12}

The occurrence and localisation of HBV DNA integration in the later phases of CHB virus infection and particularly in those considered to have quiescent disease, such as HBeAg-negative patients with low-level viraemia, remains understudied. A better understanding of this disease phase is critical in order to determine the risk of HCC development. Moreover, there is a need to define the localisation of integrated HBV DNA in genes regulating hepatocyte function other than proliferation (such as cell metabolism or recognition by the immune response) and the impact of HBV DNA integration on perturbing cellular gene expression.

In light of this knowledge gap, we studied HBeAg-negative patients across the disease spectrum, including those with low-level to moderate-level viraemia, in order to provide a comprehensive characterisation of the intrahepatic HBV reservoir and to investigate how this might determine the extent and localisation of HBV DNA integration in the whole exome. Furthermore, we report for the first time on the role of HBV DNA integration in altering the transcriptome profile of human hepatocytes in these patients. This highlights the importance of a more circumspect approach in the management of low-to-moderate viraemic HBeAg-negative CHB and whether greater consideration should be given to lowering treatment thresholds in these patients.

MATERIALS AND METHODS

Study population

Eighty-four HBeAg-negative patients with CHB were recruited from Viral Hepatitis clinics at The Royal London Hospital (Barts

Health NHS Trust) UK from 2013 to 2016. All patients underwent a liver biopsy and tissue surplus to diagnostics was used for experimental assays as described. Matched serum samples were also collected (online supplemental figure S1 depicts the patient cohort and virological analyses that were performed). All patients with CHB were treatment-naïve and monitored for ≥2 years with virological and biochemical parameters for disease stratification prior to tissue sampling. Patients were HBV mono-infected; and co-infection with HCV, HIV and hepatitis Delta virus was excluded. Patients were stratified according to serum HBV DNA levels (calculated as the mean value obtained by consecutive measurements during a 2-year period of monitoring before undergoing liver biopsy):

- Group 1: patients with serum HBV DNA persistently <2000IU/mL; low viraemia (n=12).
- Group 2: patients with serum HBV DNA between 2000 and 20 000IU/mL; moderate viraemia (n=25).
- Group 3: patients with serum HBV DNA persistently >20 000IU/mL; high viraemia (n=47).

DNA and RNA isolation from liver biopsy tissue

Total intrahepatic DNA and RNA were isolated from patients' liver tissues by using AllPrep DNA/RNA Mini Kit (QIAGEN, Hilden, Germany) (procedure described in online supplemental material (SM)).

Quantification of intrahepatic HBV markers: total HBV DNA, cccDNA and pgRNA

Intrahepatic DNA was used to perform itHBV DNA and cccDNA quantification as described in SM. pgRNA levels were determined for 41/84 liver biopsies with an available RNA sample, by applying an in-house digital droplet (dd)-PCR assay (described in SM). itHBV DNA, cccDNA and pgRNA values were normalised to cell number, according to the quantification obtained by Albumin-based ddPCR copy number assay: Alb, Human (Bio-Rad, Pleasanton, California, USA) and, thus reported as itHBV DNA, cccDNA and pgRNA copies/1000 cells. At least three negative controls and one positive control were included in each real-time and ddPCR reaction to verify the amplification efficiency and to exclude sample contamination.

Serological HBV markers: quantification of HBV DNA, HBsAg and HBcrAg

Serum HBV DNA quantification was performed by a real-time PCR using the Cobas Ampliprep/Cobas TaqmanHBV assay (Roche Diagnostics, Manheim, Germany; lower limit of quantitation (LLOQ): 20IU/mL). Serum hepatitis B surface antigen (HBsAg) was quantified with Elecsys HBsAg II kit/Cobas (Roche Diagnostics; LLOQ: 0.05 IU/mL). Hepatitis B core-related antigen (HBcrAg) was quantified by a chemiluminescent assay, Lumipulse GHcrAg assay (Fujirebio Europe, Gent, Belgium; LLOQ: 3 logU/mL).

Assessment of liver fibrosis

Liver fibrosis was assessed by a specialist liver histopathologist using the Ishak scoring system. Values from 0 to 2 were considered as absence of or mild fibrosis only, while values from 3 to 6 were considered as moderate-to-severe fibrosis/cirrhosis.

HBV genotyping

For each patient, HBsAg population-based sequencing (226 amino acids) was performed on DNA extracted from serum

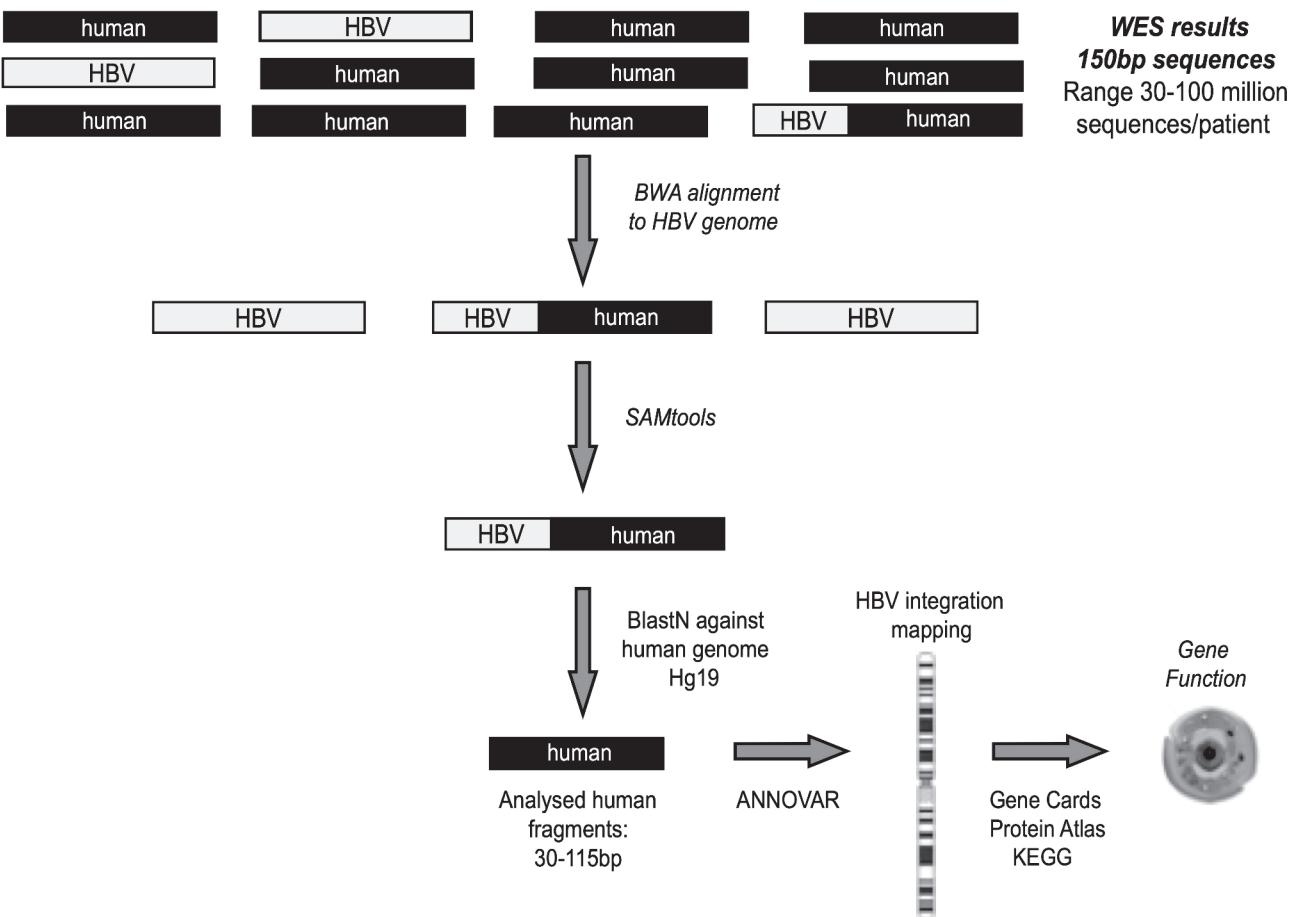


Figure 1 Bioinformatic workflow for the identification of chimeric HBV human sequences. Following quality control (by Trimmomatic), sequencing reads were aligned to virus-specific genome by Burrows Wheeler Aligner-Maximal Exact Matches (BWA-MEM) in order to extract all reads that contained HBV fragments. SAMtools software was applied to extrapolate all chimeric HBV human sequences, representing the HBV integrations into human genome. ANNOVAR software was used to map HBV integrations at chromosome and gene level. The functionality of genes involved in HBV integration was retrieved by Gene Cards, Protein Atlas and KEGG databases. WES, whole exome sequencing.

sample, following a protocol previously described.¹³ A phylogenetic approach was used to determine HBV genotype.

Whole exome sequencing

Whole exome sequencing (WES) was carried out for 40/84 liver biopsies where adequate DNA was available, by applying a well-consolidated next-generation sequencing approach, based on Illumina technology routinely used for diagnostic purposes at Eurofins Genoma group in Rome, Italy (detailed description in SM). This procedure contains a crucial step for capturing the exome-amplified fragments, allowing an in-depth analysis of the coding fraction of human genome and its flanking intronic/intergenic regions.¹⁴ Overall, median (IQR) coverage of exome sequencing was 115× (90×–140×). A stringent bioinformatic pipeline, described in SM and schematised in figure 1, was applied to determine HBV integration sites. The functionality of genes involved in HBV integration was retrieved by querying three different online available databases: Gene Cards, Protein Atlas and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (figure 1).^{15–17} The publicly available Virus Integration Site DataBase (VISDB) was used to verify if the observed HBV DNA integrations occurred in genes described as targets of HBV DNA integrations in previous studies analysing liver samples (tumour/peritumour/non-tumour). Indeed, VISDB collects information from literature on high-quality viral integration sites in

the human genome and related malignancies (<https://bioinfo.uth.edu/VISDB/index.php/homepage>). For HBV, VISDB has data on 20 558 integration sites from 45 publications.

Quantification of HBV integrations by a digital droplet PCR approach

For HBV DNA integrations occurring in intronic/exonic human genomic regions detected by WES, ad hoc ddPCR assays were designed in order to further confirm their presence and to quantify the total number of hepatocytes harbouring HBV DNA integration events (detailed description provided in SM).

Whole transcriptome analysis

Whole transcriptome sequencing was performed on a subset of 15 patients with available material: 7 patients with confirmed HBV DNA integration in the whole exome and 8 without confirmed HBV DNA integration in the whole exome in order to evaluate the differential expression of overall mRNAs in liver biopsies. RNA libraries were sequenced with Illumina Hiseq 2000 (Illumina, California, USA). Approximately 45–60 million paired-end 150 base pair reads were obtained per sample. The relative abundance of each annotated transcript for each sample was reported in units of transcripts-per-million (TPM) (detailed description in SM).

Table 1 Patient characteristics

Characteristics	Overall population	Group 1*	Group 2*	Group 3*
No. cases	84	12	25	47
Age, (years) median (IQR)	35 (28–42)	37 (28–43)	35 (28–43)	35 (29–41)
Male, n (%)	58 (69.0%)	10 (83.3%)	18 (72.0%)	30 (63.8%)
Ethnicity, n (%)				
European	24 (28.6%)	4 (33.3%)	7 (28.0%)	13 (27.6%)
Asian	41 (48.8%)	5 (41.7%)	12 (48.0%)	24 (51.1%)
African	19 (22.6%)	3 (25.0%)	6 (24.0%)	10 (21.3%)
Serum HBV DNA, logIU/mL (median, IQR)†	4.2 (3.5–5.2)	2.7 (2.0–2.8)	3.7 (3.5–3.8)	5.1 (4.5–5.5)
Serum ALT, U/L (median, IQR)†	41 (28–57)	48 (28–96)	31 (27–41)	42 (29–58)
HBsAg, logIU/mL (median, IQR)	3.7 (3.2–4.0)	3.4 (3.1–4.1)	3.6 (3.2–3.9)	3.8 (3.3–4.0)
HBcrAg negative, n (%)‡	35 (70%)	6 (66.7%)	18 (81.1%)	11 (28.2%)
HBcrAg, logU/mL (median, IQR)‡	4.0 (3.3–4.8)	3.8 (3.4–4.0)	3.4 (3.2–4.2)	4.1 (3.3–4.8)
HBV genotype, n (%)				
A	9 (10.7%)	1 (8.3%)	1 (4.0%)	7 (14.9%)
B	6 (7.1%)	1 (8.3%)	3 (12.0%)	2 (4.3%)
C	16 (19.0%)	2 (16.7%)	3 (12.0%)	11 (23.4%)
D	38 (45.2%)	6 (50.0%)	12 (48.0%)	20 (42.6%)
E	15 (17.9%)	2 (16.7%)	6 (24.0%)	7 (14.9%)
Ishak fibrosis stage, n (%)				
0–2	70 (83.3%)	11 (91.7%)	24 (96.0%)	35 (74.5%)
3–6	14 (16.7%)	1 (8.3%)	1 (4.0%)	12 (25.5%)

*Groups 1, 2 and 3 were defined according to serum HBV DNA levels (mean value obtained by consecutive measurements during the 2-year follow-up prior to performing liver biopsy; HBV DNA persistently <2000 IU/mL (group 1); HBV DNA 2000–20 000 IU/mL (group 2); HBV DNA >20 000 IU/mL (group 3)).

†The median value (IQR) of serum HBV DNA, HBsAg and ALT, calculated by using mean values observed during the 2-year monitoring period prior to performing liver biopsy.

‡Data available for 70/84 patients (9/12 in group 1; 22/25 in group 2; 39/45 in group 3).

ALT, alanine aminotransferase; HBcrAg, hepatitis B core-related antigen; HBsAg, hepatitis B surface antigen.

Statistical analysis

Mann-Whitney test for continuous variables and χ^2 test for discrete variables were applied to define statistically significant differences. The area under the receiver operating characteristics (AUROC) was used to define the thresholds of peripheral parameters with the best performance in predicting intrahepatic reservoir (cccDNA <1.5 log copies/1000 cells) and the threshold of HBsAg levels with the best performance in predicting the occurrence of HBV integration into the human genome.

RESULTS

Characterisation of serum and intrahepatic HBV markers

Eighty-four HBeAg-negative CHB treatment-naïve patients, stratified according to viral load (see 'Materials and methods' section) were included in the study. There were no significant differences in clinical and demographic data between the patient groups (table 1). Importantly, no/mild fibrosis (F0–2) was found in 83.3% of the study patients.

HBsAg levels were comparable between all patient groups, however, most patients in groups 1 and 2 displayed HBcrAg <3 logU/mL (indicating a negative result) (66.7% and 81.1%, respectively; $p=0.38$). Conversely, HBcrAg levels were significantly greater in high viraemic patients, median (IQR) HBcrAg 4.1 (3.3–4.8) logU/mL ($p=0.03$ for group 1 vs group 3 and $p<0.001$ for group 2 vs group 3) (table 1). Similarly, patients with low-level and moderate-level viraemia demonstrated comparable median levels of itHBV DNA and cccDNA (figure 2A,B), which were significantly lower than those observed for patients with high viraemia ($p=0.004$ to 0.02) (figure 2A,B). Likewise, pgRNA levels were comparable between groups 1 and 2, although the distribution of pgRNA was narrower in group 1 (figure 2C). Analogous to HBcrAg levels, patients in group 3

were characterised by significantly higher pgRNA levels than the other two groups (group 3 vs group 2, $p=0.002$; group 3 vs group 1, $p<0.001$) (figure 2C). By AUROC, the combination of serum HBV DNA <20 000 IU/mL, HBcrAg <3 logU/mL and HBsAg <1000 IU/mL identified a limited HBV reservoir (defined as cccDNA <1.5 log copies/1000 cells) with 84.3% diagnostic accuracy, 90% positive predictive value (PPV) and 83.3% negative predictive value (NPV) (online supplemental table S1).

The intrahepatic HBV reservoir was also evaluated across HBV genotypes. In particular, patients infected with HBV genotype D were characterised by a lower cccDNA and intrahepatic HBV DNA compared with genotypes A and E (for cccDNA: 1.9 (0.9–2.6) vs 2.7 (2.0–3.2) and 2.4 (2.1–2.7) log copies/1000 cells, $p=0.007$ and $p=0.004$; for intrahepatic DNA: 3.2 (2.9–3.7) vs 4.1 (3.7–4.2) and 3.9 (3.5–4.0) log copies/1000 cells, $p=0.005$ and $p=0.02$). Similarly, the profile of HBV genotype D correlated with a lower pgRNA than genotype A (3 (1.4–11) vs 19 (7–32) copies/1000 cells, $p=0.05$). Conversely, data from genotype D patients showed a comparable intrahepatic reservoir in respect to HBV genotypes B and C.

HBV integration events occur throughout the spectrum of HBeAg-negative CHB

HBV DNA integration was then investigated in the whole exome (defined by exons, exon-flanking intronic regions and intergenic regions located at <500 kilobase pairs from exons) obtained from 40 out of 84 patients, where sufficient DNA was available for WES analysis. Next-generation sequencing technologies are used by WES and have contributed to unveiling the landscape of genetic and epigenetic aberrations.¹⁸ The subset of 40 patients included in the analysis, accurately reflected the overall study population in disease stratification according to serum

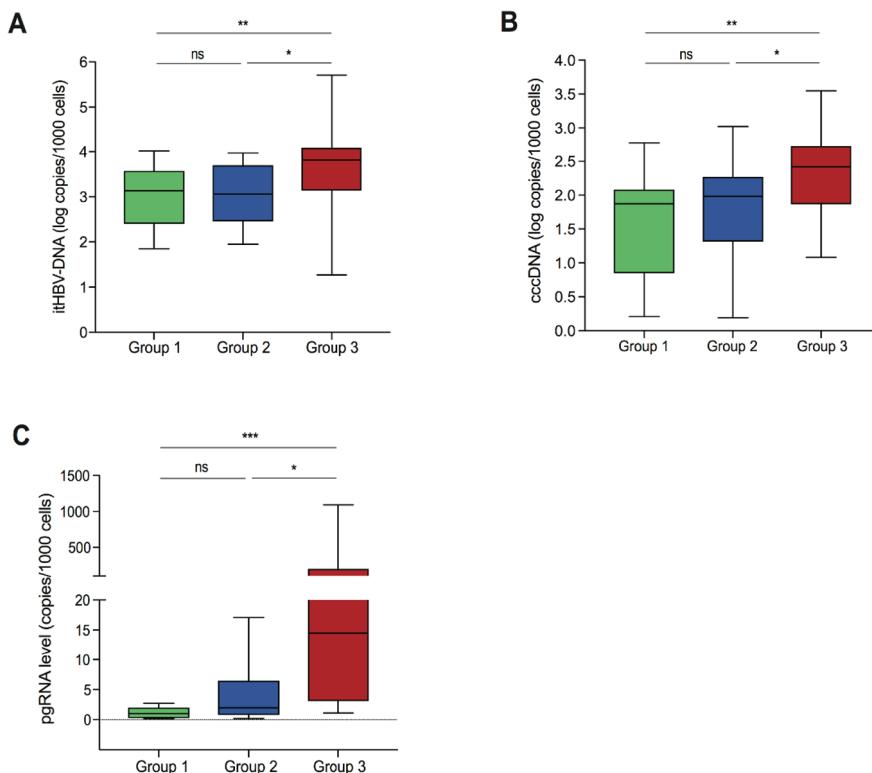


Figure 2 Distribution of intrahepatic HBV markers in the patient groups. Box plots displaying the distribution of intrahepatic total HBV DNA (itHBV DNA) (A), covalently closed circular DNA (cccDNA) (B) and pregenomic RNA (pgRNA) (C) in group 1, group 2 and group 3, respectively. Statistically significant differences were assessed by Mann-Whitney U test. Significant changes marked with asterisks, $*p<0.05$; $**p<0.01$; $***p<0.001$; ns, not significant.

HBV DNA levels (online supplemental table S2). Median (IQR) sequence coverage depth was $115\times$ ($90\times$ – $140\times$), with a median (IQR) number of paired-end reads of 70 (54–84) millions per patient, consistent with the recommendation for WES.^{19 20} At least one HBV DNA integration event in the whole exome was detected in all patient groups with the highest prevalence in group 3 (55.6%, 10/18). Notably, HBV DNA integration was also evident in 14.3% (2/14) and 25% (2/8) of patients in groups 2 and 1, respectively, despite lower viraemia and a more limited intrahepatic reservoir. HBV DNA integration profiles were also evaluated across HBV genotypes; integration was detected more frequently in genotype D patients (38.9%) followed by genotype E (33.3%) and genotype C patients (22.2%). For HBV genotype A and B, HBV DNA integration was analysed in two patients each, where integration events were detected in both (2/2) patients infected with genotype A, but in neither of those with genotype B HBV infection.

Differential localisation of HBV DNA integration events in the whole exome

The identified HBV DNA integrations were mapped with regard to the cellular chromosomes. Integration events were detected in 12 chromosomes without evidence of preferential chromosomal hotspots (figure 3A). A total of 16 HBV integration events were detected. Most HBV DNA integration events (68.7%, 11/16) occurred within exon-flanking introns. Among them, eight HBV DNA integrations were detected within or in close proximity to signal sequences necessary for the RNA splicing, an event crucial for mRNA synthesis and in turn for protein functionality (table 2). Furthermore, HBV DNA integration was also detected in exons (6.3%, 1/16) and in intergenic regions (25%, 4/16). The

HBV DNA integration events occurring in regions crucial for mRNA synthesis were also quantified by ddPCR (eight within or in close proximity to signal sequences necessary for RNA splicing and the single event in the exon) (table 3). The quantification of these HBV DNA integration events ranged from 0.5 to 157 integrations per 1000 hepatocytes (median=5 integrations per 1000 hepatocytes) (table 3).

By gene ontology, HBV DNA integrations localised in human genes regulating cell proliferation (NUP85, ANKRD52, ELAC2, COL18A1 and AGBL5) in five patients, including those patients with low-level and moderate-level viraemia (table 2). Notably, by using the publicly available database VISDB, collecting 20 558 HBV integration sites in tumour/peritumour/non-tumour liver samples from 45 publications; the ANKRD52, COL18A1 and AGBL genes involved in HBV DNA integration were previously detected in tumour or peritumour liver samples. Integrated HBV DNA was also found in genes regulating drug or lipid metabolism (CYP2UI, LMF-1) and in modulating antiviral or inflammatory responses (NR3C1, IFITM1) (table 2). According to VISDB, LMF-1 and NR3C1 genes were also involved in HBV DNA integration events.

Viral genomic regions are integrated within the human whole exome

The HBV DNA integrations were then mapped into the viral genome. Most HBV DNA integrations (62.5%, 10/16) occurred between HBV nucleotide positions 1590 and 1840 corresponding to the viral genomic region spanning the direct repeat 2 (DR2) and direct repeat 1 (DR1) (figure 3B). This region contains the enhancer II, known to upregulate the expression of viral genes²¹ and corresponds to the C-terminus of hepatitis B X protein

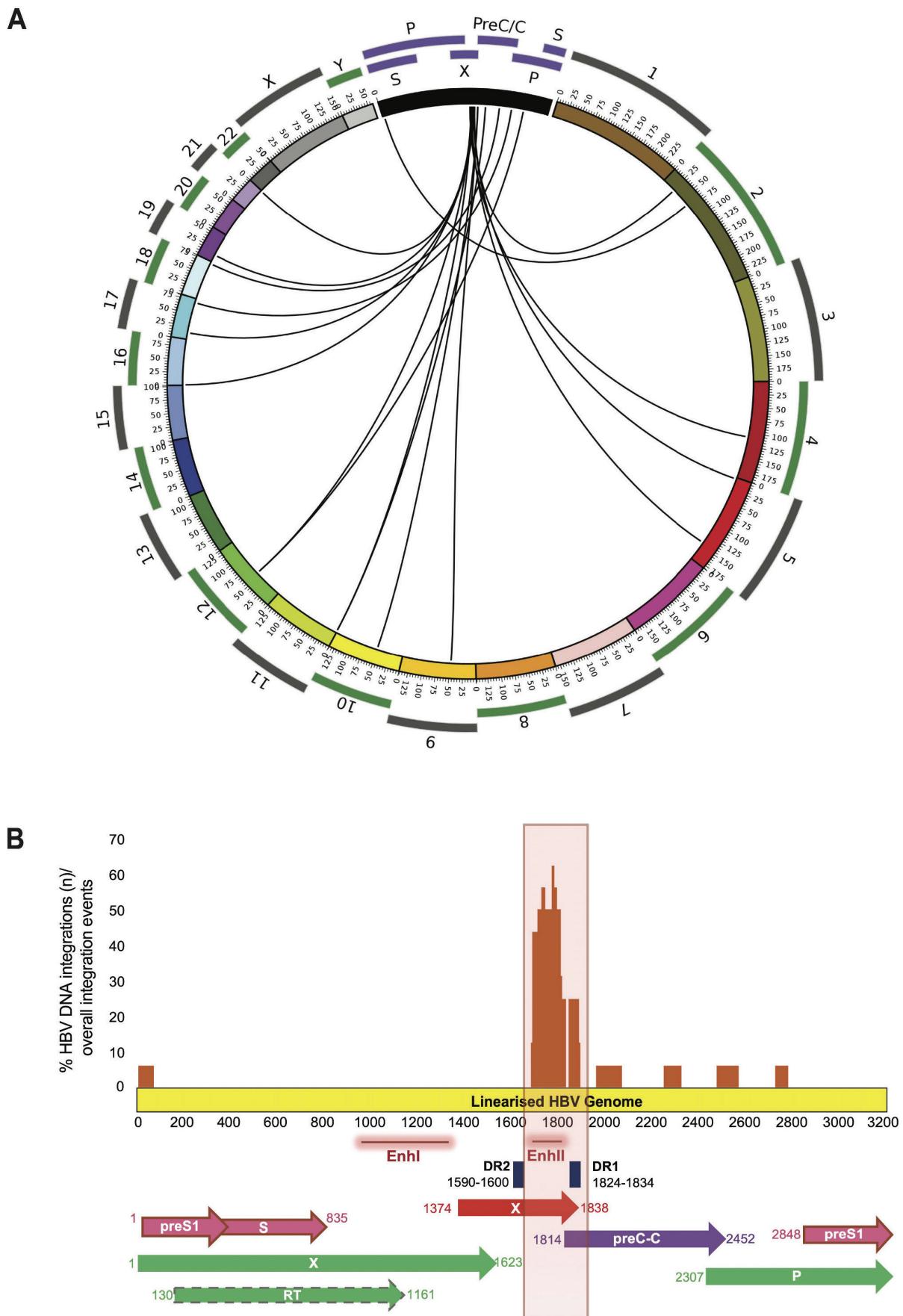


Figure 3 Visualisation of HBV DNA integrations with regard to human chromosomes and the viral genome. (A) Each line defines HBV integration breakpoints at particular loci in the human genome (hg19). (B) The histogram reports the frequency of integration for each nucleotide position of HBV genome. The locations of the genes encoding HBV polymerase (green), core (violet), S (pink) and X (red) proteins are shown.

Table 2 Location of HBV DNA integrations in coding regions of the human genome by whole exome sequencing

Patient no.	Patient group	Chimeric HBV human reads (n)	Human chromosome no.	Human genes	Gene description*	Location of HBV integrants within human genes	Distance to proximal exon†	Potential downstream effect‡
<i>HBV DNA integration in intronic or exonic regions of human genome</i>								
14	1	1	Chr_21	COL18A1§	Collagene XVIII	Intronic	14 091	Regulation of angiogenesis and tumour growth
22	1	1	Chr_16	LMF1§	Lipoprotein maturation factor 1	Intronic	7157	Protein involved in lipoprotein metabolism
33	2	1	Chr_19	ADGRL1§	G-protein coupled receptors	Intronic	7310	Involved in cell adhesion and signal transduction
39	3	1	Chr_17	ELAC2	Ribonuclease Z	Intronic	15¶	Altered levels in liver cancer
40	3	2	Chr_11	IFITM1	Interferon-induced antiviral protein	Intronic	44 and 72**	B cell receptor signalling pathway
44	2	1	Chr_17	NUP85	Nuclear pore complex	Intronic	6**	RNA transport pathway; altered levels are associated with negative prognosis in liver cancer
62	3	8	Chr_12	ANKRD52§	Ankyrin repeat domain	Intronic	6 and 32**	Altered levels are associated with negative prognosis in liver cancer
75	2	1	Chr_2	AGBL5§	Metallocarboxypeptidase	Intronic	80¶	Altered levels are associated with negative prognosis in liver cancer
83	2	2	Chr_5	NR3C1§	Glucocorticoid receptor	Exonic	0	Regulation of inflammatory responses
84	3	2	Chr_4	CYP2U1	Cytochrome P450	Intronic	99¶	Drug metabolism
<i>HBV DNA integration in intergenic regions of human genome</i>								
2	3	2	Chr_9	n.a.	n.a.	Intergenic	458 413	n.a.
10	3	1	Chr_5	n.a.	n.a.	Intergenic	69 228	n.a.
68	3	1	Chr_18	n.a.	n.a.	Intergenic	158 827	n.a.
81	3	2	Chr_10	n.a.	n.a.	Intergenic	17 498	n.a.

The table displays the 12 HBV DNA integration events observed in the 10 patients localised in intronic or exonic regions of human genome and the 4 HBV DNA integration events localised in intergenic regions.

*Genes information is retrieved by GeneCards: The Human Gene Database.

†Distance in nucleotides of HBV integrants to the most proximal exons.

‡Genes' cellular role is retrieved from the Human Protein Database and from KEGG PATHWAY Database.

§These genes were involved in HBV DNA integration events as reported in Virus Integration Sites Database (<https://bioinfo.uth.edu/VISDB/index.php/homepage>).

¶HBV integrants reside in proximity (<100 nucleotides) to donor site, localised at 5' end of introns and crucial for RNA splicing.

**HBV integrants reside within or in proximity (<50 nucleotides) to branching site, a crucial region localised at 3' end of introns and crucial for RNA splicing.

n.a., not applicable.

(HBx), the expression of which is associated with the onset of liver cancer.^{22 23} HBV integration events were also observed in the open reading frame preC/Core (ORF C) (in 3/16), ORF P (in 2/16) and in the ORF P genomic region overlapping with ORF S (in 1/16).

Two distinct HBV DNA integration events were detected in two patients. In patient no. 40, the viral genome regions encompassing the nucleotides 1665–1746 and 1951–2039 were both detected integrated in intron 1 of IFITM1 gene with a frequency of 4 and 5 integrations per 1000 hepatocytes (table 3). The viral genome regions including the nucleotides 1684–1775 and 2716–2769 were both detected integrated in intron 7 of ANKRD52 gene, with

a frequency of 158 and 157 integrations per 1000 hepatocytes in patient no. 62 (table 3). The localisation in the same human gene with a superimposable burden supports the integration of a long portion of viral genome of at least 354 and 1085 nucleotides.

HBV DNA integration events correlate with serum HBsAg levels

Intrahepatic HBV markers were correlated with HBV DNA integration events and noted to be comparable in patients with or without evidence of HBV DNA integration within the whole exome (table 4). Conversely, among serum HBV markers, HBsAg was significantly higher in patients demonstrating integrated HBV DNA in the whole exome (3.9 (3.8–4.2) logIU/mL vs 3.2 (2.9–3.7) logIU/mL, p<0.001) (table 4). Furthermore, by AUROC, HBsAg >5000 IU/mL identified the occurrence of HBV integration in the whole exome with the best diagnostic accuracy (86.5%), 70.6% PPV and 94.7% NPV (figure 4). No correlation, however, was noted between the occurrence of HBV DNA integration and ALT or Ishak fibrosis stage. Moreover, no correlation was observed for patient's age [34(26–39) years for those with HBV DNA integration vs 37 (30–43) years for patients without integration, p=0.16].

HBV DNA integration impacts the transcriptome profile of human hepatocytes

Cellular gene expression was evaluated in a subset of 15 out of 40 patients analysed for WES, where adequate tissue was available. These patients were stratified according to the presence (n=8) or absence (n=7) of HBV DNA integration as indicated in the heatmap (figure 5). A median (IQR) number of 55 (52–61) million paired end reads for each patient were obtained and a total of 52 268 expressed genes were analysed. Among them,

Table 3 Quantification of HBV DNA integrations involving intron/exon human regions by droplet digital PCR

Patient no.	Human genomic regions*	HBV regions	Number of HBV integrations/1000 hepatocytes†
39	ELAC2 intron-10	HBx	8
40	IFITM1 intron-1	HBx	4
	IFITM1 intron-1	Core	5
44	NUP85 intron-3	Pol	3
62	ANKRD52 intron-7	Pol	158
	ANKRD52 intron-7	HBx	157
75	AGBL5 intron-12	HBx/Core	7
83	NR3C1 exon-5	HBx	4
84	CYP2U1 intron-4	HBx	0.5

*The human genomic regions involved in HBV integrations were retrieved by ANNOVAR software.

†HBV integrations were quantified by digital droplet PCR assays.

HBx, hepatitis B X protein.

Table 4 Correlation of intrahepatic and serological markers with the occurrence of HBV DNA integration

Patients' characteristics	At least one integration event (n=14)	No integration events (n=26)	P value*
Serological parameters			
Median serum ALT, U/L (IQR)	41 (27–90)	42 (28–67)	0.9
Median serum HBV DNA, log ₁₀ U/mL (IQR)	4.6 (3.6–5.6)	3.7 (3.1–5.0)	0.2
Median serum HBsAg, log ₁₀ U/mL (IQR)	3.9 (3.8–4.2)	3.2 (2.9–3.7)	<0.001
Median serum HBcrAg, log ₁₀ U/mL (IQR)	4.0 (3.9–4.5)	4.5 (3.4–5.4)	0.5
Intrahepatic parameters			
Median cccDNA (log copies/1000 cells) (IQR)	2.4 (1.9–2.7)	2.0 (1.0–2.6)	0.17
Median itHBV DNA (log copies/1000 cells) (IQR)	3.6 (3.1–3.9)	3.5 (2.7–4.0)	0.6
Median pgRNA (copies/1000 cells)†	3.1 (1.9–226)	4.5 (0.6–17)	0.73
Liver damage			
Ishak fibrosis stage			
0–2	10 (32.3)	21 (66.7)	0.6
3–6	4 (44.4)	5 (55.5)	

*P values were determined by Mann-Whitney U test for continuous variables and χ^2 test for discrete variables.

†pgRNA levels were available for 21/40 patients analysed for HBV integration.

ALT, alanine aminotransferase; cccDNA, circular covalently closed DNA; HBcrAg, hepatitis B core-related antigen; HBsAg, hepatitis B surface antigen; itHBV DNA, intrahepatic total HBV DNA; pgRNA, pregenomic RNA.

hRAS (encoding the oncoprotein p21) and HGB2 (encoding fetal haemoglobin) were the only genes whose transcript levels were significantly higher in patients with HBV DNA integration than in patients with no HBV integrants (median (IQR): 255 (151–539) vs 32.1 (24.3–33.7) TPM, adjusted p=0.00002 for hRAS and

37.5 (20.1–115) vs 2.5 (0.9–3.1) TPM, adjusted p=0.00014 for HGB2, respectively). We then investigated the quantitated expression of cellular genes in those patients where there was evidence of HBV DNA integration by WES. In the patient with HBV DNA integration in the cellular gene ELAC2, its transcript levels were lower than those observed in patients without any HBV DNA integration (92 vs a median (IQR) value of 217 (186–232) TPM in those without HBV DNA integration, adjusted p=0.05).

Similarly, in the patient with HBV DNA integration in ANKRD52, the transcript levels of this gene were lower than those observed in patients without any HBV DNA integration, although not reaching statistical significance (119 vs a median (IQR) value of 220 (155–243) TPM in those without HBV DNA integration, p=0.20). Conversely, in the patient with HBV DNA integration in the cellular gene IFITM1, the transcript levels were higher (111 vs a median (IQR) value of 42 (36–50) TPM in those without HBV DNA integration, p=0.02). Indeed, no further differences in transcript levels were observed for the other genes analysed in transcriptome analysis.

DISCUSSION

In this study, we provide an in-depth characterisation of the intrahepatic HBV reservoir and its impact on HBV DNA integration in HBeAg-negative patients with CHB. Our data show that patients with low-level and moderate-level HBV DNA were characterised by a comparable intrahepatic HBV reservoir in terms of total intrahepatic HBV DNA and cccDNA, which was significantly lower than that observed in patients with higher viraemia. This finding is in keeping with previous clinical studies showing that most patients in the so-called grey zone (serum HBV DNA ranging between 2000 and 20 000 IU/mL) tend to have a benign clinical outcome characterised by limited liver disease progression.^{24 25} Interestingly, despite comparable cccDNA levels, pgRNA is widely distributed between low and moderately viraemic patients, suggesting a stricter epigenetic control of cccDNA transcriptional activity with low levels of HBV DNA. Again, this corroborates previous clinical studies showing that a number of patients with serum HBV DNA from 2000 to 20 000 IU/mL (13% within 1 year according to a recent study²⁴) progress to HBeAg-negative chronic hepatitis. Furthermore, our data show that HBcrAg along with serum HBV DNA and HBsAg can predict a limited intrahepatic HBV reservoir. This is consistent with recent data showing that in HBeAg-negative patients with CHB, HBcrAg can reflect the extent of cccDNA transcriptional activity.⁵ While this underlines the importance of incorporating serum HBV biomarkers into clinical practice for a more precise stratification of HBeAg-negative patients with CHB, less is known about the impact of the HBV reservoir on HBV DNA integration, which is an important factor to consider in disease progression and more specifically in the development of HCC figure 6.

Using WES coupled with ddPCR in this study, we provide the unique opportunity to finely unravel in HBeAg-negative patients with CHB, both the occurrence and burden of HBV DNA integration in the most relevant parts of the human genome. The whole exome represents the protein-coding regions composed of 180 000 exons where approximately 85% of genomic alterations associated with human genetic diseases are localised.^{26 27} Thus, HBV DNA integration in the whole exome has the highest probability to alter human gene expression and, in turn, to be of clinicopathological significance. Furthermore, recent studies found that tumour tissues were characterised by an irregular enrichment of HBV DNA integration events in coding/functional regions of the human genome, compared with non-tumour

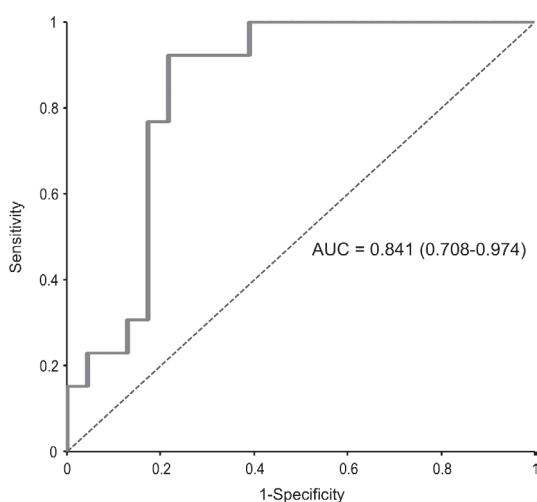


Figure 4 Area under receiver operating characteristics (AUROC) curve of hepatitis B surface antigen (HBsAg) levels and occurrence of HBV integration. HBsAg levels >5000 IU/mL were able to identify patients with HBV integration (n=14) with the best diagnostic accuracy (83.5%): positive predictive value=70.6%; negative predictive value=94.7%; area under the curve (AUC)=0.841 (0.708–0.974); sensitivity=92.3%; specificity=78.3%.

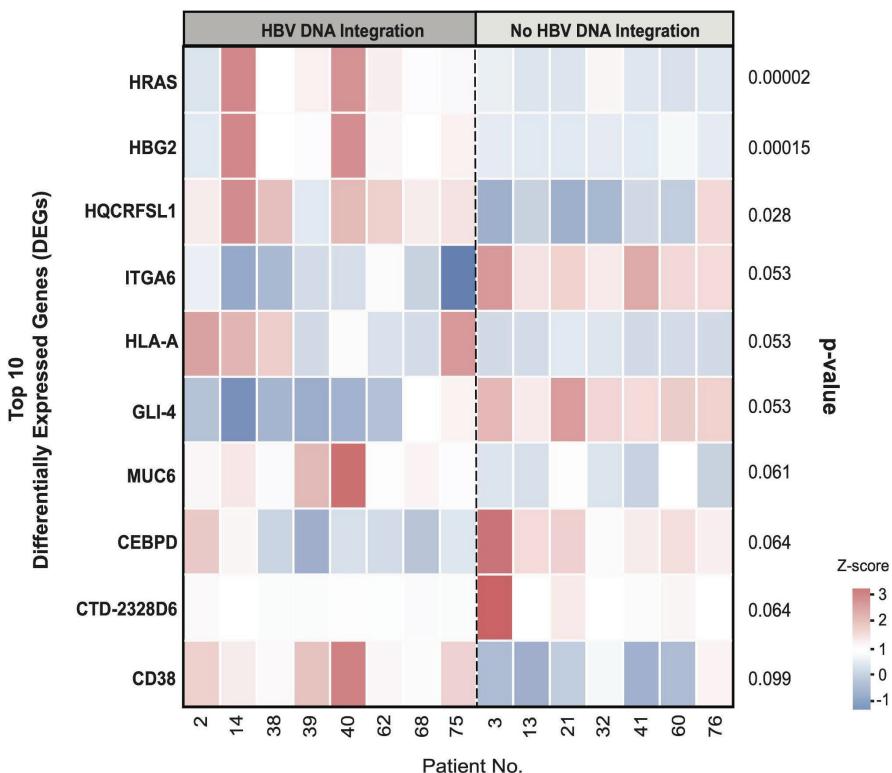


Figure 5 Heatmap of the top 10 differentially expressed genes in the presence or absence of HBV DNA integration. The colour from red to blue refers to the z-score calculated for each gene in each patient according to the following formula: $X - (\text{mean } X)/\text{SD}$, where X stands for transcripts per million. The heatmap reports the top 10 genes with differential expression in patients with or without evidence of HBV DNA integration and their respective p values.

tissues, suggesting that HBV integration in these regions could confer a selective advantage during tumourigenesis.^{7,28}

Our data show that HBV integration in the whole exome occurs in a notable proportion of patients with high HBV DNA (55.6%), and at significant levels in patients with low (25%) and moderate viraemia (14.3%), despite their limited HBV reservoir. By ddPCR, we found that HBV DNA integrations occurred with a prevalence ranging from 0.5 up to 158 events per 1000 hepatocytes, supporting a clonal expansion of hepatocytes harbouring these HBV DNA integrations. Notably, HBV DNA integration was detected in human genes such as NUP85, ANKRD52, ELAC2 and AGBL5 involved

in the regulation of cell proliferation. It is conceivable that the enrichment of HBV DNA integration in these genes may enhance hepatocyte survival thus allowing persistence of an intrahepatic viral reservoir and long-term production of infectious progeny, as previously shown for other viral infections.^{29,30} At the same time, this may pose the basis for clonal selection of hepatocytes harbouring these HBV DNA integrations, contributing to the development of liver cancer. In line with this concept, the altered expression of the aforementioned genes has been associated with the onset or worse prognosis of HCC.^{16,31–34} Notably, by interrogating the publicly available database VISDB, we confirmed some of

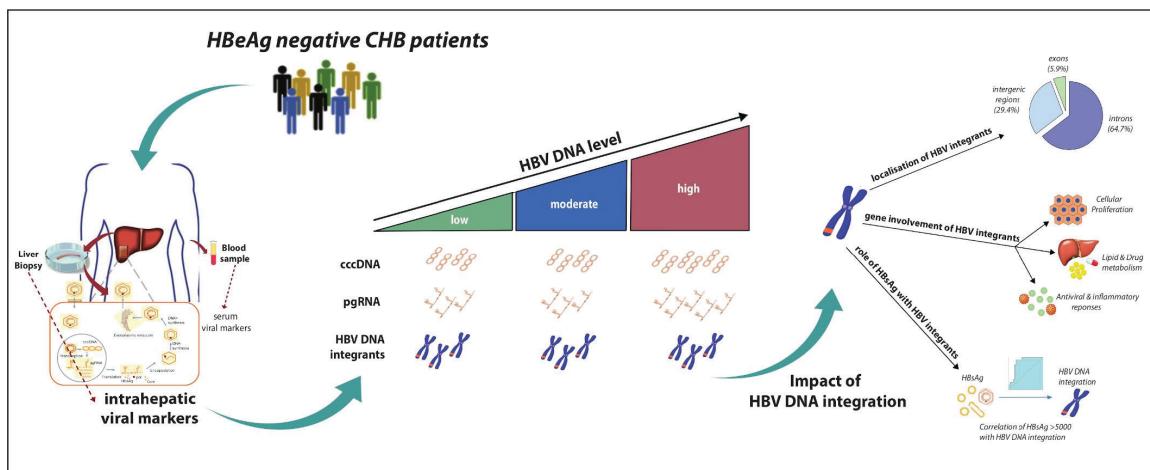


Figure 6 Infographic summary highlighting the key points of the data.

these genes were involved in HBV DNA integration events detected in tumour or peritumour liver samples in previous studies, supporting their role in hepatocarcinogenesis.

Most HBV DNA integrations were detected in exon-flanking intronic regions where signal sequences crucial for RNA splicing (such as the branching site) are localised. A plausible concept is that HBV DNA integration in these regions can affect the correct synthesis of mRNAs, thus altering cellular gene expression. In line with this hypothesis, among the genes noted, transcriptome analysis revealed a significantly lower expression of ELAC2 in the patient with HBV DNA integration within this gene. This result could be explained by the fact that HBV DNA integration resides within the so-called branching site, an intronic region crucial for the proper folding of RNA during the splicing process. By applying the 'Human Splicing Finder' (a consolidated and robust algorithm to predict the branching site within a given intron), we found that HBV DNA integration in ELAC2 determines a 3-nucleotide shift in the branching site of the intron involved in HBV integration and reduces stability of this domain.

Interestingly, recent studies have shown that an aberrant expression of ELAC2 plays a role in the neoplastic transformation of different cell types including hepatocytes.^{31 35}

Analysing the whole transcriptome in patients according to the presence or absence of HBV DNA integration, a significant upregulation of the cellular genes hRAS and HBG2 (encoding fetal haemoglobin) was highlighted in those with HBV DNA integration. Both these genes are associated with increased cell proliferation,^{15 16} in addition, the upregulation of hRAS, a proto-oncogene, has been associated with neoplastic transformation in several human cancers.^{36 37} Furthermore, it has been demonstrated that the aberrant activation of cellular signalling processes such as Ras/Raf/MAPK/72 plays a pivotal role in HCC development.³⁸ Similarly, a moderate-to-strong cytoplasmic positivity to fetal haemoglobin (usually not expressed in adults) was described in patients with HCCs.¹⁶ The increased expression of these genes could be favoured by genome instability, promoted by HBV DNA integration.^{11 39} This concept is in keeping with a recent study showing the capability of HBV DNA integration to act as bridge for interchromosomal translocations.⁴⁰ hRAS upregulation can also be related to the overexpression of the viral protein HBx, whose ORF was found integrated in the vast majority of patients. Indeed, this is in line with previous data showing that an elevated HBx production can favour the cytoplasmic accumulation of p21 (encoded by hRAS), providing the basis for the neoplastic transformation of hepatocytes.⁴¹ Notably, an interesting case is patient 40, characterised by a co-localisation of integrated HBV DNA (within the gene encoding IFITM1) along with the differentially expressed genes HRAS, HBG2 and MUC6 in chromosome 11. In this patient, the variation in the expression levels of HRAS, HBG2 and MUC6 was particularly marked, further supporting the role of HBV DNA integration in altering transcriptome profiles. Finally, changes in transcriptome profiles were also observed in patients with HBV DNA integration in intergenic regions. This may be attributable to the potential of HBV DNA integrations to modify the accessibility of cell chromatin and/or to alter the production of regulatory RNAs, typically encoded by intergenic regions of human genome.⁴² Overall, these findings support the concept that HBeAg-negative CHB patients, even with low levels of HBV DNA, still carry a potential risk of neoplastic transformation of hepatocytes.

It is noteworthy that HBV DNA integration was also detected in genes regulating antiviral immunity and hepatocyte metabolism. In this regard, transcriptome analysis highlights a higher

expression of the interferon-associated gene IFITM1 in the patient with HBV DNA integration involving this gene. Notably, by analysing the region of HBV genome integrated within IFITM1, we found that this event involves an HBV region enriched of promoters (nucleotides: 1665–1746), including the enhancer II, recognised by transcription factors critical in favouring gene expression. A plausible concept is that HBV DNA integration in the first intron of IFITM1 (upstream of the gene) can enhance the expression level of IFITM1, being responsible for the increased transcript levels observed in our study.

These findings shed new light on the potential impact of HBV DNA integration on cellular function. Thus, despite being a replicative dead-end product of HBV replication,⁸ HBV DNA integration can have multiple downstream effects on cell homeostasis, metabolism and the innate antiviral immune response. Regarding integrated portions of the HBV genome, as previously discussed, most HBV DNA integrations involved the HBV genomic region spanning DR1 and DR2; recognised as the recombination proficient region of HBV genome.⁴³ This region also encodes the HBx C-terminus, of which the overexpression has been reported to induce stem cell-like properties,⁴⁴ transformation and inhibition of apoptosis.⁴⁵ This further supports the potential risk of neoplastic transformation of hepatocytes even in patients with low levels of HBV DNA. Notably, the integration of a long portion of viral genome of at least 354 and 1085 nucleotides was detected in two patients. This is consistent with recent data highlighting integrations are not restricted to partial regions, but also involve the full viral genome, further supporting the contribution of integrated HBV DNA to human genomic alterations.⁴⁰

In conclusion, in HBeAg-negative CHB, HBV DNA integration in the whole exome occurs frequently in highly viraemic patients, and is observed in a considerable proportion of patients with low viraemia. Localisation of HBV integrations suggests that this event is not restricted to hepatocarcinogenesis and but can also be implicated in mechanisms regulating antiviral immunity, the inflammatory response and hepatocyte metabolism (figure 6). Our findings underline the clinical challenges associated with the management of HBeAg-negative CHB. In particular, the evidence of HBV integration in low viraemic patients with a limited HBV reservoir is a timely reminder that these patients who do not meet treatment criteria, remain at risk of disease progression and the development of liver cancer.

Author affiliations

¹Department of Experimental Medicine, University of Rome Tor Vergata, Roma, Lazio, Italy

²Barts Liver Centre, Immunobiology, Blizard Institute, Barts and The London School of Medicine & Dentistry, Queen Mary University of London, London, UK

³Department of Microbiology and Virology, University of Milan, Milano, Lombardia, Italy

⁴Molecular Genetics Laboratory, Eurofins GENOMA, Roma, Lazio, Italy

⁵Department of Oncology and Haematology, University of Milan, Milano, Lombardia, Italy

Correction notice This article has been corrected since it published Online First. Figure has been added.

Twitter Upkar S Gill @druppygill and Patrick T F Kennedy @drpkennedy

Acknowledgements The authors are grateful for helpful discussions, advice and critical review of this manuscript from Professor William Mason (Fox Chase Cancer Center) and Dr Carla Usai (Blizard Institute, Barts and The London SMD, QMUL). The authors would also like to thank all patients, their families and the staff at The Royal London Hospital who have supported this work.

Contributors Study concept and design: VS, PTFK. Acquisition of data: VS, RSA, LP, AB, LCo, RSC, USG. Analysis and Interpretation of data: VS, RSA, LP, AB, LCo, RSC, USG, PTFK. Mathematical and statistical analysis: VS, RSA, LCa. Obtained funding: VS, USG, PTFK. Administrative/technical/material/ethics support: VS, RSA, LCa, MS, VC, AN, NH, FC-S, CFP, USG, PTFK. Study supervision: VS, PTFK. Drafting of manuscript:

VS, RSA, USG, PTFK. Critical revision of manuscript: VS, RSA, LP, LCA, AB, LCo, RSc, MS, VC, AN, NH, FC-S, CFP, USG, PTFK.

Funding This work was supported by funding from Italian Ministry of Instruction, University and Research (FIRB project: RBAP11Y57K_001), National Research Council (Progetto Bandiera PB05) to VS, A Wellcome Trust Clinical Research Training Fellowship (107389/Z/15/Z), NIHR Academic Clinical Lectureship (018/064/A), Academy of Medical Sciences Starter Grant (SGL021/1030) and Seedcorn funding Rosetrees/Stoneygate Trust (A2903) awarded to USG; Barts Charity Project Grants (723/1795 and MGU/0406) and an NIHR Research for patient benefit award (PB-PG-0614-34087) to PTFK.

Competing interests PTFK has collaborative grant funding from Gilead, participates in advisory boards/provides consultancy to Gilead, Janssen and is an investigator for industry-led trials with Gilead, Janssen, Alere, Assembly Biosciences, GSK and Roche.

Patient consent for publication Not required.

Ethics approval The study was approved by the local research ethics committee (Brent Research Ethics Committee, reference: 09/H0717/32) and complied with the Declaration of Helsinki. All participants provided written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. All clinical data have been collected in an anonymous database available at The Blizard Institute, Barts and The London SMD, QMUL. Responsible: PTFK

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

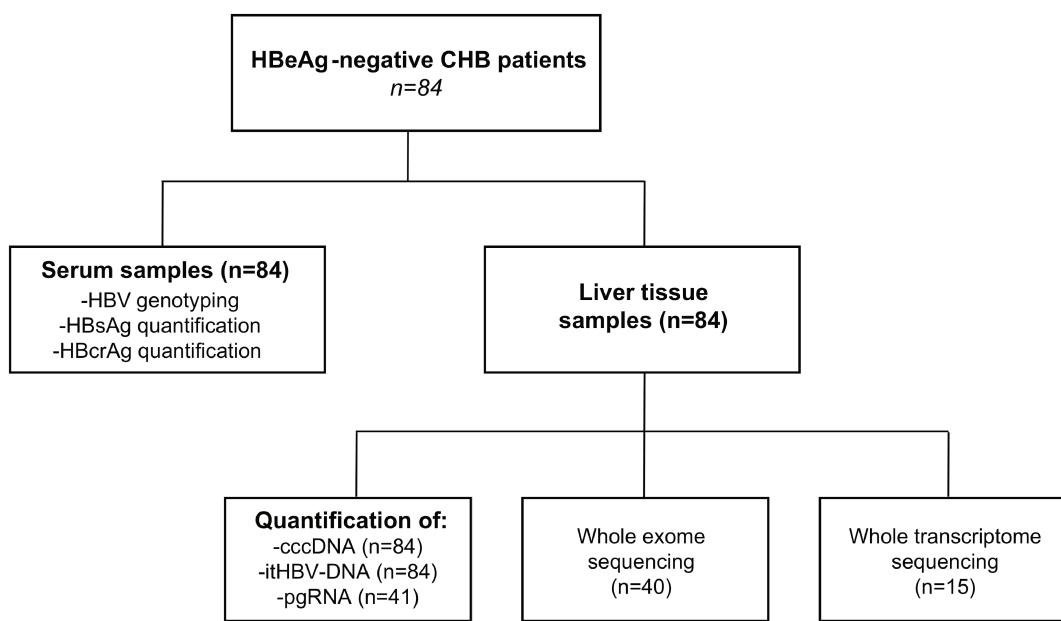
Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Patrick T F Kennedy <http://orcid.org/0000-0001-9201-0094>

REFERENCES

- Lampertico P, Agarwal K, Berg T, et al. EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. *J Hepatol* 2017;67:370–98.
- Liang L-B, Zhu X, Yan L-B, et al. Quantitative intrahepatic HBV cccDNA correlates with histological liver inflammation in chronic hepatitis B virus infection. *Int J Infect Dis* 2016;52:77–82.
- Cheng P-N, Liu W-C, Tsai H-W, et al. Association of intrahepatic cccDNA reduction with the improvement of liver histology in chronic hepatitis B patients receiving oral antiviral agents. *J Med Virol* 2011;83:602–7.
- Höner zu Siederdissen C, Maasoumy B, Cornberg M. New viral biomarkers for hepatitis B: are we able to change practice? *J Viral Hepat* 2018;25:1226–35.
- Testoni B, Leboscé F, Scholtes C, et al. Serum hepatitis B core-related antigen (HBcrAg) correlates with covalently closed circular DNA transcriptional activity in chronic hepatitis B patients. *J Hepatol* 2019;70:615–25.
- Jiang Z, Jhunjhunwala S, Liu J, et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Res* 2012;22:593–601.
- Zhao L-H, Liu X, Yan H-X, et al. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun* 2016;7:12992.
- Tu T, Budzinska M, Shackel N, et al. Hbv DNA integration: molecular mechanisms and clinical implications. *Viruses* 2017;9:75.
- Mason WS, Gill US, Litwin S, et al. Hbv DNA integration and clonal hepatocyte expansion in chronic hepatitis B patients considered immune tolerant. *Gastroenterology* 2016;151:986–98.
- Budzinska MA, Shackel NA, Urban S, et al. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg Microbes Infect* 2018;7:1–12.
- Tu T, Budzinska MA, Shackel NA, et al. Conceptual models for the initiation of hepatitis B virus-associated hepatocellular carcinoma. *Liver Int* 2015;35:1786–800.
- Tu T, Mason WS, Clouston AD, et al. Clonal expansion of hepatocytes with a selective advantage occurs during all stages of chronic hepatitis B virus infection. *J Viral Hepat* 2015;22:737–53.
- Salpini R, Alteri C, Cento V, et al. Snapshot on drug-resistance rate and profiles in patients with chronic hepatitis B receiving nucleos(t)ide analogues in clinical practice. *J Med Virol* 2013;85:996–1004.
- Guo Y, Long J, He J, et al. Exome sequencing generates high quality data in non-target regions. *BMC Genomics* 2012;13:194.
- Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics* 2016;54:1.30.1–1.30.33.
- Thul PJ, Lindskog C. The human protein atlas: a spatial map of the human proteome. *Protein Science* 2018;27:233–44.
- Kanehisa M, Sato Y, Kawashima M, et al. Kegg as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;44:D457–62.
- Adams DR, Eng CM. Next-Generation sequencing to diagnose suspected genetic disorders. *N Engl J Med* 2018;379:1353–62.
- Clark MJ, Chen R, Lam HYK, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 2011;29:908–14.
- Sims D, Sudbery I, Illott NE, et al. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014;15:121–32.
- Su H, Yee JK. Regulation of hepatitis B virus gene expression by its two enhancers. *Proc Natl Acad Sci U S A* 1992;89:2708–12.
- Rawat S, Bouchard MJ. The hepatitis B virus (HBV) HBx protein activates Akt to simultaneously regulate HBV replication and hepatocyte survival. *J Virol* 2015;89:999–1012.
- Tang H, Oishi N, Kaneko S, et al. Molecular functions and biological roles of hepatitis B virus X protein. *Cancer Sci* 2006;97:977–83.
- Oliveri F, Surace L, Cavallone D, et al. Long-Term outcome of inactive and active, low viraemic HBeAg-negative-hepatitis B virus infection: benign course towards HBsAg clearance. *Liver Int* 2017;37:1622–31.
- Bonacci M, Lens S, Mariño Z, et al. Anti-Viral therapy can be delayed or avoided in a significant proportion of HBeAg-negative Caucasian patients in the grey zone. *Aliment Pharmacol Ther* 2018;47:1397–408.
- Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009;106:19096–101.
- Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet* 2010;42:13–14.
- Wang A, Wu L, Lin J, et al. Whole-Exome sequencing reveals the origin and evolution of hepatobiliary carcinomas. *Nat Commun* 2018;9:894.
- Maldarelli F, Wu X, Su L, et al. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* 2014;345:179–83.
- Wagner TA, McLaughlin S, Garg K, et al. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* 2014;345:570–3.
- Yu M-C, Lee C-W, Lee Y-S, et al. Prediction of early-stage hepatocellular carcinoma using OncoScan chromosomal copy number aberration data. *World J Gastroenterol* 2017;23:7818–29.
- Falcon T, Freitas M, Mello AC, et al. Analysis of the cancer genome atlas data reveals novel putative ncRNAs targets in hepatocellular carcinoma. *Biomed Res Int* 2018;2018:1–9.
- Beck M, Schirmacher P, Singer S. Alterations of the nuclear transport system in hepatocellular carcinoma – new basis for therapeutic strategies. *J Hepatol* 2017;67:1051–61.
- Fu L, Jiang Z, Li T, et al. Circular RNAs in hepatocellular carcinoma: functions and implications. *Cancer Med* 2018;7:3101–9.
- Noda D, Itoh S, Watanabe Y, et al. Elac2, a putative prostate cancer susceptibility gene product, potentiates TGF-β/Smad-induced growth arrest of prostate cells. *Oncogene* 2006;25:5591–600.
- Murugan AK, Grieco M, Tsuchida N. Ras mutations in human cancers: roles in precision medicine. *Semin Cancer Biol* 2019;59:23–35.
- Khan AQ, Kuttikrishnan S, Siveen KS, et al. Ras-Mediated oncogenic signaling pathways in human malignancies. *Semin Cancer Biol* 2019;54:1–13.
- Torresi J, Tran BM, Christiansen D, et al. Hbv-Related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. *BMC Cancer* 2019;19:707.
- Budzinska MA, Shackel NA, Urban S, et al. Cellular genomic sites of hepatitis B virus DNA integration. *Genes* 2018;9:365.
- Ramirez R, Van Buuren N, Suri V, et al. Targeted long read sequencing reveals the comprehensive architecture and expression patterns of integrated HBV DNA in Chb liver biopsies. *J Hepatol* 2020;73:S6–7.
- Yano M, Ohkoshi S, Aoki Y-hei, Aoki Y, et al. Hepatitis B virus X induces cell proliferation in the hepatocarcinogenesis via up-regulation of cytoplasmic p21 expression. *Liver Int* 2013;33:1218–29.
- Liang H-W, Wang N, Wang Y, et al. Hepatitis B virus-human chimeric transcript HBx-LINE1 promotes hepatic injury via sequestering cellular microRNA-122. *J Hepatol* 2016;64:278–91.
- Hino O, Ohtake K, Rogler CE. Features of two hepatitis B virus (HBV) DNA integrations suggest mechanisms of HBV integration. *J Virol* 1989;63:2638–43.
- Ng K-Y, Chai S, Tong M, et al. C-Terminal truncated hepatitis B virus X protein promotes hepatocellular carcinogenesis through induction of cancer and stem cell-like properties. *Oncotarget* 2016;7:24005–17.
- Ma N-F, Lau SH, Hu L, et al. CooH-Terminal truncated HBV X protein plays key role in hepatocarcinogenesis. *Clin Cancer Res* 2008;14:5061–8.

Figure S1

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

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'-ACCTCTCTTACGCGG-3'; Reverse, 5'-ACAGCTTGGAGGCTTGAA-3' and Probe, 5'-6FAM-CTCCCCGTCTGTGCCTCTCATC-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 10⁶ 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'-CTTCGGAGTGTGGATTG-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™ 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	<i>GGCATAAATTGGCTCGGCC</i>	<u>CAGGGGAGCTGAGTGTTGAG</u>	GCACACAGGCCTCAGTCACGTGC
40	IFITM1	<u>GAGAGGAGATGGTGAGGGGA</u>	<i>GGAGACTCTAAGGCTTCCCG</i>	CCAGGATCCCCAGCTTTGCCT
40	IFITM1	<i>TGTCAACGACCGACCTTGAG</i>	<u>GGAGGTGGTCCCTGATCTCA</u>	AGACTGGGAGGAGTTGGGGAGGAG
44	NUP85	<u>GCCAAGTTGCCCTTATCT</u>	<i>CCCGTAAAGTTCCCACCTT</i>	GGGCCTCCAGAGAATTGACG
62	ANKRD52	<i>TGTGGGTCTGGAAGTAATGA</i>	<u>GCTCAATAGCCACAGCATCC</u>	GCACATCGCCTGCTACCTGGGCC
62	ANKRD52	<i>CCGACCTTGAGGCCTACTTC</i>	<u>GGCTACATGTGCTGGCTTCT</u>	AGGACTGGGAGGAGTTGGGGAGG
75	AGBL5	<i>ACTAGGAGGCTGTAGGCATAA</i>	<u>GTTGCCAGCCTCCATCCAT</u>	GGTCTGCGCACAGCACCATGCA
83	NR3C1	<i>ACGACCGACCTTGAGGCATA</i>	<u>ATGATAGCTCTGTTCCAGACTCAA</u>	GGAGGAGTTGGGGAGGAGA
84	CYP2U1	<i>GACCGACCTTGAGGCATACT'</i>	<u>AAACCGGAGGATTCTACCC</u>	TCCTGGTCATCCAGAAATCGA

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: $\text{TPM} = A \frac{1}{\sum(A)} 10^6$, where A is $\frac{\text{total reads mapping to a specific gene} * 10^3}{\text{gene length in bp}}$ [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>.

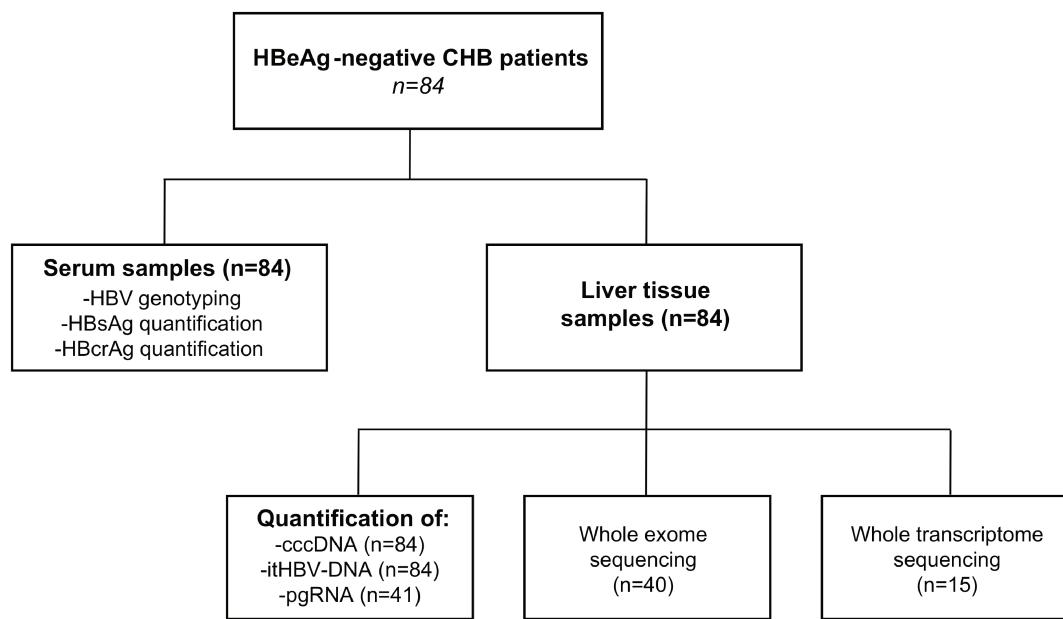
Supplementary Table 1. Diagnostic performance of serological markers in identifying a limited intrahepatic reservoir

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%)

Figure S1

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

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'-ACCTCTCTTACGCGG-3'; Reverse, 5'-ACAGCTTGGAGGCTTGAA-3' and Probe, 5'-6FAM-CTCCCCGTCTGTGCCTCTCATC-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 10⁶ 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'-CTTCGGAGTGTGGATTG-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™ 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	<i>GGCATAAATTGGCTGCGCC</i>	<u>CAGGGGAGCTGAGTGTTGAG</u>	GCACACAGGCCAGTCACGTGC
40	IFITM1	<u>GAGAGGAGATGGTGAGGGGA</u>	<i>GGAGACTCTAAGGCTTCCCG</i>	CCAGGATCCCCAGCTTTGCCT
40	IFITM1	<i>TGTCAACGACCGACCTTGAG</i>	<u>GGAGGTGGTCCCTGATCTCA</u>	AGACTGGGAGGAGTTGGGGAGGAG
44	NUP85	<u>GCCAAGTTGCCCTTATCT</u>	<i>CCCGTAAAGTTCCCACCTT</i>	GGGCCTCCAGAGAATTGACG
62	ANKRD52	<i>TGTGGGTCTGGAAGTAATGA</i>	<u>GCTCAATAGCCACAGCATCC</u>	GCACATGCCCTGCTACCTGGGCC
62	ANKRD52	<i>CCGACCTTGAGGCCTACTTC</i>	<u>GGCTACATGTGCTGGCTTCT</u>	AGGACTGGGAGGAGTTGGGGAGG
75	AGBL5	<i>ACTAGGAGGCTGTAGGCATAA</i>	<u>GTTGCCAGCCTCCATCCAT</u>	GGTCTGCGCACAGCACCATGCA
83	NR3C1	<i>ACGACCGACCTTGAGGCATA</i>	<u>ATGATAGCTCTGTTCCAGACTCAA</u>	GGAGGAGTTGGGGAGGAGA
84	CYP2U1	<i>GACCGACCTTGAGGCATACT'</i>	<u>AAACCGGAGGATTCTACCC</u>	TCCTGGTCATCCAGAAATCGA

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: $\text{TPM} = A \frac{1}{\sum(A)} 10^6$, where A is $\frac{\text{total reads mapping to a specific gene} * 10^3}{\text{gene length in bp}}$ [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%)