OPEN ACCESS

Original research

Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma

Camille Péneau ,[1,2] Sandrine Imbeaud ,[1,2] Tiziana La Bella,[1,2] Theo Z Hirsch ,[1,2] Stefano Caruso,[1,2] Julien Calderaro,[3] Valerie Paradis,[4,5] Jean-Frederic Blanc,[6,7,8] Eric Letouzé,[1,2] Jean-Charles Nault ,[1,2,9] Giuliana Amaddeo,[10] Jessica Zucman-Rossi [1,2,11]

► Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2020-323153).

For numbered affiliations see end of article.

Correspondence to
Professor Jessica Zucman-Rossi, Centre de Recherche des Cordeliers, Sorbonne Université, Inserm, Université de Paris, INSERM, Paris 75006, France; jessica.zucman-rossi@inserm.fr

Received 21 September 2020
Revised 8 January 2021
Accepted 11 January 2021

To cite: Péneau C, Imbeaud S, La Bella T, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2020-323153

ABSTRACT

Objective Infection by HBV is the main risk factor for hepatocellular carcinoma (HCC) worldwide. HBV directly drives carcinogenesis through integrations in the human genome. This study aimed to precisely characterise HBV integrations, in relation with viral and host genomics and clinical features.

Design A novel pipeline was set up to perform viral capture on tumours and non-tumour liver tissues from a French cohort of 177 patients mainly of European and African origins. Clonality of each integration event was determined with the localisation, orientation and content of the integrated sequence. In three selected tumours, complex integrations were reconstructed using long-read sequencing or Bionano whole genome mapping.

Results Replicating HBV DNA was more frequently detected in non-tumour tissues and associated with a higher number of non-clonal integrations. In HCC, clonal selection of HBV integrations was related to two different mechanisms involved in carcinogenesis. First, integration of viral enhancer nearby a cancer-driver gene may lead to a strong overexpression of oncogenes. Second, we identified frequent chromosome rearrangements at HBV integration sites leading to cancer-driver genes (*TERT, TP53, MYC*) alterations at distance. Moreover, HBV integrations have direct clinical implications as HCC with a high number of insertions develop in young patients and have a poor prognosis.

Conclusion Deep characterisation of HBV integrations in liver tissues highlights new HBV-associated driver mechanisms involved in hepatocarcinogenesis. HBV integrations have multiple direct oncogenic consequences that remain an important challenge for the follow-up of HBV-infected patients.

Significance of this study

What is already known on this subject?

- HBV infection is the main risk factor for hepatocellular carcinoma (HCC) worldwide.
- HBV is a small DNA virus which can integrate in the human genome and thus play a role in promoting carcinogenesis.
- Studies using next-generation sequencing on liver tumours from Asian patients pointed out that HBV promotes HCC development through insertional mutagenesis.

What are the new findings?

- Replicating HBV DNA was more frequently detected in non-tumour tissues, and associated with a higher number of integrations.
- HCC development through HBV insertional mutagenesis is linked to the presence of a viral enhancer in the integrated sequence in the proximity of a cancer-driver gene.
- HBV-induced carcinogenesis can also be driven by frequent copy number alterations of cancer-driver genes associated with distant viral integrations.
- The number of HBV integrations is an independent prognostic factor in HBV-related HCC.

INTRODUCTION

Despite the implementation of vaccination programmes and the strong decrease of new infections among children, the percentage of people living with chronic HBV infection worldwide remained as high as 3.5% of the global population in 2015.[1] Moreover, only 9% of infected people were diagnosed and only 8% of them were under treatment, which makes hepatitis B a major public health threat now prioritised by WHO.[1 2] The burden of HBV infection is closely related to the development of hepatocellular carcinoma (HCC), the most frequent primary liver cancer and the fourth cause of cancer death worldwide.[3] Persistent HBV infection is actually responsible for >50% of all HCC cases worldwide and up to 85% in some areas where the infection is endemic.[4] Even patients receiving antiviral treatments who have maintained viral suppression remain at risk of developing an HCC.[5–8]

HBV-related HCC occurs in the setting of cirrhosis and in normal liver,[9] underlying that the virus has its own oncogenic properties besides the induction of chronic inflammation in the liver. HBV is a small 3.2 kb DNA virus that can integrate in human DNA and promote cell transformation by insertional mutagenesis or through expression of

Significance of this study

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

- A high number of HBV integrations is associated with viral replication in non-tumour liver tissues and with poor prognosis in tumours, showing the importance of providing efficient antiviral therapy early during life to limit the number of HBV integrations and fight against direct HBV-related HCC development.
- This study underlines that HBV-related HCC are promoted by distinct mechanisms, highlighting the heterogeneity of these tumours and the importance of molecular characterisation to identify new specific therapeutic opportunities.

viral oncoproteins such as the protein HBx.[10–12] As HBV integrations occur early during infection,[13] comparing the insertions occurring in normal hepatocytes and in tumour cells may enable to identify new genomic defects associated with tumour development.

The development of next-generation sequencing led to a more precise characterisation of the role of HBV insertions in HCC initiation and development.[14–18] Cancer-related genes such as *TERT*, *CCNE1* and *KMT2B* have been identified as recurrently targeted by HBV insertions in HCC, with specific functional consequences and clinical outcomes.[16 19–21] Thus, the presence of integrated HBV DNA in hepatocytes has been pointed out as a driver of hepatocarcinogenesis through the alteration of the expression or function of these genes.[12] But up to now, such identifications of HBV insertions in HCC have been mainly performed in Asian populations where HBV genotypes B and C are the most prevalent HBV strains.[22] Furthermore, as a significant proportion of HBV-related HCC does not contain any integration in a known cancer-related gene, two other independent mechanisms involving integrated HBV DNA have been proposed: induction of chromosomal instability and persistent expression of altered HBV genes.[23] However, the interactions between the three mechanisms mentioned and the precise role they might play individually or synergistically in HCC initiation and development remain to be fully elucidated.

Therefore, in this work, we developed a reliable pipeline of analysis based on viral capture, to characterise HBV integrations in a cohort of 177 patients, together with genome rearrangements, clinical features and viral characteristics in tumours and their corresponding non-tumour liver tissues. We also investigated the oncogenic consequences of recurrent HBV integrations on gene expression and chromosomal instability, involved in hepatocarcinogenesis.

MATERIALS AND METHODS

(Extended in online supplemental materials and methods section).

A series of tissue samples containing 1128 frozen HCC and 1063 non-tumour counterparts from 1128 patients was assembled (online supplemental figure 1). Clinical and serological data were collected from each centre (online supplemental table 1). Viral DNA screening was performed according to protocols previously described[24] with specific HBV probes (online supplemental table 2).

HBV viral capture was performed for 177 frozen HCC and 170 matched non-tumour liver tissues using single-stranded biotinylated probes (SeqCap EZ Designs, Roche NimbleGen, Madison, Wisconsin, USA) to target 40 references distributed on the eight main HBV genotypes and four human regions (*TERT* promoter, *CCNE1*, *CCNA2* and *KMT2B*; online supplemental table 3). DNA libraries of 1 kb fragments were prepared and multiplexed before following the double capture steps with SeqCap-EZ-HyperCap Workflow (Roche). Sequencing was performed with Illumina MiSeq instrument to produce paired-end reads of 2×250 nucleotides.

Analysis of integration events was performed following a pipeline described in a flowchart (online supplemental figure 2). Briefly, sequencing data were aligned on HBV and human references (online supplemental table 3). HBV copy number per cell was evaluated from the mean read coverage of the HBV genome, including integrated and episomal HBV (online supplemental figure 3A). Paired and chimeric reads were extracted using htslib package to identify HBV/human junction breakpoints and assess clonality of integration events. The 8809 integration events were classified in three categories using the k-means method: (1) 'clonal integrations' (>48% of cells), (2) 'unique integrations' (<3% of cells) and (3) 'subclonal integrations' (3%–48% of cells) (online supplemental figure 3B). For 20 pairs of tumours and non-tumour tissues, integration events identified with viral capture and whole genome sequencing (WGS) were comparable (online supplemental figure 3C,D). Integration breakpoints were annotated for replication timing in HepG2 (ENCODE),[25] top 20% highly expressed genes in HBV-positive livers, common and rare fragile sites,[26] repetitive sequences and CpG islands and chromatin structure in adult liver (ROADMAP).[27] Insertions were named 'classic' if the two breakpoints clustered within 25 kb in opposite direction, 'local inversion' if two breakpoints clustered within 25 kb in the same direction, 'large rearrangements' if only one breakpoint was identified and 'cluster of insertions' if more than two breakpoints were identified.

Genomic DNA sequencing of driver genes were performed in 265 tumours and their adjacent liver tissues with high coverage WGS (n=62) or whole exome sequencing (WES, n=203) and re-sequenced at *TERT* promoter with Sanger or MiSeq.

Detection of HBV episomal form was performed with specific DNAse/TaqMan-based assay[24] in 162 tumours and 155 non-tumour liver tissues (online supplemental table 2).

Viral mRNA and specific genes mRNA screening was performed by quantitative reverse transcriptase (qRT)-PCR using 10 probe sets covering 8 HBV regions from the main HBV genotypes, and probe sets to detect HCV, HDV or TERT expression (online supplemental table 2).

HBV replicative forms were considered as present in one tissue if both HBV episomal DNA and HBV pregenomic RNA (pgRNA) forms were identified. When episomal forms were detected in the absence of pgRNA, the tissue was considered as containing 'episomal not replicative HBV DNA'.

RNA-sequencing (RNA-seq) and transcriptomic analyses were performed in 265 tumours (130 HBV-positive and 135 HBV-negative) and 24 HBV-positive non-tumour tissues using Illumina TruSeq or Illumina TruSeq Stranded mRNA kit on HiSeq2000 by IntegraGen, Evry, France.[19] We used the Bioconductor limma package[28] to test for differential expression in all expressed genes with an in-house adaptation of the gene set enrichment analysis (GSEA) method.[29] Tumours were classified in G1-G6 transcriptomic groups as previously described by qRT-PCR of 190 genes using BioMark PCR system.[30 31]

HBV variants analysis from viral capture was performed according to nomenclature previously described.[22 32]

Cell culture, transfection and dual luciferase assay were performed in HuH7 cells purchased from the American Type Culture Collection as previously described.[33]

Long-read sequencing were performed in three HBV-positive tumours (#1151T, #1597T, #1994T) using the Single Molecule Real Time technology, PacBio-SMRTcell system (ICGex platform, Curie Institute, France) and the Consensus Circular Sequencing algorithm (Pacific Biosciences). *Whole genome mapping* was performed to analyse the same samples by Bionano Genomics, La Jolla, California, USA.

Statistical analyses were performed with R V.3.6.0 (http://www.R-project.org), various statistical tests were applied with respect to the type of variable. Survival analysis was performed in patients treated for a primary HCC tumour by R0 liver resection as previously described.[34]

RESULTS

HBV integrations occur in open chromatin regions and relate to viral replication in the liver

We performed HBV viral capture on 177 HCC and 170 non-tumour liver tissues from 177 HBV-positive patients to analyse HBV integrations in the human genome with a new bioinformatics pipeline and identify precisely the structure of insertion events and their clonality (see 'Materials and methods' section and online supplemental table 1). In 84% of non-tumour liver tissues (143 out of 170), we identified 6610 HBV integration breakpoints at HBV/human junctions corresponding to unique (82%), subclonal (17%) or clonal (1%) events, involving similarly all HBV genotypes (figure 1A; online supplemental figure 4; online supplemental table 4). In the human genome, HBV integration breakpoints were enriched in active and open chromatin regions,

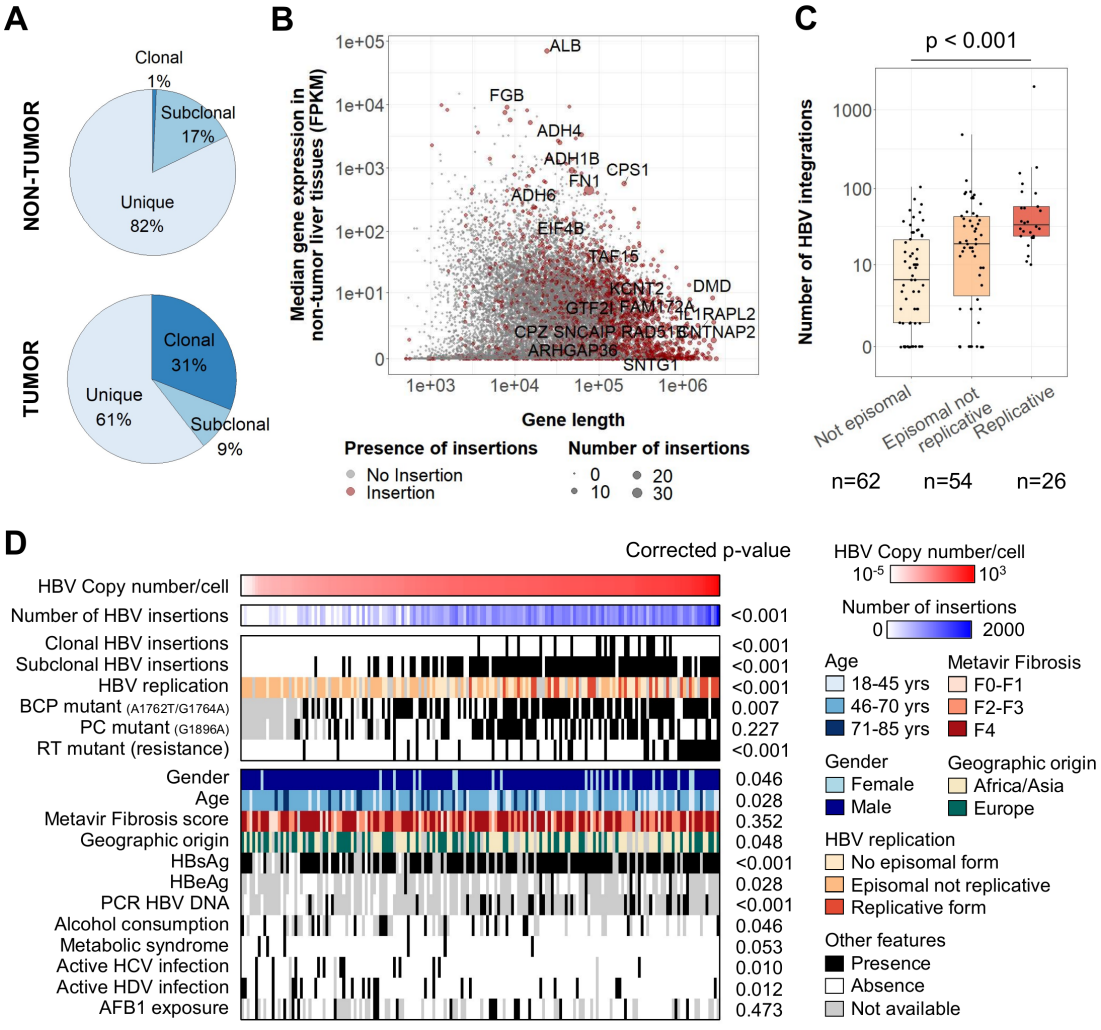


Figure 1 HBV integrations in non-tumour tissues are associated with viral replication and more frequent in large highly expressed genes. (A) Repartition of clonality between all HBV integration events detected in viral capture (n=8809). (B) Coding genes with or without HBV integrations according to the gene length and the median gene expression in non-tumour liver tissues. Genes with recurrent clonal or subclonal HBV integrations are annotated. (C) Number of HBV integrations identified in non-tumour samples (n=142) according to the presence of episomal HBV DNA and replicative HBV DNA (Jonckheere's trend test). (D) Correlations between HBV copy number/cell in 170 non-tumour liver tissues assessed by viral capture and clinical or molecular features. Positivity for hepatitis B surface antigen (HBsAg), hepatitis B e antigen (HBeAg), HBV DNA (by PCR) in the patients' serum and duration of antiviral treatments were obtained from clinical data. Wilcoxon signed-rank, Kruskal-Wallis or Pearson's correlation statistical tests were applied with respect to the type of variable. P values were adjusted for multiple testing using the Benjamini-Hochberg method (false discovery rate). AFB1, aflatoxin-B1; BCP, basal core promoter; FPKM, fragments per kilobase of exons per million reads; PC, PreCore; RT, reverse transcriptase.

and they were close to early replicated and highly expressed genes. Integrations were also more frequent in simple repeats, in particular in telomeric and subtelomeric regions, but not in Alu motifs, fragile sites or long/short interspersed nuclear elements (LINE/SINE; online supplemental figure 5A). Reconstruction of a selection of 394 events of subclonal or clonal integrations in non-tumour tissues showed that they all contained a large part of the whole HBV sequence (median size=2912 nucleotides) and 42% of them were bordered by 1-to-8 nucleotide homology between the targeted human sequence and the integrated HBV sequence, suggesting the involvement of microhomology-mediated end joining (online supplemental figure 5B,C).

In the non-tumour liver tissues, recurrent clonal/subclonal insertions were identified in 20 genes with the following characteristics: highly expressed (7/20, fragments per kilobase of exons per million reads >100, p<0.001) or very large genes (11/20, >100 kb, p<0.001), such as *FN1* (14 cases), *CPS1* (4 cases), *KCNT2* or *ADH1B* (3 cases), *ADH4 or ADH6 or ALB* (2 cases, figure 1B). Including all insertion events in *FN1*, 78 HBV breakpoints were distributed all over the gene, mainly in introns (72/78) and without specific hotspots (online supplemental figure 6A). RNA-seq analysis of 26 HBV integrations in the *FN1* locus identified two major types of fusion transcripts: (1) out-of-frame HBx-FN1 transcripts in the same or in the opposite direction (9/26) and (2) in-frame HBs-FN1 transcripts generated by a cryptic splice site at position 458 in the S ORF (11/26; online supplemental figure 6A). Overall, the diversity of the fusion transcripts HBs-FN1 starting at different exons of *FN1* and the global lack of overexpression of the genes targeted by HBV integrations indicated that most of the clonal/subclonal insertions in non-tumour liver tissues did not argue for a functional effect (online supplemental figure 6B). Interestingly, clonal insertions suggesting a local proliferation of non-neoplastic hepatocytes were observed in 11% of the 170 non-tumour samples, all in fibrotic livers (F2-F4), with a decreased inflammatory response and decreased nuclear factor-κB/tumour necrosis factor signalling, detected by RNA-seq of a subset of non-tumour tissues (online supplemental figure 6C).

In 170 analysed non-tumour liver tissues, high HBV copy number per cell was significantly associated with female gender, young age, African or Asian geographic origin and positivity for hepatitis B surface antigen (HBsAg), hepatitis B e antigen (HBeAg) and HBV DNA in the serum. In contrast, presence of cofactors of chronic liver disease such as active HCV or HDV infection, alcohol consumption or metabolic syndrome was associated with lower copy number of HBV and less HBV integrations. Only 18% of the samples contained replicative HBV, they were enriched in genotype A and showed a higher number of integrations, underlying that HBV replication and integration were linked processes (figure 1C,D; online supplemental figure 4). Also, samples with high HBV copy number showed frequent mutations of the basal core promoter (BCP; A1762T/G1764A) or of the RT region: mutations known to favour viral replication[32] (figure 1D).

Complex HBV integrations and viral genome rearrangements are frequent in HCC

In 88% of 177 analysed HBV-related HCC, we identified HBV integrations in human genome totalising 2199 breakpoints at HBV/human junctions, and 31% of them corresponded to clonal events (figure 1A; figure 2A). A vast majority (82%) of tumours harboured at least one clonal HBV integration, compared with only 11% in non-tumour liver tissues (p<0.001, figure 2B). Overall, the number of HBV integrations was highly correlated with the HBV copy number per cell (figure 2C). However, whereas tumours showed a higher HBV copy number per cell, they harboured a lower number of HBV breakpoints per sample than in their corresponding non-tumour tissues (mean 12 vs 39, p<0.001) (figure 2D; online supplemental figure 7). This may reflect the high diversity of unique HBV integrations in a large number of hepatocytes during the infection, and the decrease in diversity induced by clonal expansion of transformed cells. In both types of tissues, the most frequent hotspot of HBV integration was observed in the C-terminal region of the HBx gene around the DR1 sequence (1817–1836) corresponding to the ends of the double-strand linear DNA form of HBV (figure 2E). In tumours, only a part of the HBV genome was detected, corresponding to integrated HBV DNA with frequent truncation of the HBx gene (online supplemental figure 8A). In addition, HBV genomes in tumours contained a higher number of structural variants within the viral sequence (deletions, duplications, inversions; figure 2F) and only 3% of these structural variants in tumours were observed in the adjacent liver tissues (17/514). Finally, the localisation and orientation of HBV/human junctions in the human genome showed frequent chromosome rearrangements in tumours, suggesting a more complex integration process or postintegration structural modifications than in non-tumour liver tissues (online supplemental figure 8B).

HBV replication was less frequent in tumours (7%) than in non-tumours (18%, p<0.001) (figure 2G). In addition, analysis of HBV transcripts revealed that even in tumours containing an episomal form of the virus, truncated HBx transcripts derived from integrated HBV DNA were predominant over complete transcripts, suggesting that episomal forms in tumours account for only a small proportion of HBV DNA (online supplemental figure 8C,D). Moreover, the majority of tumours with clonal HBV integrations showed a high mRNA expression level in the S region (supplementary figure 8E). Finally, by analysing the variant allele frequency of the HBV variants in the viral genome, the same major genotype was identified in the tumours and their corresponding non-tumour liver tissues in 138 out of 140 cases. Surprisingly, we observed a negative selection of HBV variants affecting HBeAg production (A1762T and G1764A BCP variants, G1896A PC variant) and antiviral resistance-associated mutations (RT region) in the tumours, suggesting that emergence of variants in HBV sequence aimed to improve the virus fitness but did not give specific advantage for tumour development (online supplemental figure 8F). Overall, these results suggest that HBV integrations in non-tumour and tumour liver tissues may reflect different viral dynamics and selection processes not directly correlated.

Human chromosome rearrangements were recurrently delimited by HBV integrations

Among 504 clonal integrations identified from the capture, WES or WGS sequencing of 121 HCC, 179 events (36%) precisely matched boundaries of chromosome copy number alterations (CNA) in the human genome, showing a direct relationship between viral insertion events and chromosome structural rearrangements (figure 3A). CNA-associated integrations were clonally selected, suggesting that these specific alterations are positively selected and constitute cancer driver events (figure 3A). Indeed, three major types of CNA bordered by HBV integrations were observed in >10 samples: (1) large deletions of the chromosome 17p including *TP53* (15 tumours) and frequently associated with centromeric insertions (8/15),

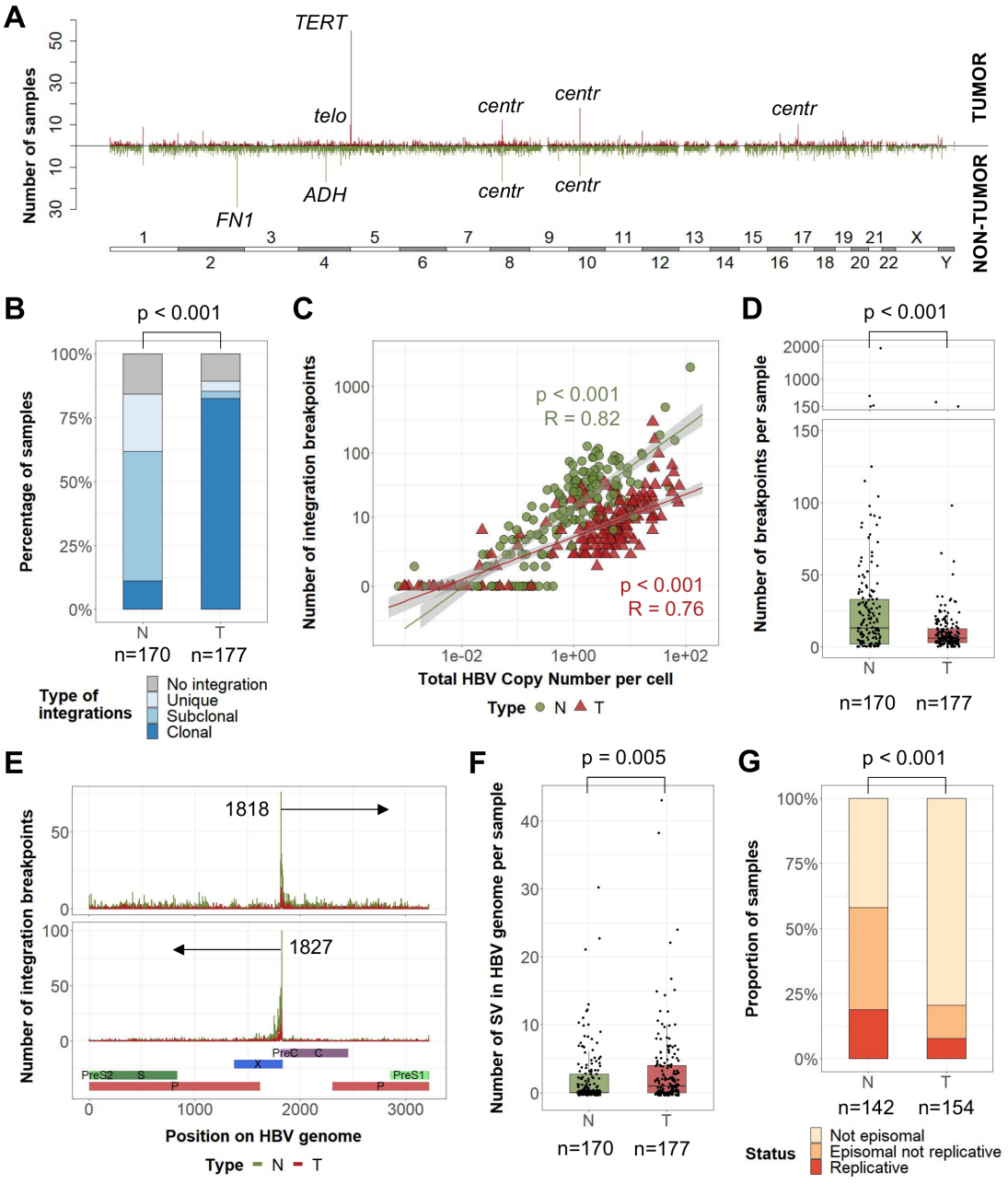


Figure 2 Tumours and non-tumour tissues have different HBV integration profiles and structures of HBV sequences. (A) Pan-genomic view of genomic locations of all HBV integration breakpoints in tumours (up) or non-tumour tissues (down) in 347 HBV-positive samples (non-tumour liver, n=170, and tumour samples, n=177). A line corresponds to a 1M-bin region. (B) Proportion of non-tumour tissues and tumours harbouring only unique integrations, at least one subclonal integration or at least one clonal integration (χ^2 test). (C) Correlation between the HBV copy number per cell and the number of HBV integration breakpoints (Pearson's correlation). (D) Number of HBV integration breakpoints per sample (Wilcoxon signed-rank test). (E) Localisation of HBV integration breakpoints along the HBV genome in tumours and non-tumour samples according to the orientation of the integrated sequence. (F) Number of structural variants in HBV genome per sample (Wilcoxon signed-rank test). (G) Proportion of non-tumour and tumour samples containing replicative HBV DNA, episomal non-replicative HBV DNA or no HBV episomal form (χ^2 test). centr, centromeric; SV, structural variant; telo, telomeric.

(2) focal gains of *TERT* at 5p (14 tumours) and (3) large gains at chromosome 8q including *MYC* (13 tumours) and related to centromeric insertions (4/13) (figure 3B; see examples #1733T and #1597T in figure 3C,D).

To investigate how integrations and chromosome alterations were associated, we analysed three selected cases with long-read sequencing (PacBio) and Bionano mapping to generate genomic assemblies and investigate rearrangements at a larger scale. In tumour #1151T, we identified two different HBV integrations on chromosome 8q that together with a whole duplication of the human genome resulted in a large eight copies amplification of 57 Mb that included *MYC* oncogene. We reconstructed the two distinct integration events: (1) HBV integration resulted in an inter-chromosome t(8q21;17p11) translocation inducing a large 17p deletion and (2) HBV integration located in the 8q centromeric region bordering a duplication-inverted-like event potentially reflecting an isochromosome iso(8q) (online supplemental figure 9A). In tumour #1597T (figure 3D), we identified a *TERT*

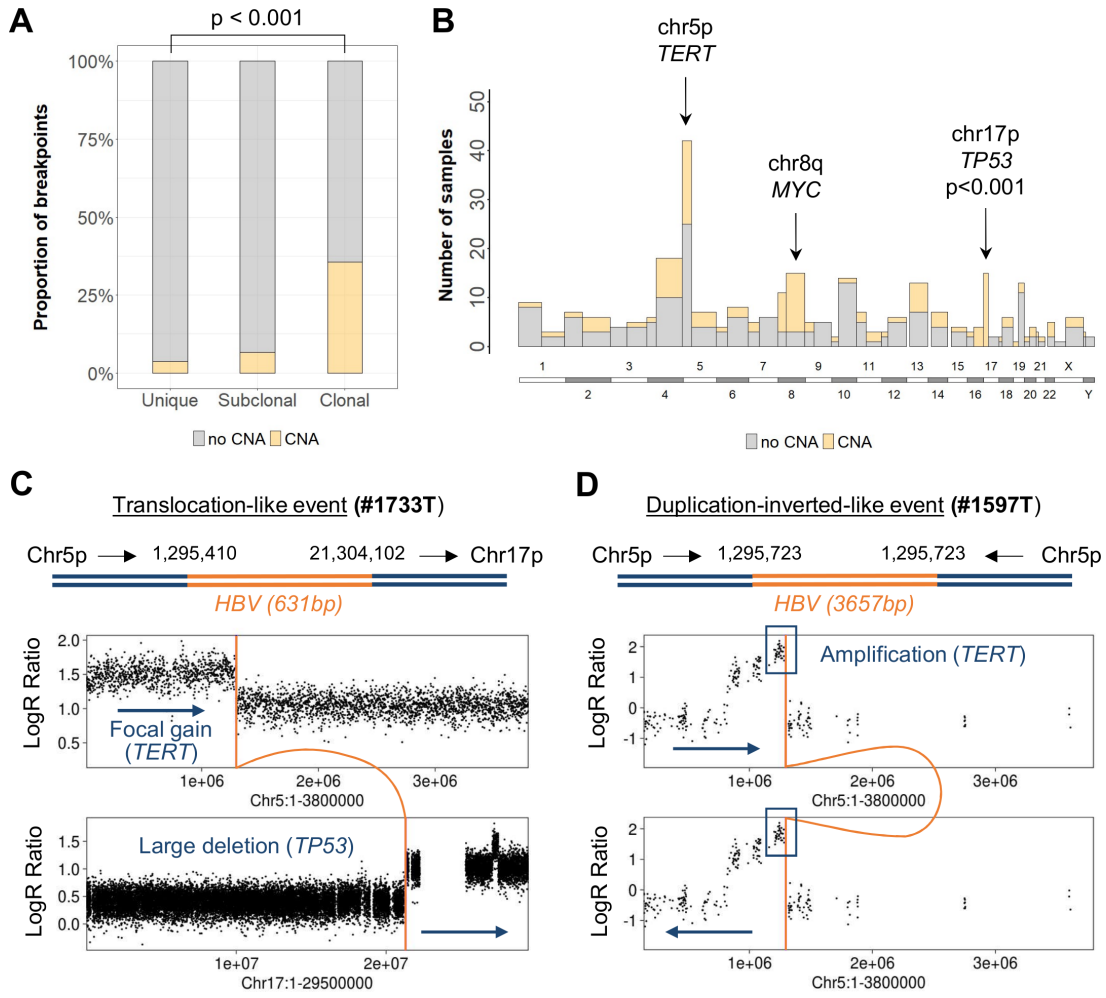


Figure 3 HBV integrations induce chromosomal rearrangements and distant driver oncogenic alterations. (A) Proportion of HBV integration breakpoints (n=1436) associated with a copy number alteration (CNA) in 121 HBV-positive tumours (χ^2 test). (B) Pan-genomic view of the number of HBV integration breakpoints according to their association with a CNA, split by chromosome arms. The three genomic regions containing the higher number of HBV integrations associated with CNA are annotated. Fisher's exact tests were performed to compare the number of integrations with or without CNA and p-values were adjusted for multiple testing. (C) Translocation-like event in tumour #1733T: HBV integration is associated with a focal gain on chr5p and a large deletion on chr17p. (D) Duplication-inverted-like event in tumour #1597T, reconstructed with long-read sequencing: HBV integration is associated with a focal amplification including *TERT*.

amplification at eight copies associated with multiple structural variants including two HBV integrations located within chromosome 1q (online supplemental figure 9B).

Surprisingly, even if HCC with integration-associated rearrangements had a significantly higher number of both HBV integrations and CNA in their genome, these two features were not correlated (online supplemental figure 10A), meaning that only a subset of the viral integrations were associated with chromosome rearrangements, in particular integrations located at centromeric or telomeric regions. Interestingly, centromeric integrations were enriched in young patients with African origin (online supplemental figure 10B); telomeric HBV integrations usually occurred directly or in proximity to telomere repeats (eg, #4229T and #4268T in online supplemental figure 10C) and were enriched in tumours with HBV replication and a high number of CNA (online supplemental figure 10A). As shown in tumours #1994T using Bionano sequencing, telomeric HBV integration can support the translocation of a large chromosome region at the extremity of another chromosome (online supplemental figure 10D). Overall, 41% of clonal HBV integrations identified in tumours were involved in large rearrangements of the human genome and 48% of them were associated with CNA. Functionally, integration-associated rearrangements frequently altered a cancer driver gene, either at distance when associated with recurrent *TP53* or *MYC* alterations, either locally with *TERT* alterations.

HBV genomes can integrate in cancer driver genes with *cis*-activating consequences

In tumours, HBV integrations followed a different distribution along the human genome when compared with non-tumour tissues and HBV copy number/cell was associated with the number of clonal events (figure 2A; figure 4A). These clonal insertions were enriched around cancer driver genes, suggesting a strong functional selection of cells containing such events (online supplemental figure 11A). However, among the 229 genes harbouring clonal integrations in their vicinity (±25 kb), only three genes were recurrently found in more than two HCC: *TERT* (n=48), *CCNE1* (n=4), *KMT2B* (n=3). Four other genes had HBV integrations in two samples: *AHRR*, *NRG1*, *TRIM16L* and *ST18* (figure 2A; online supplemental table 4).[35]

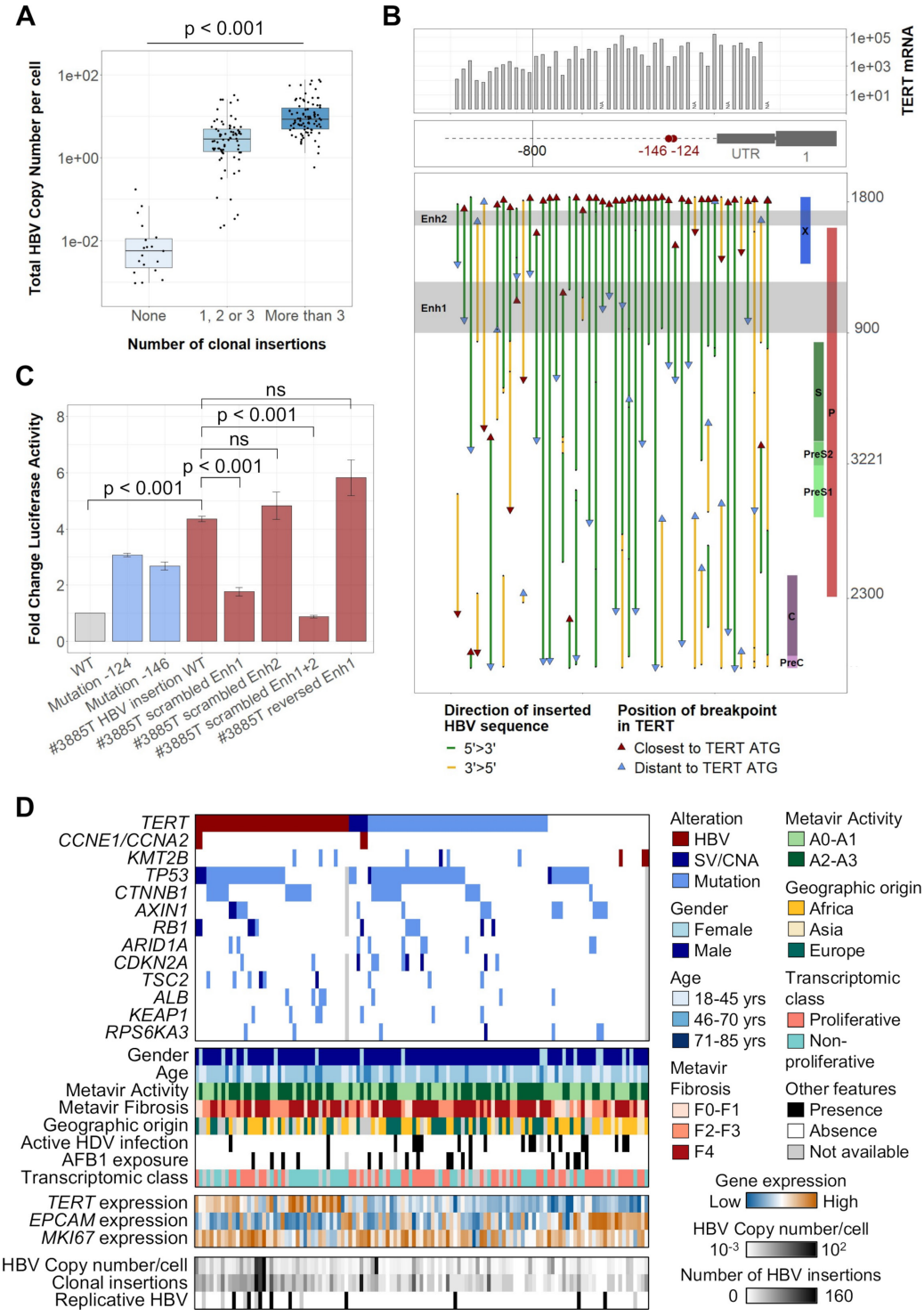


Figure 4 HBV integrations in the *TERT* promoter induce a strong activation of *TERT* promoting hepatocellular carcinoma (HCC) development. (A) HBV copy number/cell in 177 tumours from the capture series according to their number of clonal HBV integration breakpoints (Jonckheere's trend test). (B) Integrated HBV sequences located in the *TERT* promoter in 48 tumours harbouring clonal HBV integrations at this locus. The position of the breakpoints, of the viral enhancer regions and the orientation of the integrated sequences are annotated along the HBV genome. mRNA expression (-ddCT) from qRT-PCR data are represented above. (C) The impact of HBV integration in the *TERT* promoter was evaluated using promoter luciferase assays in Huh7 liver cell lines. Constructs of *TERT* promoter containing different HBV-integrated sequences with or without scrambled enhancer regions were compared with the WT promoter and to the *TERT* promoter with mutations at the −124 or −146 hotspots. Error bars correspond to SD of three independent transfections for each plasmid (Student's t-test). (D) Molecular profile of 121 HBV-positive tumours with alterations in 14 HCC-associated genes (HBV integration, SV/CNA or mutation), according to clinical and other molecular features. mRNA expression of *TERT*, *EPCAM* and *MKI67* were obtained from RNA-seq. AFB1, aflatoxin B1; Enh, enhancer; CNA, copy number alteration; SV, structural variant; WT, wild type; ns, not significant.

All HBV integrations at the *TERT* locus were located in the promoter region or in the 5' untranslated region (figure 4B; online supplemental figure 11B). Most of the promoter integrations were in the same 5'>3' orientation and contained both viral enhancer regions (35 out of 48). Increased mRNA *TERT* expression was higher for viral integrations situated closer to the *TERT* ATG (<800 bp) but was independent of the orientation of the inserted sequence (figure 4B, online supplemental figure 11C). Moreover, HBV insertions induced a higher overexpression than promoter mutations at the −124 and −146 hotspots and structural variants inducing CNA without viral insertions (online supplemental figure 11D). Modelling *TERT* promoter integration identified in tumour #3885T in luciferase reporter vector (online supplemental figure 11E,F), we confirmed that *TERT* activation was caused by the two viral enhancers with enhancer 1 providing a stronger activation compared with enhancer 2 (figure 4C). HBV insertions in *TERT* promoter were exclusive from mutations or structural variations of the same region, and more frequent in tumours with a higher HBV copy number per cell ($p<0.001$) driven by more HBV clonal insertions ($p<0.001$) and more HBV replication ($p=0.01$; figure 4D).

We previously described two tumours with HBV insertions in *CCNE1* or *CCNA2* genes as part of a homogenous group of HCC (CCN-HCC) characterised by a signature of specific chromosome rearrangements (templated-insertion cycles) induced by stress replication.[19] We identified three additional HBV integrations in the promoter of *CCNE1* using viral capture and RNA-seq showed a significant overexpression of normal *CCNE1* transcripts (online supplemental figure 12A,B). Three *KMT2B* integrations were also identified between exons 3–6, which increased the mRNA expression and altered the transcript structure through alternative splicing or intron retention (online supplemental figure 12A,C). Overall, *TERT* promoter integrations were enriched in non-proliferative tumours ($p=0.007$; figure 4D). In contrast, tumours harbouring HBV integrations in *CCNE1*, *CCNA2* or *KMT2B* were more frequently identified in patients without cirrhosis (5/8) with African or Asian origin (6/8), belonging to the proliferative class of HCC, and associated with G1 (for *KMT2B*-integrated HCC) or G3 (for CCN-HCC) transcriptomic subgroup. Interestingly, CCN-HCC showed *TERT* promoter alterations whereas *KMT2B*-integrated HCC did not show *TERT* promoter or any other driver genes alterations, suggesting different processes of carcinogenesis (figure 4D).

Integrated analysis between clinical features, HBV and human genomic alterations

To integrate all the genomic alterations identified in the tumours, we analysed a subgroup of 130 HBV-positive HCC compared with 135 HBV-negative tumours sequenced in WGS or WES and RNA-seq[19 34] (see online supplemental table 1 for description of the series). Patients with HBV HCC were younger, enriched in African or Asian origin, and they showed specific cofactors: HDV infection or carcinogen exposure signatures (figure 5A, first column). Indeed, the two sporadic mutational signatures 22 and 24, characteristics of acid aristolochic (AA) and aflatoxin B1 (AFB1) exposure respectively, were more frequent within HBV-positive patients. Transcriptomic and genomic analysis revealed an association between HBV infection and G3 transcriptomic class, more frequent *TP53* mutations and less *CTNNB1* mutations in tumours.

In addition, in the series of 177 HBV-positive patients analysed in capture, 60 had an African origin; they were younger with frequent AFB1-mutational signature in HCC characterised by a high expression of progenitor markers such as *EPCAM* (figure 5A). No specific viral feature identified in tumours were significantly associated with the geographic origin of patients, except for the HBV genotype. Among patients with cofactors, whereas AFB1 exposure was strongly associated with *TP53* R249S mutation, HBV/HDV-related HCC harboured less HBV-associated alterations in *TERT* promoter or any other driver genes. Finally, 10 patients had a negative HBsAg serology, they were negative for HBV DNA in the serum but positive in the liver; their tumours showed a tendency for low numbers of HBV copies per cell and of HBV integrations and these HCC were enriched in G1 transcriptomic group. Of note, two of these HCCs had a clonal HBV integration in the *TERT* promoter (#4229T and #4265T).

Tumours with a high number of HBV integrations were associated with a poor prognosis, independently from other features such as tumour size, microvascular invasion, differentiation status and transcriptomic groups (figure 5B–5C; online supplemental table 5). This association between the high number of integrations in tumours and poor survival was independent from the other features of HBV infection. Interestingly, these patients were significantly younger and with tumours harbouring a high proportion of HBV integrations affecting cancer driver genes such as *TERT* ($p=0.02$) through an insertion in the promoter, *TP53* ($p<0.001$) or *MYC* ($p=0.03$) through HBV-associated CNA (figure 5A).

DISCUSSION

This study presents an integrative analysis of HBV genomes in non-tumour and tumour liver tissues of a large cohort of patients mainly coming from Europe and Africa. On one hand, HBV integrations in non-tumour tissues reflect an active viral replication and the expansion of hepatocytes in liver tissues with a lower expression of inflammation-associated genes. On the other hand, HBV-integrated sequences in tumours highlight a functional selection of cells harbouring HBV-associated structural rearrangements or insertional mutagenesis as driver alterations.

Our method of identification of HBV integrations based on viral capture coupled with WGS and WES on frozen tissues enabled to characterise integrations based on their clonality defining precisely the detected events as 'clonal', 'subclonal' or 'unique' with a quantitative method. In contrast with previous studies and in accordance with the clonal cell expansion during carcinogenesis with selection of functional genomic alterations, here, more clonal HBV insertions were identified in tumours compared with the non-tumour liver tissues.[14 15 21 36] In the same line, we observed an important enrichment of HBV integrations in simple repeats, underlying the importance of precisely filtering insertional events detected in viral capture to remove duplicates. Therefore, our study based on a highly confident description of HBV integrations provides a comprehensive view of the integration process in liver tissues.

In non-tumour liver, we confirmed that integration events occur more frequently in regions of the human genome with open chromatin,[21 37] resulting in part from microhomology-mediated end joining.[15 38 39] More importantly, viral replication is related to the number of integrations reflecting either an increased number of infected cells, either a multisteps integration process within a single cell. Both hypotheses are probably occurring concurrently in the liver. A recent in vitro study[13] has shown that the integration rate within the first week after infection is not altered by replication suppression, suggesting that the majority of integration events occurs at primo-infection.

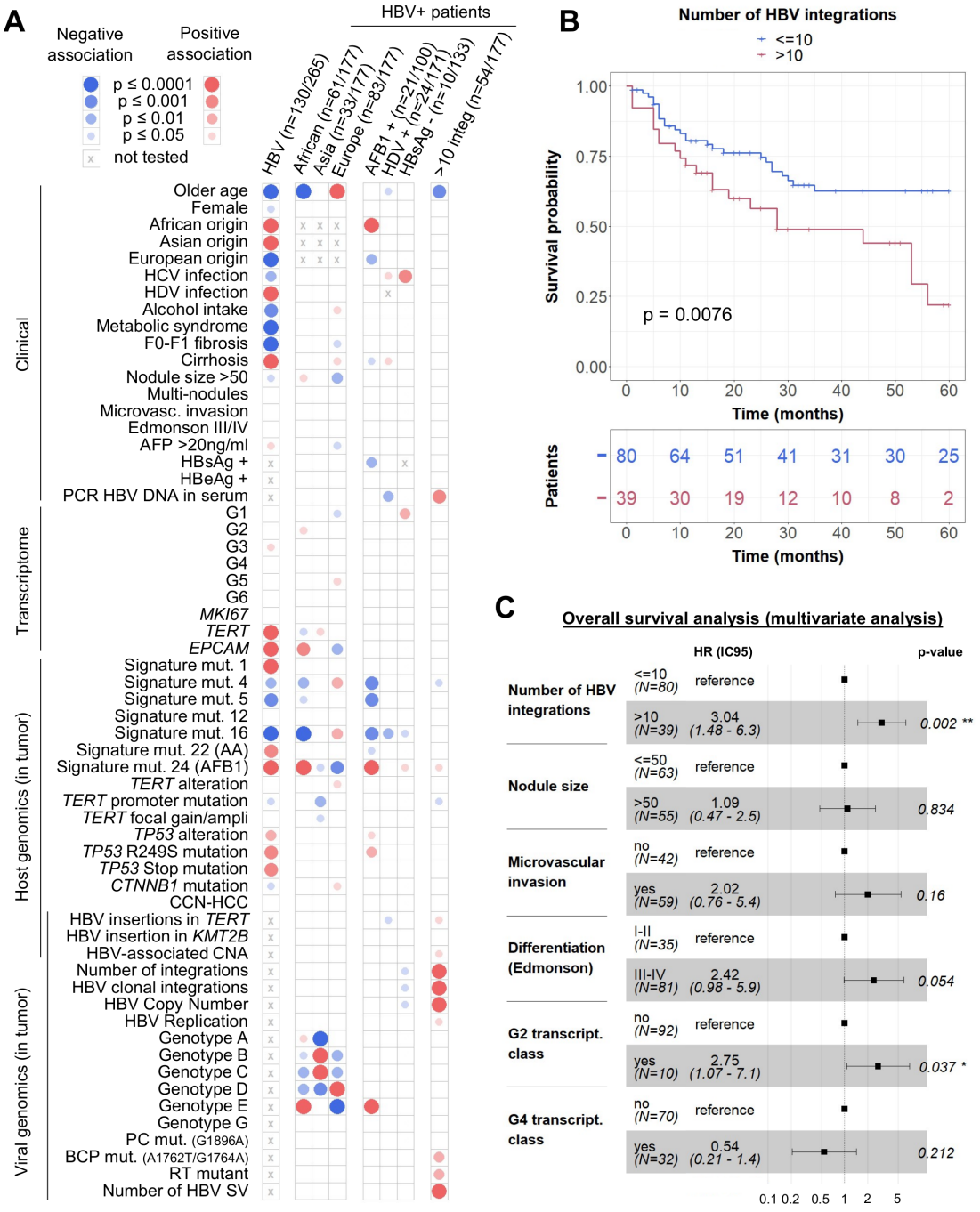


Figure 5 Integrative analysis reveals that a high number of HBV integrations is associated with poor survival of patients. (A) Mosaic plot: association of HBV infection with patient, tumour and viral characteristics in a series of 265 HCC (next-generation sequencing series) and association of geographic origin, aflatoxin B1 exposure, HDV infection, hepatitis B surface antigen (HBsAg) negativity or number of HBV integrations with patient, tumour and viral characteristics in a series of 177 HBV-positive HCC (capture series). Blue and red circles indicate negative and positive associations, respectively. Colour intensities represent different levels of statistical significance. Statistical analysis was performed using χ^2 test, Wilcoxon signed-rank test or Pearson's correlation with respect to the type of variable. (B) Kaplan-Meier curves for 5-year overall survival from 119 patients after curative R0 resection. (C) Multivariate Cox regression model for overall survival analysis. AA, aristolochic acid. *p<0.05; **p<0.01.

However, all NGS-based studies, including ours, identified the presence of several clonal integrations in some cellular clones. Therefore, several HBV integrations must accumulate at a low rate in a single cell, promoted by active replication.

The number of HBV integrations may increase over time within hepatocytes in normal liver and as in vitro and in vivo studies have shown that HBV integrations occur early after infection,[13 40 41] a majority of events could be considered as passenger. Nevertheless, the presence of clonal integrations (ie, present in all cells of a given frozen sample) reflects the expansion of hepatocytes and this process appears to be different in non-tumour tissues and in tumours. In non-tumour tissues, it might be explained by a selective pressure induced by chronic inflammation and the emergence of cirrhotic nodules in the liver but this process seems to be independent from integrations and viral replication as it has already been suggested.[42–44] Indeed, in our study, clonal integrations in these tissues do not argue for functional consequences but are located in regions recurrently

targeted by integrations (large, highly expressed genes), and are not associated with viral replication. Yet, as non-tumour tissues harbouring clonal integrations showed a downregulation of genes involved in inflammatory response, it suggests that some hepatocytes with a selective advantage to escape immune antiviral response might undergo active proliferation and obtain a protective profile against malignant transformation.

Although the majority of HBV integrations are passenger events and do not have any functional consequences, some can be driver and promote HCC initiation in cirrhotic or non-cirrhotic livers. While integrations had already been associated with CNA in previous studies,[14 45 46] we identified for the first time recurrent structural rearrangements associated with viral insertions in HCC. Through integrations often located around centromeric or telomeric regions, HBV is involved in large complex structural rearrangements frequently inducing gains of chr8q (including *MYC* locus) and chr5p (including *TERT* locus) and losses of chr17p (including *TP53* locus). Thus, in tumours harbouring integration-associated rearrangements, HBV integrations are associated with alterations of cancer-driver genes located at distance. As HBV DNA integration occurs at double-strand breaks,[23 47] the presence of HBV-integrated sequences delimitating CNA may result from opportunistic mechanism. However, our data showed no correlation between the numbers of HBV integrations and of CNA in tumours; it highlighted the role of specific alterations due to integration-associated rearrangements in promoting selection of hepatocytes and HCC development.

HBV integrations may also drive carcinogenesis by altering the closest gene from the viral-integrated sequence through insertional mutagenesis. Our study confirmed that the *TERT* promoter is the main HBV integration hotspot in HCC, as one-third of HBV-related HCC harboured clonal integration at this hotspot. In these tumours, strong activation of *TERT* was due to the presence of the viral enhancers as it had already been reported,[16 48] at proximity of the ATG start codon and in an orientation-independent manner. In addition, HBV integrations in *CCNA2* or *CCNE1* genes have already been investigated in a previous study of our group.[19] HBV-integrated sequence is either altering directly the structure of the protein (for *CCNA2*) or inducing a strong overexpression due to its enhancer sequences (for *CCNE1*). Alterations of one of these two genes induce strong proliferation, replicative stress and a signature of rearrangements directly promoting the development of a tumour belonging to a specific subgroup of HCC developed on non-cirrhotic livers (CCN-HCC). Finally, even if further investigations are needed to fully understand the functional consequences, tumours with HBV integrations in *KMT2B* are also mainly developing on non-cirrhotic liver with a high expression of progenitor markers, suggesting another strong direct mechanism induced by HBV-integrated sequence.

Overall, this study underlines the complexity of the interplay between HBV integrations, HBV replication and chromosomal instability in hepatocarcinogenesis, in addition to other cofactors. On one side, aflatoxin B1 exposure, mainly present in Africa, may initiate directly HBV-related HCC development through *TP53* R249S mutation.[49] On the other side, HDV infection limits HBV replication[50] and viral integrations but accelerates chronic inflammation and fibrosis in young patients. Importantly, our series showed that the number of HBV integrations in tumours is the only marker within viral features associated with poor prognostic.

In conclusion, structural rearrangements associated with HBV integrations are a mechanism to drive carcinogenesis by altering cancer-driver genes at distance, different from *cis*-activating insertional mutagenesis. This underlines the heterogeneity of HBV-related HCC and the importance of molecular characterisation to identify new specific therapeutic opportunities.

Author affiliations
[1]Centre de Recherche des Cordeliers, Sorbonne Université, INSERM, Université de Paris, Paris, France
[2]Functional Genomics of Solid Tumors laboratory, équipe labellisée Ligue Nationale contre le Cancer, Labex OncoImmunology, Paris, France
[3]Service d'Anatomopathologie, Hôpital Henri Mondor, APHP, Institut Mondor de Recherche Biomédicale, Créteil, France
[4]Service de Pathologie, Hôpital Beaujon, APHP, Clichy, France
[5]Université Paris Diderot, CNRS, Centre de Recherche 27 sur l'Inflammation (CRI), Paris, France
[6]Service Hépato-Gastroentérologie et Oncologie Digestive, Hôpital Haut-Lévêque, CHU de Bordeaux, Bordeaux, France
[7]Service de Pathologie, CHU Bordeaux GH Pellegrin, Bordeaux, France
[8]Université Bordeaux, Inserm, Research in Translational Oncology, BaRITOn, Bordeaux, France
[9]Service d'Hépatologie, Hôpital Avicenne, Hôpitaux Universitaires Paris-Seine-Saint-Denis, APHP, Bobigny, France
[10]Service d'Hépato-Gastro-Entérologie, Hôpital Henri Mondor, APHP, Université Paris Est Créteil, Inserm U955, Institut Mondor de recherche biomedicale, Creteil, Île-de-France, France
[11]Hôpital Européen Georges Pompidou, AP-HP, Paris, France

Twitter Jessica Zucman-Rossi @Zucmanrossi

Acknowledgements The authors would like to thank Alain Nicolas, Sylvain Baulande and Sonia Lameiras at Institut Curie for their help in analysing PacBio results and setting up viral capture. The authors would like to thank Quentin Bayard and Karl Hong for their help in analysing Bionano sequencing results. The authors would also like to thank Gabrielle Couchy, Iadh Mami, Bénédicte Noblet, Massih Ningarhari, Jill Pilet and Jie Yang for their help in molecular biology experiments.

Contributors JZ-R conceived and directed the research. CP and JZ-R designed the study and wrote the manuscript. CP and TLB performed the experiments. CP, SI, TZH, SC, EL and JZ-R analysed and interpreted the data. JC, VP, J-FB, J-CN and GA provided essential biological resources and collected clinical data. All authors approved the final manuscript and contributed to critical revisions to its intellectual context.

Funding This work was supported by ANRS (French national agency for research on AIDS and viral hepatitis). The group is supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d'avenir), grant IREB, Coup d'Elan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM, FRM prix Rosen, Ligue Contre le Cancer Comité de Paris (prix René et André Duquesne) and Fondation Mérieux. CP was supported by a fellowship from ANRS (French national agency for research on AIDS and viral hepatitis). TLB was supported by an "Attractivite IDEX" fellowship from IUH, TZH by a fellowship from Cancéropole Ile de France and Fondation d'Entreprise Bristol-Myers Squibb pour la Recherche en Immuno-Oncologie, and SC by CARPEM and the Labex OncoImmunology.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the institutional review board (IRB) committees.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The sequencing data of the LICA-FR cohort reported in this paper have been deposited to the European Genome-phenome Archive (EGA) database (RNA-seq fastq files accessions (EGAS00001001284), (EGAS00001002879), (EGAS00001003310), (EGAS00001003837) and (EGAS00001004629); WES bam files accessions (EGAS00001000217), (EGAS00001001002), (EGAS00001003063), (EGAS00001003130), (EGAS00001003837) and (EGAS00001004629); WGS bam files accessions (EGAS00001000706), (EGAS00001002408), (EGAS00001002888), (EGAS00001003063), (EGAS00001003837) and (EGAS00001004629), through the International Cancer Genome Consortium (ICGC) data access committee. Data are available for reuse and can be consulted at the following address: https://ega-archive.org/studies/EGASxxx with access permission from ICGC Data Access Compliance Office.

ORCID iDs
Camille Péneau http://orcid.org/0000-0002-5479-1288
Sandrine Imbeaud http://orcid.org/0000-0001-8439-6732
Theo Z Hirsch http://orcid.org/0000-0003-4428-2997
Jean-Charles Nault http://orcid.org/0000-0002-4875-9353
Jessica Zucman-Rossi http://orcid.org/0000-0002-5687-0334

REFERENCES

1 World Health Organization. Global hepatitis report, 2017; 2017.
2 Thomas DL. Global elimination of chronic hepatitis. *N Engl J Med* 2019;380:2041–50.
3 Bray F, Ferlay J, Soerjomataram I, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
4 Kew MC. Hepatocellular carcinoma: epidemiology and risk factors. *J Hepatocell Carcinoma* 2014;1:115.
5 Papatheodoridis GV, Chan HL-Y, Hansen BE, *et al*. Risk of hepatocellular carcinoma in chronic hepatitis B: assessment and modification with current antiviral therapy. *J Hepatol* 2015;62:956–67.
6 Wu C-Y, Lin J-T, Ho HJ, *et al*. Association of nucleos(t)ide analogue therapy with reduced risk of hepatocellular carcinoma in patients with chronic hepatitis B: a nationwide cohort study. *Gastroenterology* 2014;147:143–51.
7 Wong GL-H, Chan HL-Y, Mak CW-H, *et al*. Entecavir treatment reduces hepatic events and deaths in chronic hepatitis B patients with liver cirrhosis. *Hepatology* 2013;58:1537–47.
8 Hosaka T, Suzuki F, Kobayashi M, *et al*. Long-Term entecavir treatment reduces hepatocellular carcinoma incidence in patients with hepatitis B virus infection. *Hepatology* 2013;58:98–107.
9 Yang JD, Hainaut P, Gores GJ, *et al*. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol* 2019;16:589–604.
10 Valaydon ZS, Locarnini SA. The virological aspects of hepatitis B. *Best Pract Res Clin Gastroenterol* 2017;31:257–64.
11 Lamontagne RJ, Bagga S, Bouchard MJ. Hepatitis B virus molecular biology and pathogenesis. *Hepatoma Res* 2016;2:163.
12 Levrero M, Zucman-Rossi J. Mechanisms of HBV-induced hepatocellular carcinoma. *J Hepatol* 2016;64:S84–101.
13 Tu T, Budzinska MA, Vondran FWR, *et al*. Hepatitis B Virus DNA Integration Occurs Early in the Viral Life Cycle in an *In Vitro* Infection Model via Sodium Taurocholate Cotransporting Polypeptide-Dependent Uptake of Enveloped Virus Particles. *J Virol* 2018;92:e02007-17–17.
14 Sung W-K, Zheng H, Li S, *et al*. Genome-Wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012;44:765–9.
15 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.
16 Li C-L, Li C-Y, Lin Y-Y, *et al*. Androgen receptor enhances hepatic telomerase reverse transcriptase gene transcription after hepatitis B virus integration or point mutation in promoter region. *Hepatology* 2019;69:498–512.
17 Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu, Cancer Genome Atlas Research NetworkAlly A. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 2017;169:1327–41.
18 Fujimoto A, Totoki Y, Abe T, *et al*. Whole-Genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 2012;44:760–4.
19 Bayard Q, Meunier L, Peneau C, *et al*. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun* 2018;9:5235.
20 Dong H, Zhang L, Qian Z, *et al*. Identification of HBV-MLL4 integration and its molecular basis in Chinese hepatocellular carcinoma. *PLoS One* 2015;10:e0123175.
21 Furuta M, Tanaka H, Shiraishi Y, *et al*. Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. *Oncotarget* 2018;9:25075–88.
22 Tong S, Revill P. Overview of hepatitis B viral replication and genetic variability. *J Hepatol* 2016;64:S4–16.
23 Tu T, Budzinska MA, Shackel NA, *et al*. HBV DNA integration: molecular mechanisms and clinical implications. *Viruses* 2017;9:75.
24 La Bella T, Imbeaud S, Peneau C, *et al*. Adeno-Associated virus in the liver: natural history and consequences in tumour development. *Gut* 2020;69:737–47.
25 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
26 Debacker K, Kooy RF. Fragile sites and human disease. *Hum Mol Genet* 2007;16 Spec No. 2:R150–8.
27 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, *et al*. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–30.
28 Ritchie ME, Phipson B, Wu D, *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47–8.
29 Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
30 Hirsch TZ, Negulescu A, Gupta B, *et al*. BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA. *J Hepatol* 2020;72:S0168827819307184.
31 Boyault S, Rickman DS, de Reyniès A, *et al*. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 2007;45:42–52.
32 Lazarevic I. Clinical implications of hepatitis B virus mutations: recent advances. *World J Gastroenterol* 2014;20:7653.
33 Nault J-C, Datta S, Imbeaud S, *et al*. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat Genet* 2015;47:1187–93.
34 Nault J-C, Martin Y, Caruso S, *et al*. Clinical impact of genomic diversity from early to advanced hepatocellular carcinoma. *Hepatology* 2020;71:164–82.
35 Caruso S, Calatayud A-L, Pilet J, *et al*. Analysis of liver cancer cell lines identifies agents with likely efficacy against hepatocellular carcinoma and markers of response. *Gastroenterology* 2019;157:760–76.
36 Yang L, Ye S, Zhao X, *et al*. Molecular characterization of HBV DNA integration in patients with hepatitis and hepatocellular carcinoma. *J Cancer* 2018;9:3225–35.
37 Li X, Zhang J, Yang Z, *et al*. The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. *J Hepatol* 2014;60:975–84.
38 Yoo S, Wang W, Wang Q, *et al*. A pilot systematic genomic comparison of recurrence risks of hepatitis B virus-associated hepatocellular carcinoma with low- and high-degree liver fibrosis. *BMC Med* 2017;15:214.
39 Yan H, Yang Y, Zhang L, *et al*. Characterization of the genotype and integration patterns of hepatitis B virus in early- and late-onset hepatocellular carcinoma. *Hepatology* 2015;61:1821–31.
40 Yang W, Summers J. Integration of hepadnavirus DNA in infected liver: evidence for a linear precursor. *J Virol* 1999;73:9710–7.
41 Summers J, Jilbert AR, Yang W, *et al*. Hepatocyte turnover during resolution of a transient hepadnaviral infection. *Proc Natl Acad Sci U S A* 2003;100:11652–9.
42 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.
43 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.
44 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.
45 Zapatka M, Borozan I, Brewer DS, *et al*. The landscape of viral associations in human cancers. *Nat Genet* 2020;52:320–30.
46 Totoki Y, Tatsuno K, Covington KR, *et al*. Trans-Ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet* 2014;46:1267–73.
47 Bill CA, Summers J. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc Natl Acad Sci U S A* 2004;101:11135–40.
48 Sze KM*et al*. HBV- *TERT* Promoter Integration Harnesses Host ELF4 Resulting in *TERT* Gene Transcription in Hepatocellular Carcinoma. *Hepatology hep* 2020;31231.
49 Amaddeo G, Cao Q, Ladeiro Y, *et al*. Integration of tumour and viral genomic characterizations in HBV-related hepatocellular carcinomas. *Gut* 2015;64:820–9.
50 Mentha N, Clément S, Negro F, *et al*. A review on hepatitis D: from virology to new therapies. *J Adv Res* 2019;17:3–15.

Supplementary Information

Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma

Péneau C. et al.

Supplementary Materials and Methods

Supplementary Figures

Supplementary Tables

Supplementary table 1 - Clinical description of the series.
Supplementary table 2 - List of probe sets and primers.
Supplementary table 3 - List of HBV genotypes and human regions targeted in viral capture.
Supplementary table 4 - Characterization of 8809 HBV integration breakpoints identified by viral capture.
Supplementary table 5 - Survival analysis in patients treated with R0 curative resection.

SUPPLEMENTARY MATERIALS AND METHODS

Patients and tissue samples

A first series of 1128 hepatocellular carcinoma and 1063 non-tumor counterparts from 1128 patients was assembled and the study was approved by our institutional review board (IRB) committees (CCPRB Paris Saint-Louis, 1997 and 2004; Bordeaux 2010-A00498-31, Ile-de-France VII: projects C0-15-003 and PP 16–001). Patients were involved as advisers in hospital tumor collection boards and the definition of inform consent in IRB. They were not involved in the design of this study. Samples were collected in 12 academic hospitals in France (LICA-FR cohort) and frozen immediately at -80°C after resection or biopsy. DNA was systematically extracted, as RNA when the quantity of tissue was sufficient. The majority of tumor samples was from primary tumors except for 12 samples collected at relapse. Clinical features and serological data were gathered from each center, classified as previously described[1], and are summarized in **supplementary table 1**. A flowchart of the study inclusion at each step is provided in **supplementary figure 1**.

Viral DNA screening

Genomic DNA were screened to detect the presence of HBV DNA by quantitative RT-PCR (qRT-PCR) on Fluidigm 96.96 dynamic arrays using the BioMark Real-Time PCR system with TaqMan probe sets designed with Primer3Plus software, as previously described[2]. In total eleven probes were designed to detect specifically five viral regions for all the eight main HBV genotypes (**supplementary table 2**). Results were analyzed using the Fluidigm Real-Time PCR Analysis software (V.4.1.3) and reported to HMBS (Hydroxymethylbilane Synthase) as reference gene. The quantification was expressed in viral copy number/cell and based on the results obtained for a series of HBs and HBx plasmids with known concentration. The values obtained were tested for bimodal distribution using normalmixEM function of mixtools package in R[3].

HBV viral capture

Viral capture was performed for a selection of 177 frozen HCC and 170 matched non-tumor liver tissues from 177 patients with a known HBV etiology. 13 tumors and 8 non-tumor liver tissues from 13 HBV-negative patients were used as negative controls. Single-stranded biotinylated probes were designed with Roche NimbleGen (SeqCap EZ Designs, Roche NimbleGen Inc., Madison, WI, USA), to target 40 references representing the 8 main HBV genotypes and 4 human regions (*TERT* promoter, *CCNE1*, *CCNA2* and *KMT2B*). The reference of sequences used is shown in **supplementary table 3**.

Samples containing 1ug of DNA were used for DNA library preparation. DNA was sheared mechanically in fragments with an average length of 1kb, before following the SeqCap EZ HyperCap Workflow developed by Roche. We used KAPA Dual-Indexed Adapters for sample indexing and we multiplexed library samples by using one capture probe set for 12 to 42 samples. DNA samples were selected for multiplexing based on the HBV copy number per cell measured by qRT-PCR, in order to have similar viral loads between all library samples in one

capture pool. A sample not containing HBV DNA was added in each pool as negative control. We performed a double capture with two successive hybridization steps to enrich twice for DNA targets and increase the specificity of capture, following an adapted protocol developed by Roche. Final libraries were quantified with KAPA Library Quantification Kit and controlled with a DNA High Sensitivity assay on Caliper LabChip GX. Two final libraries were merged to be sequenced using an Illumina MiSeq instrument with paired-end reads of 2 × 250 nt.

Analysis of integration events and clonality definition

A flowchart described the overall pipeline used for the analysis of integration events is provided in **supplementary figure 2**.

Raw capture data obtained after sequencing were post-processed by trimming low-quality bases with Trimmomatic[4] and removing duplicates. Filtered reads were first aligned on the reference used (40 HBV sequences and 4 regions of HG19 genome) using Burrows-Wheeler Aligner (V.0.7.15)[5]. A normalization factor was assessed based on the mean coverage of the targeted human regions and was corrected to remove capture biases. The main HBV genotype of each sample was determined as the sequence with the higher number of mapped reads. Read pairs with at least one read aligned on a HBV sequence were extracted using samtools (V.1.3)[6], and realigned to a custom reference genome including the HG19 reference and the HBV sequence of the main genotype identified for each sample (one single reference was chosen for each of the 8 genotypes A-H) (**supplementary table 3**). For genotype A, the X02763 sequence was renumbered to use the EcoR1 restriction site as the +1. To compare HBV breakpoints from different genotypes, all the positions in HBV genomes were converted using this numbering.

HBV copy number per cell was assessed from the ratio (multiplied by 2) of the mean read coverage of the HBV genome and the normalization factor, and was compared with the measures obtained by qRT-PCR (**supplementary figure 3A**). HBV copy number per cell includes all HBV forms: both integrated HBV sequences and HBV episomal forms. We identified the HBV/human chimeric regions with at least one chimeric read supporting the junction and one read pair with a read mapping on HBV and its mate mapping on HG19. Coverage of reads were computed with bedtools multicov utility[7] and the characteristics of junction breakpoints (position and sequence) were extracted from hard clipped reads with htslib package. Clonality of integration events was assessed from the ratio (multiplied by 2) of the coverage of reads at the junction breakpoint and the normalization factor. All viral insertions with a clonality greater than 0.03 were validated by visual inspection on IGV (Integrative Genomics Viewer). Similar integration events (same position on HG19 and on HBV and same sequence) found in two independent samples within one single capture (and with at least a 10-fold difference in clonality) were considered as contaminations intra-capture and the event with the lowest clonality was removed from the analysis. Similar integration events found in the non-tumor and tumor tissues of the same patients were inspected specifically and considered as contaminations based on the very few number of reads detected in the non-tumor tissues compared to the tumors. Integration events located in repeated regions of the human genome with the same breakpoint and same orientation in HBV genome were considered as one unique event. The 8809 integration events identified were classified in 3 categories using the k-means method: (1)“clonal integrations” observed in more than 48% of cells, (2)“unique integrations”

observed in less than 3% of cells and (3)"subclonal integrations" with intermediate copy numbers (**supplementary figure 3B**).

Integration events identified from viral capture were compared with those identified from whole genome sequencing (WGS) for 20 tumors and 20 non-tumor liver tissues (7 pairs tumor/non-tumor had already been published) (**supplementary figure 3C-3D**). WGS was performed and analyzed as already described[8]. The bioinformatics analysis of HBV integrations was similar for WGS as for capture except for the normalization factor that was corresponding to the mean sequencing coverage for WGS (30x for non-tumor tissues and 60x or 90x for tumors).

We annotated the integration breakpoints with the following genomic features: replication timing in HepG2 cell line (ENCODE[9]), highly expressed (top 20%) genes in non-tumor liver from HBV-positive patients, common and rare fragile sites[10], repetitive sequences and CpG islands, and chromatin structure in adult liver (ROADMAP[11]). Insertions were named "classic" if the two breakpoints clustered within 25kb in opposite direction, "local inversion" if two breakpoints clustered within 25kb in the same direction, "large rearrangements" if only one breakpoint was identified and "cluster of insertions" if more than two breakpoints were identified.

Genomic DNA sequencing of driver genes

A selection of 265 tumor samples were sequenced with WGS (for 62 tumors with their adjacent liver tissues: 43 already published and 19 new cases) or Whole Exome Sequencing (WES, for 203 tumors with their adjacent liver tissues: 134 already published and 69 new cases. Protocols and methods of analysis have previously been described[1,8,12]. In particular, all tumor samples without WGS were re-sequenced to search for TERT promoter mutation with Sanger or MiSeq sequencing as previously described[13].

Detection of HBV episomal form

A specific DNAse/TaqMan-based assay was adapted from protocol by Werle-Lapostolle et al to detect HBV episomal form and performed as previously described for AAV2 episomal form detection[2]. All samples from the Capture series with enough DNA material were screened for the presence of HBV episomal form: 162 tumors and 155 non-tumor liver tissues from 172 patients. The HBV probes used are listed in **supplementary table 2**. The difference between HBV Ct values without and with PS-DNase digestion was analyzed to determine the presence of episomal form (viral DNA detected before and after digestion) or the absence of episomal form (viral DNA detected before but not after digestion). When the results did not enable to conclude directly due to close CT values, a second step of digestion was performed. The results obtained were only used as qualitative results.

Viral mRNA and specific genes mRNA screening

For viral mRNA, ten specific probe sets covering eight regions of the HBV genome were designed to detect HBV mRNA from the main eight HBV genotypes; they are listed in **supplementary table 2**. The presence of a complete transcript was defined as the positivity of all probes along the transcript. When not all probes were positive until the viral polyA, we

considered the transcript as partial or truncated. Probe sets also designed to detect HCV and HDV mRNA. For TERT expression analysis, we used human catalogue TaqMan probes (Hs00972656_m1). We performed qRT-PCR using BioMark Real-Time PCR system. Expression data were normalized with the $2^{-\Delta Ct}$ method relative to ribosomal 18S (Hs03928990_g1). Five normal tissues were used as reference.

HBV replicative forms analysis

We assessed the presence of HBV replicative form in a tissue as the presence of HBV episomal forms and the presence of pregenomic RNA (pgRNA). When episomal forms were detected in the absence of pgRNA, the tissue was considered as containing "Episomal not replicative HBV DNA".

RNAseq and transcriptomic analysis

A selection of 265 tumors (130 HBV-positive and 135 HBV-negative) and 24 HBV-positive non-tumor tissues were sequenced using Illumina TruSeq or Illumina TruSeq Stranded mRNA kit on HiSeq2000 sequencer by IntegraGen, Evry, France[14]. Among the 24 non-tumor samples, we used the Bioconductor limma package[15] to test for differential expression, between samples harboring a HBV clonal integration and other HBV-positive non-tumor tissues, of all genes expressed. We applied a q-value threshold of ≤ 0.05 to define differentially expressed genes. We used an in-house adaptation of the GSEA method[16] to identify gene sets from the MSigDB (v. 6.0) database overrepresented among up- and down-regulated genes.

In addition, qRT-PCR of 190 genes using BioMark Real-Time PCR system was performed on 133 HBV-positive tumors sequenced in viral capture, in order to classify them in the G1-G6 classification as previously described for the LICA-FR cohort[17,18].

HBV variants analysis

From viral capture, we extracted the number of each base per position of the HBV sequence. The positions on all HBV genotypes were modified to use the EcoR1 restriction site as the +1. We considered the positions 1762 and 1764 of the Basal Core Promoter region, 1896 of the PreCore region and the positions 646, 667, 670, 671, 739, 741, 836 and 877 of the Reverse transcriptase (RT) domain of HBV polymerase. For RT mutants, we considered the mutations inducing the following amino acids changes: V173L, L180M, A181T, A181V, M204V, M204I, N236T and M250V[19,20]. Only samples with a coverage greater than 20 reads at the selected position were used for analysis. A sample was considered as mutated if the number of reads harboring the mutation was greater than 20 or represented more than 10% of the total coverage.

Cell culture, transfection and dual luciferase assay

HuH7 cells were purchased from the American Type Culture Collection (ATCC) and cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 100U/mL penicillin/streptomycin. Cells were co-transfected using Lipofectamine 3000 (Life Technologies) with a pGL3 plasmid containing the wild-type *TERT* promoter or promoter with the two hotspot mutations, or different HBV sequences (normal or scrambled) controlling a

firefly luciferase reporter gene and a plasmid encoding Renilla luciferase (Promega). Luminescence from firefly luciferase was normalized on the corresponding Renilla luciferase activity as previously described[21]. A fold change was then calculated relative to the values obtained for the construct containing the wild-type *TERT* promoter.

Long-read sequencing

Three HBV-positive tumors (#1151T, #1597T, #1994T) were selected to perform long-read sequencing using the technology SMRT (single molecule real time technology) with a PacBio-SMRTcell system (ICGex NGS platform, Curie Institute, Paris, France). Samples were selected based on the results of Fragment Analyzer to have an average length of DNA fragments of 10kb. Libraries were prepared based on the PacBio template Prep kit protocol. For analysis, error-prone raw subreads were firstly merged to obtain unique polymerase reads with the Consensus Circular Sequencing (CCS2) algorithm (Pacific Biosciences). The consensus reads obtained were aligned using minimap2 aligner (V.2.16)[22] on a custom reference genome including the HG19 reference and the HBV sequence of the main genotype previously identified for each sample.

Bionano whole-genome mapping

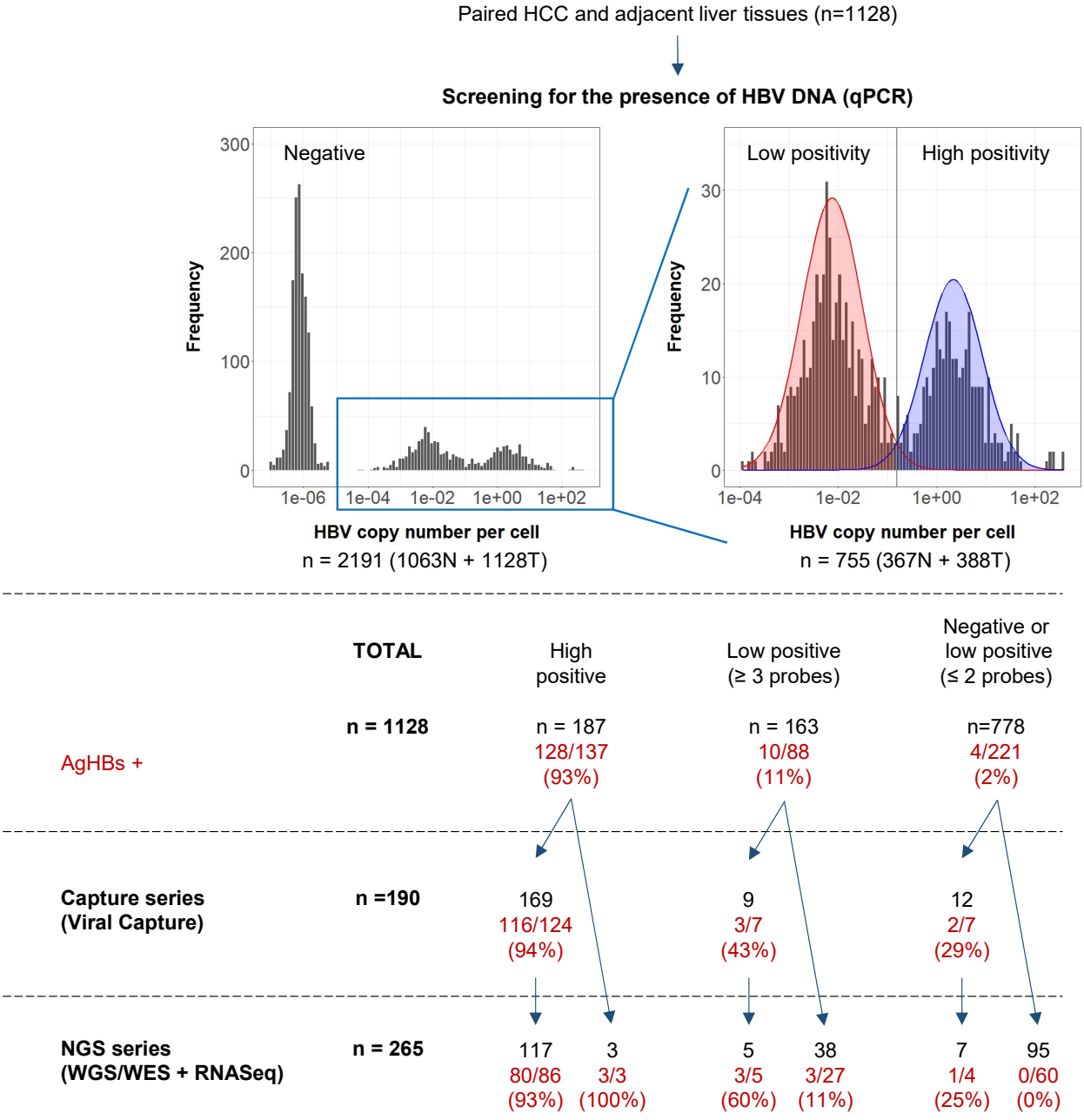
The three tumors analyzed with long-read sequencing (#1151T, #1597T, #1994T) were selected with their corresponding non-tumor liver tissues to perform whole-genome mapping using Bionano technology (Bionano Genomics, La Jolla, California, USA). 30mg of tissue were used to extract long molecules of DNA, labeled with Bionano reagents by incorporation of fluorophores at a specific sequence motif along the genome. The labeled genomic DNA was linearized in the SaphyrChip using NanoChannel arrays and single molecules were imaged and digitized. As molecules are uniquely identifiable by distinct distribution of sequence motif labels, they were then assembled by pairwise alignment into de novo genome maps. Genome-wide Structural Variant (SV) calling was performed in tumor samples by aligning these maps and molecules to a reference genome, and SV calls are annotated using the corresponding non-tumor tissue from the same patient.

Statistical analysis

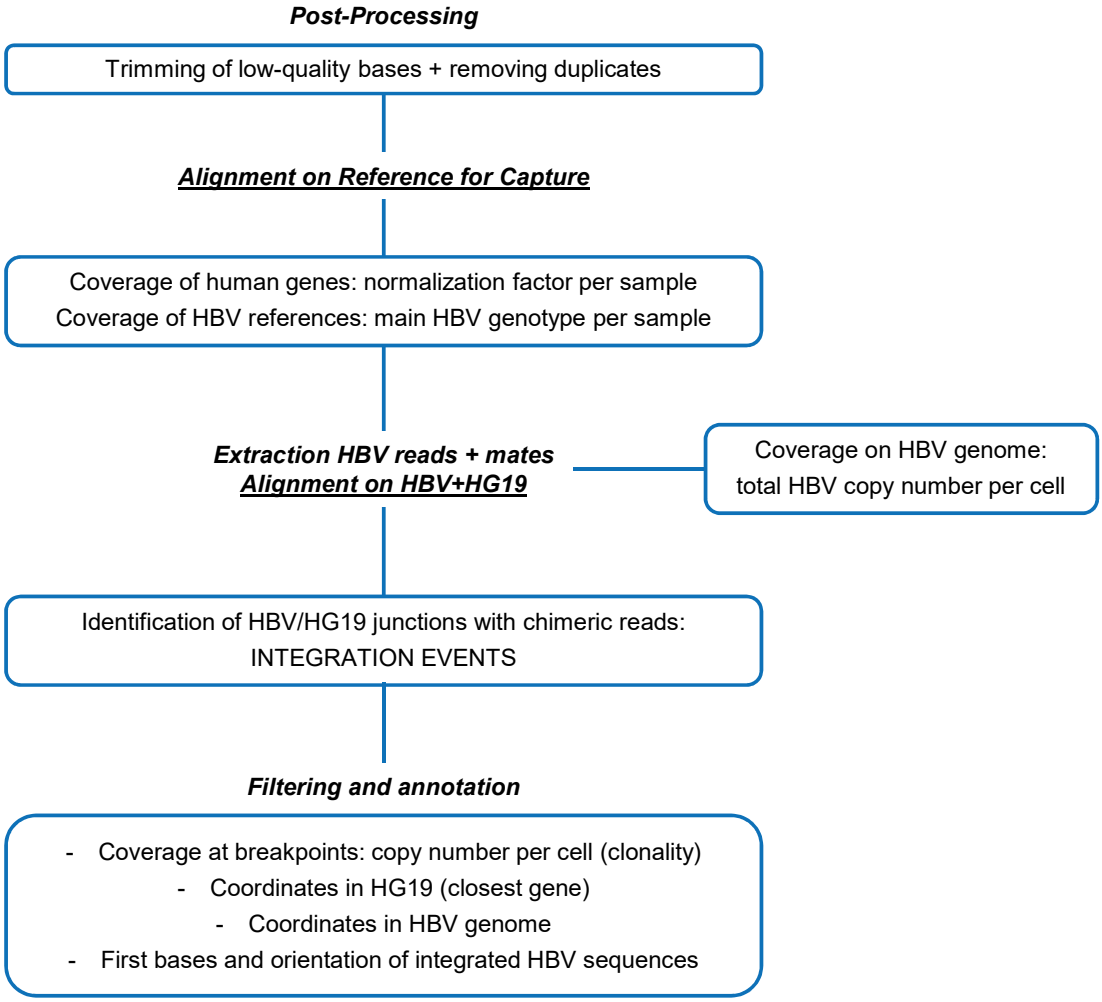
For statistical analysis, we used R version 3.6.0 (R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, (http://www.R-project.org). Wilcoxon, Fisher, Chi-square or Pearson's correlation statistical tests were applied with respect to the type of variable. Survival analysis was performed in patients treated for a primary HCC tumor by R0 liver resection as previously described[12]. We assessed overall survival defined by the interval between surgery and death. Survival curves were represented using the Kaplan-Meier method compared with the Log Rank test. Multivariate analysis was performed using Cox model. A p value <0.05 was considered as statistically significant.

References Supplementary Materials and Methods

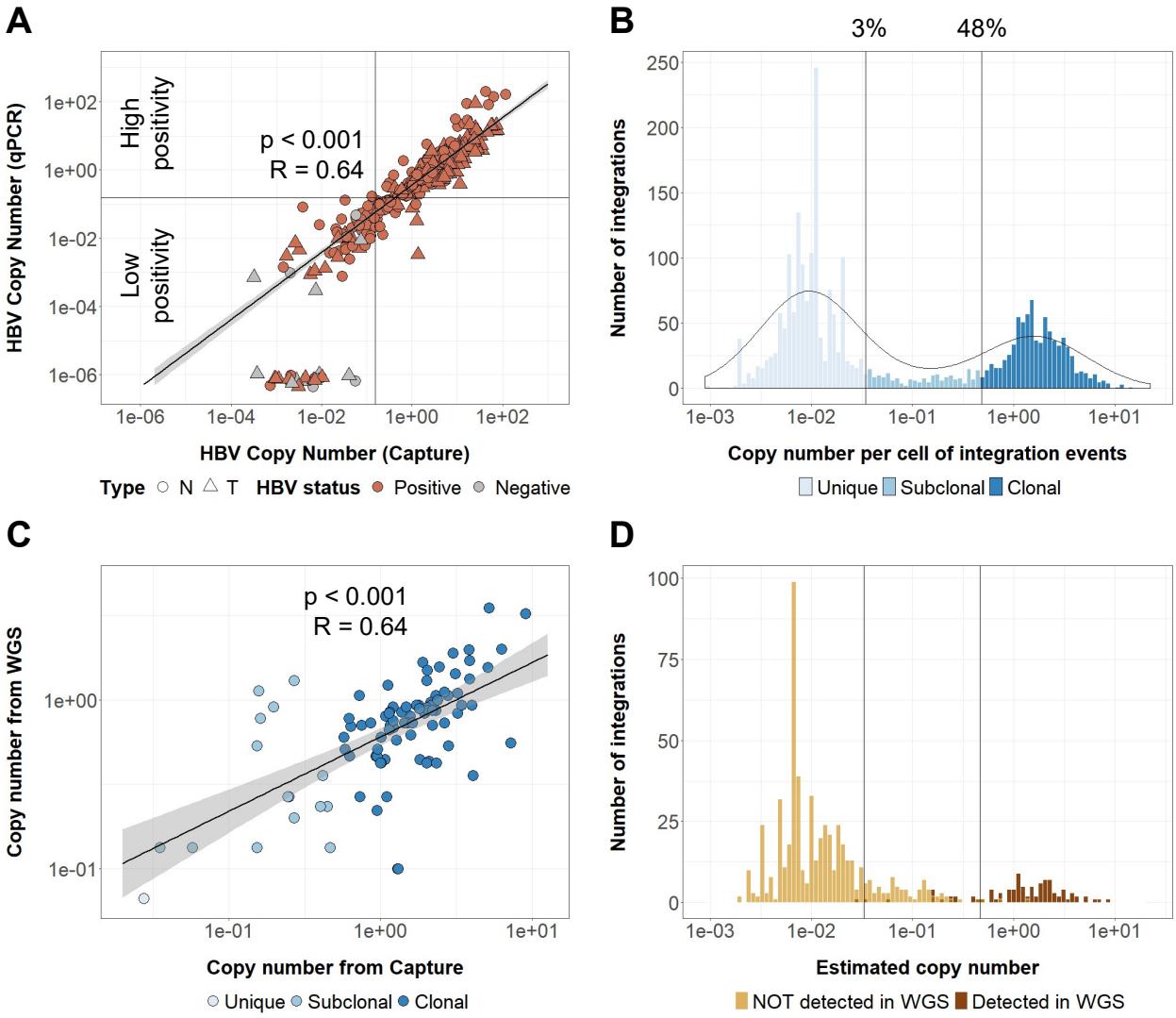
1. Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47**, 505–511 (2015).
2. La Bella, T. *et al.* Adeno-associated virus in the liver: natural history and consequences in tumour development. *Gut* **69**, 737–747 (2020).
3. Benaglia, T., Chauveau, D., Hunter, D. R. & Young, D. **mixtools** : An *R* Package for Analyzing Finite Mixture Models. *J. Stat. Softw.* **32**, (2009).
4. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
5. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
6. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
7. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
8. Letouzé, E. *et al.* Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* **8**, 1315 (2017).
9. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
10. Debacker, K. & Kooy, R. F. Fragile sites and human disease. *Hum. Mol. Genet.* **16**, R150–R158 (2007).
11. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
12. Nault, J. *et al.* Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. *Hepatology* **71**, 164–182 (2020).
13. Nault, J. C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* **4**, 2218 (2013).
14. Bayard, Q. *et al.* Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat. Commun.* **9**, 5235 (2018).
15. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
16. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
17. Hirsch, T. Z. *et al.* BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA. *J. Hepatol.* S0168827819307184 (2019) doi:10.1016/j.jhep.2019.12.006.
18. Boyault, S. *et al.* Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* **45**, 42–52 (2007).
19. Lazarevic, I. Clinical implications of hepatitis B virus mutations: Recent advances. *World J. Gastroenterol.* **20**, 7653 (2014).
20. Tong, S. & Revill, P. Overview of hepatitis B viral replication and genetic variability. *J. Hepatol.* **64**, S4–S16 (2016).
21. Nault, J.-C. *et al.* Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* **47**, 1187–1193 (2015).
22. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).



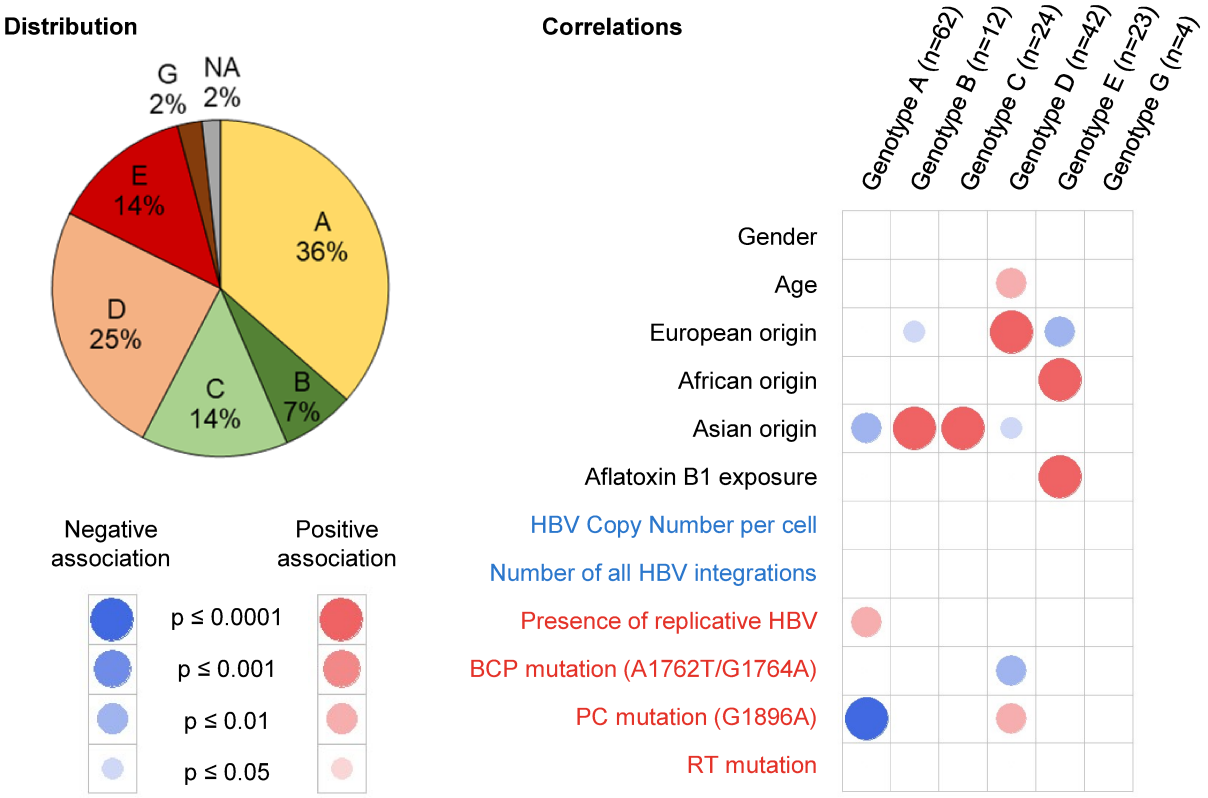
Supplementary figure 1 | Flow-chart of the study. Frozen paired HCC and adjacent liver tissues from 1128 patients were screened to detect the presence of HBV DNA by quantitative PCR with Taqman probes targeting 5 regions of the viral genome. Positive values were defined as low positive or high positive based on a bimodal distribution. Patients were then classified into 3 groups by considering for one patient the highest value between the corresponding two paired samples (tumor and adjacent liver tissue): high positive (at least 1 positive probe with a high positive value), low positive (at least 3 positive probes with low positive values), negative or low positive (0, 1 or 2 probes with low positive values). The positivity for AgHBs in patients' serum is indicated in red when available (See patient's characteristics in Supplementary Table 1). *HCC, hepatocellular carcinoma; N, non-tumor; T, tumor; NGS, next-generation sequencing; WGS, whole-genome sequencing; WES, whole-exome sequencing; RNASeq; RNA sequencing.*



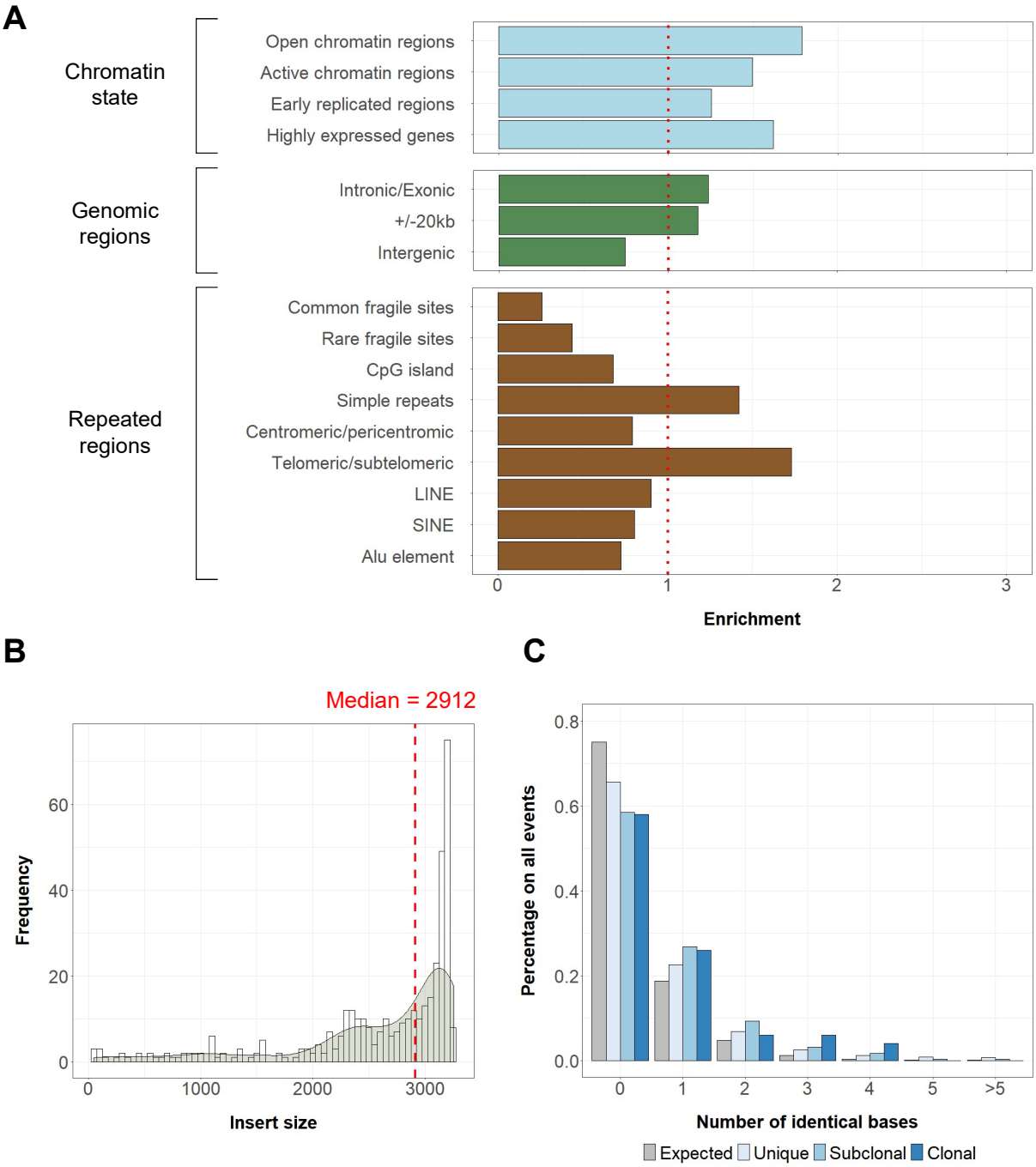
Supplementary figure 2 | Flow-chart of the HBV integrations analysis pipeline. The different steps are described in **Supplementary Materials and Methods**, with the references of the tools used. The list of the 8809 integration events obtained after filtering is detailed in **Supplementary table 4**.



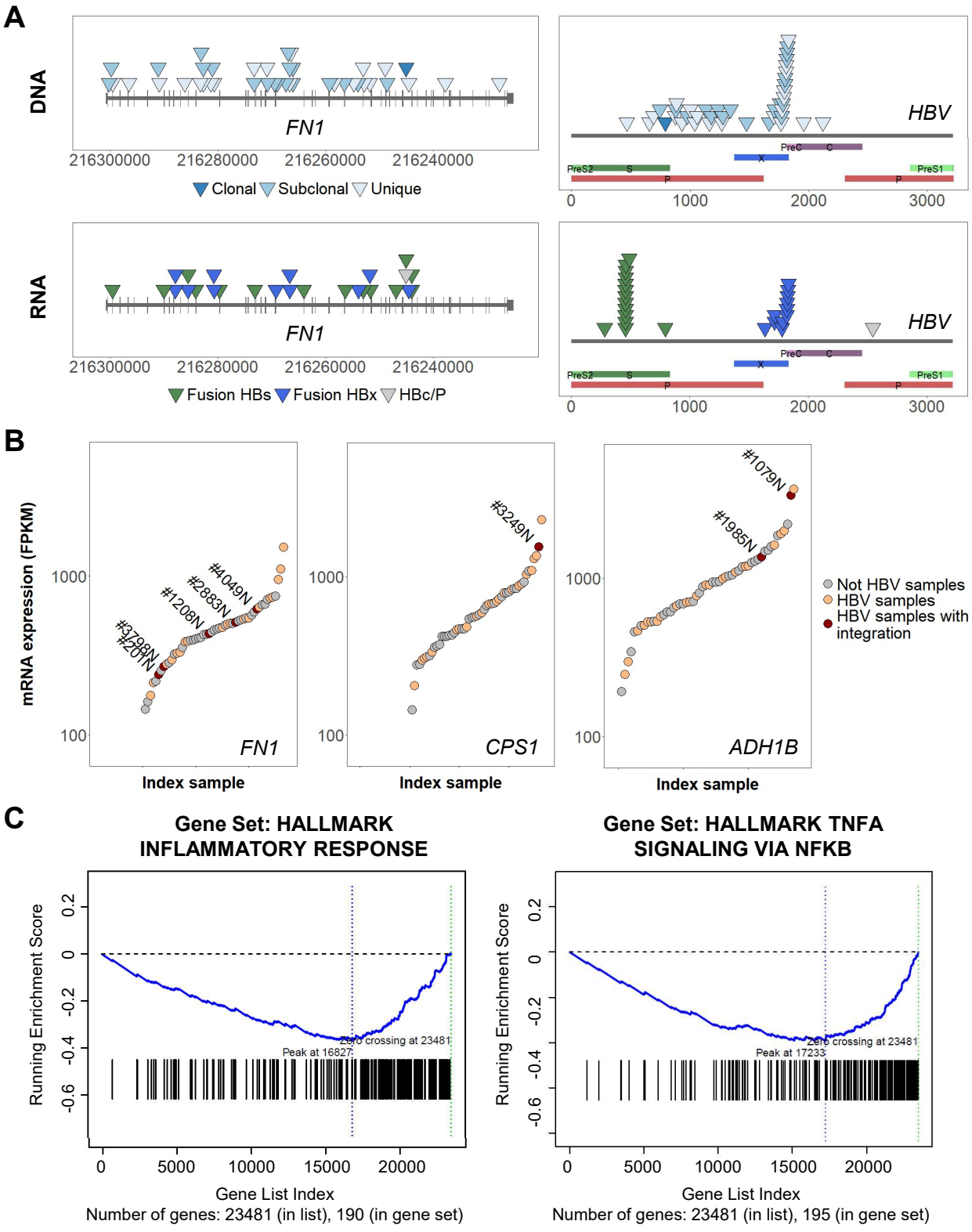
Supplementary figure 3 | Definition of clonality from viral capture. (A) Correlation between the values of HBV copy number per cell obtained from qPCR and viral capture in 347 samples from the Capture series of 190 patients (non-tumor liver, n=170, and tumor samples, n=177). The HBV status is annotated based on clinical data. The two techniques show a good correlation (Pearson correlation). (B) Definition of clonality with all 2199 integration events detected in tumors based on the k-means method. (C) Correlation between the values of copy number per cell for integration events obtained from WGS and from viral capture in 40 samples (non-tumor liver, n=20, and tumor samples, n=20). The two techniques show a good correlation (Pearson correlation). (D) Distribution of integration events detected in viral capture in 40 samples sequenced in WGS, based on their copy number per cell.



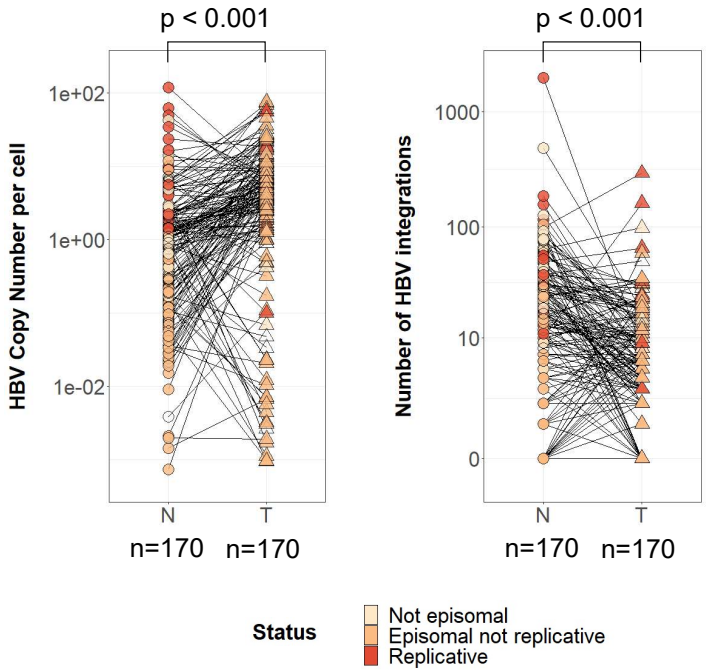
Supplementary figure 4 | HBV genotypes in non-tumor tissues. Correlation plot between the HBV genotypes identified from viral capture data and clinical or molecular features of non-tumor samples from the capture series (n=170). Blue circles indicate a negative association between features; red circles indicate positive associations. Color intensities represent different levels of statistical significance. Statistical analysis was performed using the chi-square test or Wilcoxon signed-rank test with respect to the type of variable. P values were adjusted for multiple testing using the Benjamini-Hochberg method (false discovery rate). A pie chart showing the distribution of HBV genotypes within the series is represented on the left side.



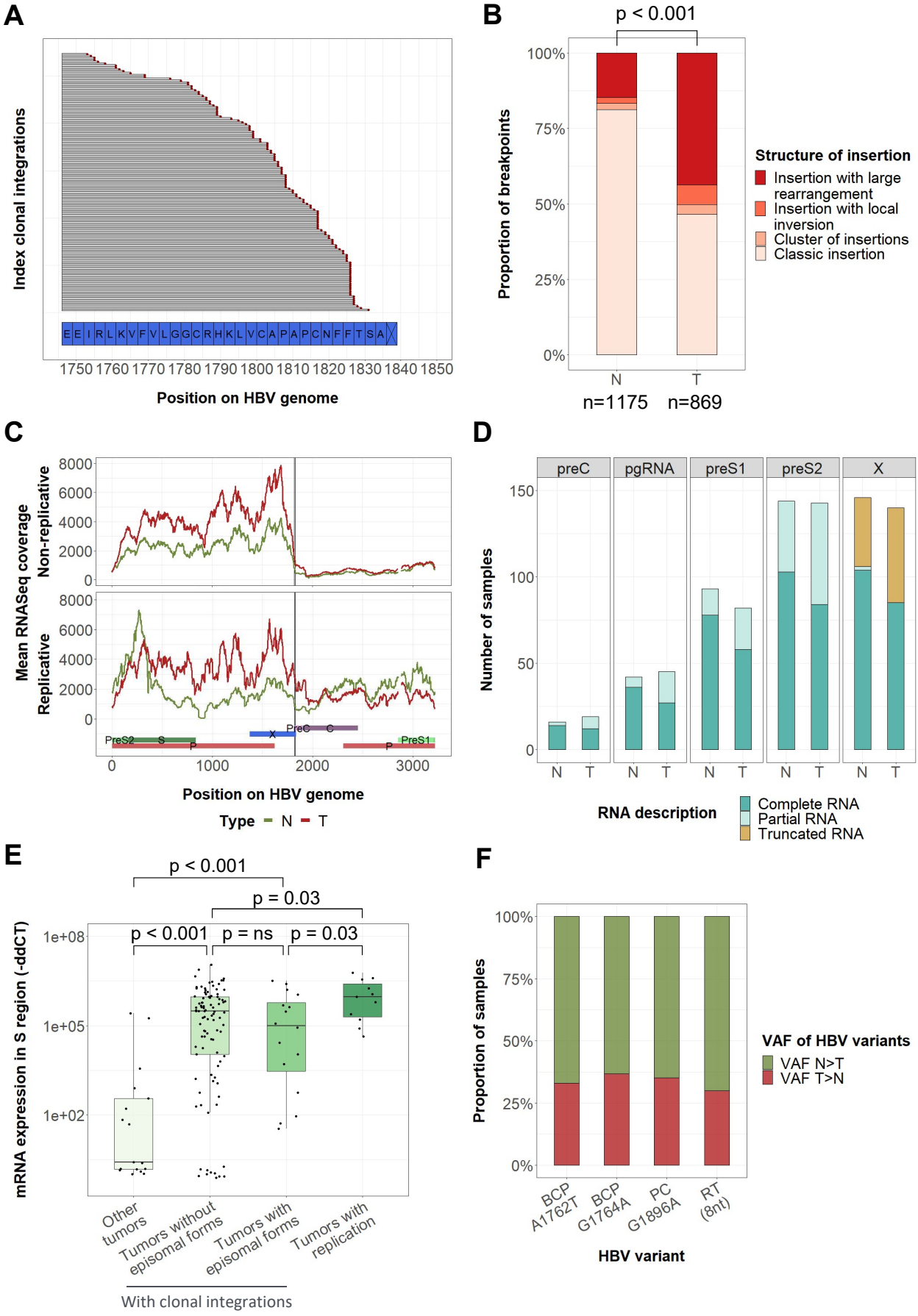
Supplementary figure 5 | Characterization of localization and sequences of HBV integrations in non-tumor samples. (A) Density of HBV integration breakpoints in a specific region compared to the complementary region. Regions were defined by chromatin state, genomic features and repeated motifs. (B) Size distribution of integrated sequences in simple subclonal or clonal integrations (n=394) identified in non-tumor tissues from the Capture series. (C) Homology analysis at HBV integration breakpoints between human genomic sequences and HBV integrated sequences, according to the clonality of the events.



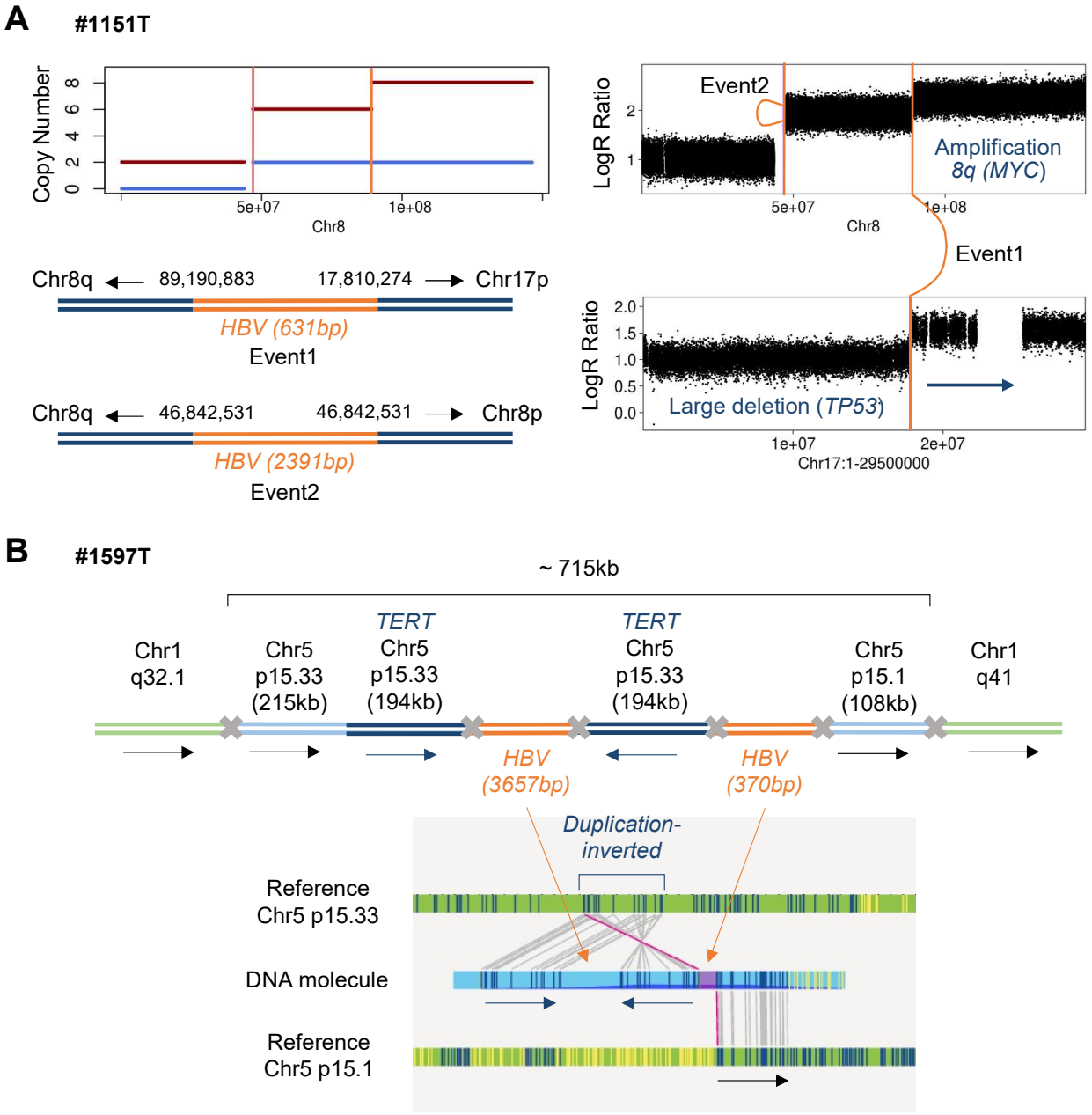
Supplementary figure 6 | Subclonal/clonal expansion in non-tumor tissues. (A) Breakpoints of HBV integrations in human genome at the *FN1* locus (left) and in HBV genome (right). Coordinates of breakpoints of HBV integrations in genomic DNA were identified from viral capture (up) and in transcripts from RNA-Seq (down). (B) mRNA expression for the 3 genes with integrations in more than two samples: *FN1*, *CPS1* and *ADH1B*, for 24 HBV-positive and 29 HBV-negative non-tumor liver samples. No RNAseq data was available for samples with integrations in *KCTN2*. (C) Gene-set enrichment analysis from RNA-Seq data, to compare HBV-positive non-tumor samples with a clonal HBV integration (n=5) or without (n=19). Two gene sets are shown to be downregulated in the first group: one gene set containing 190 genes defining inflammatory response (left) and one gene set containing 195 genes regulated by NF-κB in response to TNF (right).



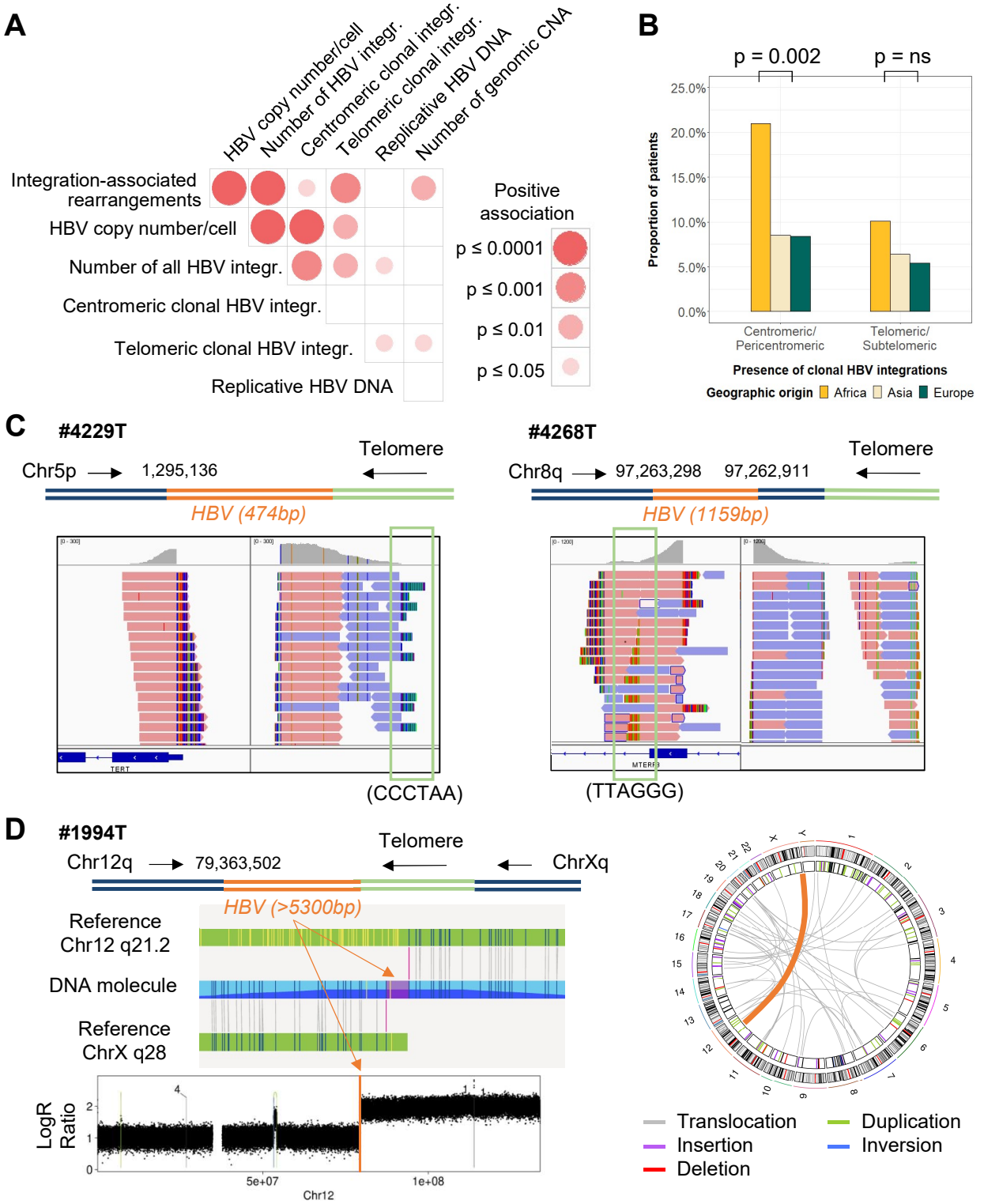
Supplementary figure 7 | Comparison of non-tumor and tumor samples. HBV copy number per cell (left) and number of HBV integration breakpoints (right) of paired tumor and non-tumor tissues of 170 HBV-positive patients from the Capture series (paired Wilcoxon signed-rank test).


A
Index clonal integrations
E E I R L K V F V L G G C R H K L V C A P A P C N F F T S A X
1750 1760 1770 1780 1790 1800 1810 1820 1830 1840 1850
Position on HBV genome
B
p < 0.001
Proportion of breakpoints
100%
75%
50%
25%
0%
N
T
n=1175 n=869
Structure of insertion
Insertion with large rearrangement
Insertion with local inversion
Cluster of insertions
Classic insertion
C
Mean RNASeq coverage
Non-replicative
Replicative
8000
6000
4000
2000
0
PreS2 S P X PreC C P PreS1
0 1000 2000 3000
Position on HBV genome
Type N T
D
preC pgRNA preS1 preS2 X
Number of samples
150
100
50
0
N T N T N T N T N T
RNA description
Complete RNA
Partial RNA
Truncated RNA
E
p < 0.001
p = 0.03
p < 0.001 p = ns p = 0.03
mRNA expression in S region (-ddCT)
1e+08
1e+05
1e+02
Other tumors
Tumors without episomal forms
Tumors with episomal forms
Tumors with replication
With clonal integrations
F
Proportion of samples
100%
75%
50%
25%
0%
BCP A1762T
BCP G1764A
PC G1896A
RT (8nt)
HBV variant
VAF of HBV variants
VAF N>T
VAF T>N

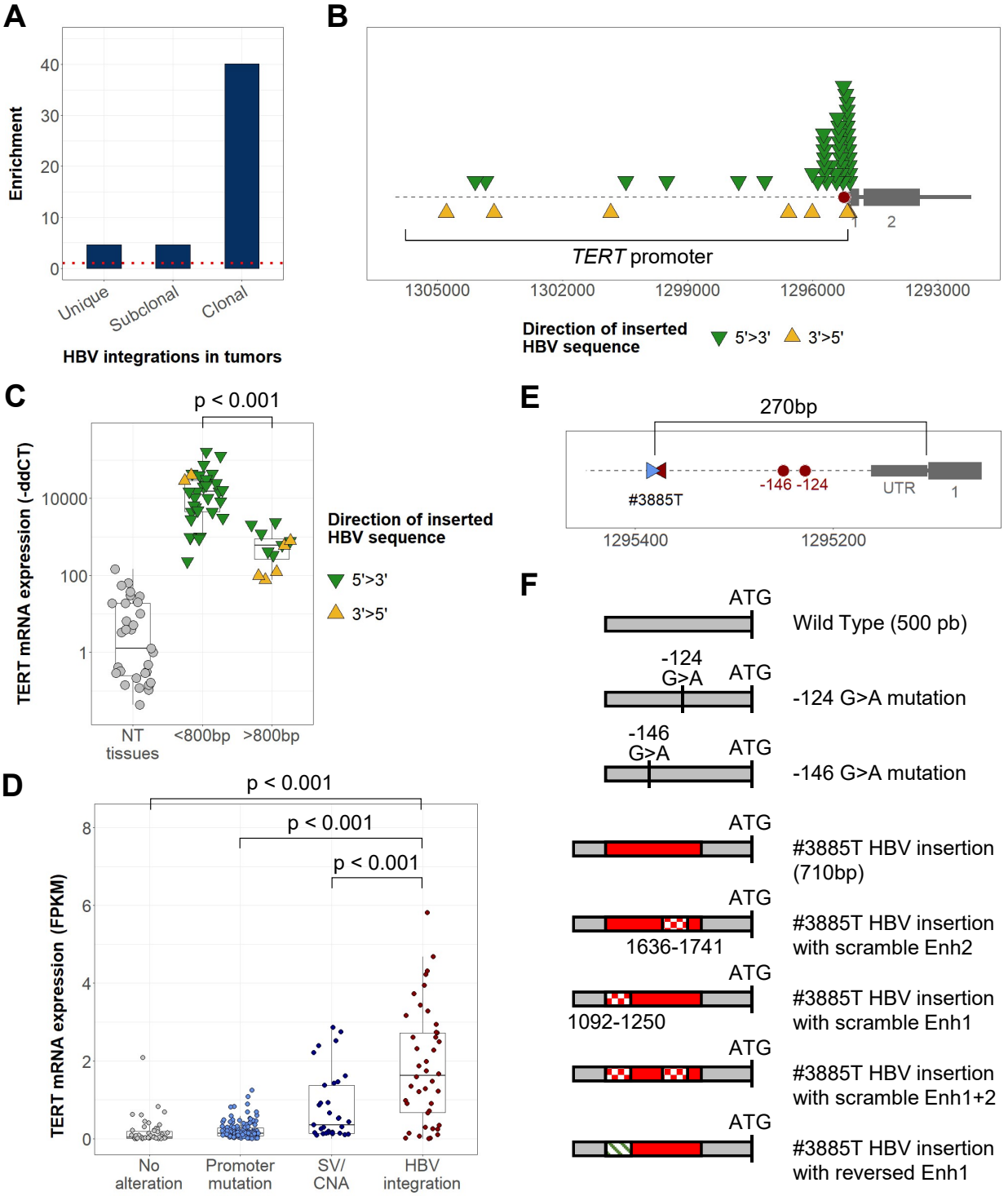
Supplementary figure 8 | Differences in HBV integrated sequences between non-tumor tissues and tumors. (A) HBV integrated sequences from 136 clonal events in tumors harboring a HBV/HG breakpoint around the 3' extremity of HBx (1750-1850) and with a 3'>5' orientation. Each line represents a different clonal integration and the breakpoints between viral and human genomes are represented with red dots. The last 30 amino acids of HBx protein are annotated below the graph. (B) Distribution of breakpoints in non-tumor and tumor tissues according to the structure of the clonal/subclonal integration (see **Materials and Methods** for definitions) (Chi-square test). (C) Mean HBV coverage from RNA-Seq of non-tumor samples (n=22) and tumors (n=73) according to the presence of replicative HBV DNA. (D) Number of samples positive for HBV mRNA according to the localization of the HBV probes. RNA are annotated as complete if a sample is positive for all probes until the polyA region on the HBV genome. (E) HBV mRNA expression (-ddCT) in the S region of HBV genome in tumors (Wilcoxon signed-rank test). (F) Proportion of non-tumor and tumor samples with HBV variants at different positions of the HBV genome. Variant Allele Frequency was determined in paired HCC and adjacent non-tumor tissue for 57 to 98 patients according to the position coverage on the HBV genome. *VAF, Variant Allele Frequency; BCP, Basal Core Promoter; PC, PreCore; RT, Reverse-Transcriptase; ns, not significant.*



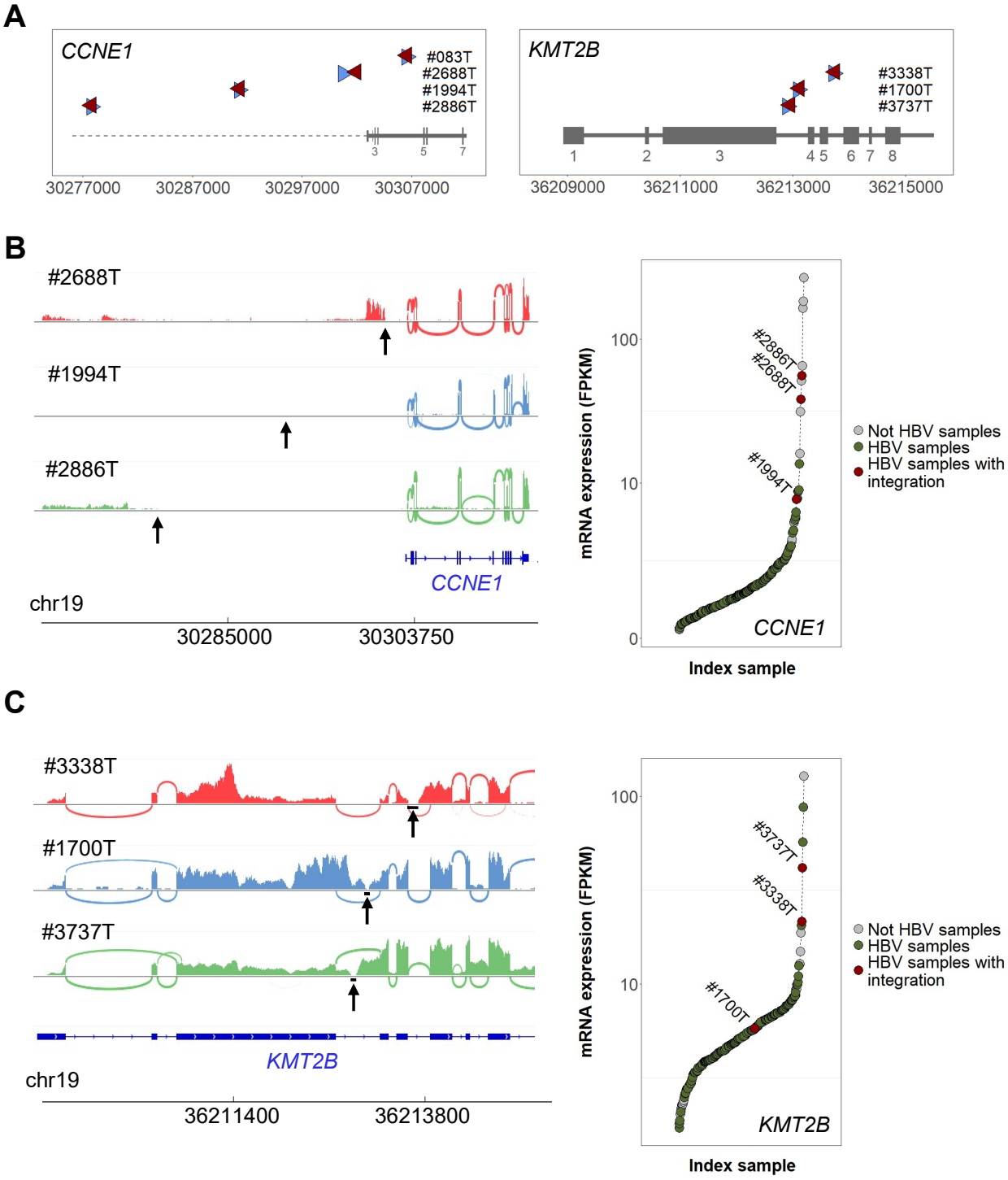
Supplementary figure 9 | Complex rearrangements involving integrations reconstructed with long-read and Bionano sequencing. (A) Reconstruction with long-read sequencing of two clonal integration events in tumor #1151T located in chr8q and inducing copy number alterations. Profiles of copy number (total in red, minor allele in blue) and LogR ratio are represented. Event 1 is a translocation-like event between chr8q and chr17p. Event 2 is a duplication-inverted-like event in the centromere of chr8. (B) Reconstruction with Bionano whole-genome sequencing of a complex rearrangement in tumor #1597T. The rearrangement is composed of two translocations between chr5p and chr1q, and two HBV clonal integrations inducing an amplification of *TERT*.



Supplementary figure 10 | HBV insertions in centromeric/telomeric regions. (A) Correlation plots between the presence of integration-associated rearrangements, viral features and the number of genomic CNA. All associations are positive and color intensities represent different levels of statistical significance. Statistical analysis was performed using the chi-square test, Wilcoxon signed-rank test or Pearson correlation with respect to the type of variable. P values were adjusted for multiple testing using the Benjamini-Hochberg method (false discovery rate). (B) Proportion of HCC according to the patients' geographic origin harboring HBV integrations in centromeric or telomeric regions (Fisher test). (C) Reconstruction of HBV clonal integration events involving telomeric regions in tumors #4229T and #4268T. Views from Integrative Genomics Viewer are shown to visualize telomeric sequences (green square). (D) Reconstruction with Bionano whole-genome sequencing of a translocation in tumor #1994T between chr12q and the telomeric region of chrXq. The logR Ration in chr12q is shown to see the CNA at HBV integration breakpoint. *CNA, Copy Number Alteration; ns, not significant.*



Supplementary figure 11 | HBV insertional mutagenesis: *TERT* activation. (A) Enrichment of HBV integration breakpoints in tumors around 72 HCC-associated genes (+/- 25kb). The HCC-associated genes were selected as previously described[45]. (B) Localization of HBV clonal integration breakpoints in the human genome in the promoter of the *TERT* gene. Red dots indicate the positions of classic mutations of the promoter (-124 and -146 from ATG) (C) mRNA expression (-ddCT from RT-qPCR) of *TERT* in HCC harboring a clonal HBV integration in the *TERT* promoter (n=48) and in the adjacent non-tumor liver samples (n=31) (Wilcoxon signed-rank test). (D) mRNA expression (log10(1+FPKM) from RNA-Seq data) of TERT in HCC harboring different alterations of the TERT promoter in the NGS Series (n=265) (Wilcoxon signed-rank test). (E) Localization of HBV clonal integration breakpoints in the promoter of the *TERT* gene for tumor #3885T. Red dots indicate the positions of mutations of the promoter (-124 and -146 from ATG). (F) Representation of plasmids constructs of TERT promoter (wild type, harboring -124 or -146 promoter mutations, harboring the HBV insertion identified in tumor #3885T wild type or containing scrambled/reversed Enhancer sequences). *SV, Structural Variant; CNA, Copy Number Alteration; Enh, Enhancer.*



Supplementary figure 12 | HBV insertional mutagenesis: *CCNE1* and *KMT2B*. (A) Localization of HBV clonal integration breakpoints in the human genome at the loci of *CCNE1* gene (left) and *KMT2B* gene (right). HBV integration in tumor #1994T have been previously described[21]. (B) Consequences of HBV integrations on *CCNE1* mRNA in 3 tumors. IGV-Sashimi plots show RNA-seq alignments (left) and sorted mRNA expression (FPKM) from RNA-Seq data of the NGS series (n=265) is represented (right). (C) Consequences of HBV integrations on *KMT2B* mRNA in 3 tumors. IGV-Sashimi plots show RNA-seq alignments (left) and sorted mRNA expression (FPKM) from RNA-Seq data of the NGS series (n=265) is represented (right). For IGV-Sashimi plots, alignments in exons are represented as read density, and alignments to splice junctions are shown as an arc connecting a pair of exons, where arc width is proportional to the number of reads aligning to the junction. The canonical transcripts for the genes are shown below. Black arrows indicate the position of HBV integration breakpoints. *FPKM, fragments per kilobase of exons per million reads.*