Virologica Sinica 39 (2024) 655–666

Contents lists available at ScienceDirect

Virologica Sinica

journal homepage: www.keaipublishing.com/en/journals/virologica-sinica
www.virosin.org

Research Article

Detection of HBV DNA integration in plasma cell-free DNA of different HBV diseases utilizing DNA capture strategy

Zerui Yang [a,1], Jingyan Zeng [b,1], Yueyue Chen [a,1], Mengchun Wang [a], Hongchun Luo [b], Ai-Long Huang [a,*], Haijun Deng [a,*], Yuan Hu [a,*]

[a] Key Laboratory of Molecular Biology on Infectious Diseases, Ministry of Education, Chongqing Medical University, Chongqing, 400016, China
[b] Department of Infectious Diseases, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, China

ARTICLE INFO

Keywords:
Hepatitis B virus (HBV)
DNA integration
Cell-free DNA (cfDNA)
DNA capture

ABSTRACT

The landscape of hepatitis B virus (HBV) integration in the plasma cell-free DNA (cfDNA) of HBV-infected patients with different stages of liver diseases [chronic hepatitis B (CHB), liver cirrhosis (LC), and hepatocellular carcinoma (HCC)] remains unclear. In this study, we developed an improved strategy for detecting HBV DNA integration in plasma cfDNA, based on DNA probe capture and next-generation sequencing. Using this optimized strategy, we successfully detected HBV integration events in chimeric artificial DNA samples and HBV-infected HepG2-NTCP cells at day one post infection, with high sensitivity and accuracy. The characteristics of HBV integration events in the HBV-infected HepG2-NTCP cells and plasma cfDNA from HBV-infected individuals (CHB, LC, and HCC) were further investigated. A total of 112 and 333 integration breakpoints were detected in the HepG2-NTCP cells and 22 out of 25 (88%) clinical HBV-infected samples, respectively. *In vivo* analysis showed that the normalized number of support unique sequences (*nnsus*) in HCC was significantly higher than in CHB or LC patients (P values < 0.05). All integration breakpoints are randomly distributed on human chromosomes and are enriched in the HBV genome around nt 1800. The majority of integration breakpoints (61.86%) are located in the gene-coding region. Both non-homologous end-joining (NHEJ) and microhomology-mediated end-joining (MMEJ) interactions occurred during HBV integration across the three different stages of liver diseases. Our study provides evidence that HBV DNA integration can be detected in the plasma cfDNA of HBV-infected patients, including those with CHB, LC, or HCC, using this optimized strategy.

1. Introduction

Hepatitis B virus (HBV) infection remains a severe global public health issue (Nguyen et al., 2020). Although vaccines can prevent viral infection, over 240 million people worldwide still suffer from chronic hepatitis B infection (CHB), a significant risk factor for liver cirrhosis (LC) and hepatocellular carcinoma (HCC) (Revill et al., 2016). Current antiviral therapies for CHB, such as nucleotide analogs (NAs) or interferon, cannot completely cure CHB due to the following reasons (Pollicino and Caminiti, 2021): (1) the persistent existence of covalently closed circular DNA (ccDNA); (2) chimeric proteins produced by integrated HBV DNA; and (3) the evasion of host immunity.

HBV genome integration has been reported to occur in the early stage of HBV infection (Chauhan et al., 2017). Double-stranded linear DNA (dslDNA), a minor species of HBV DNA formed during reverse transcription, is presumed to be the source of integration into the host genome (Yang and Summers, 1999; Tu et al., 2017). Circumstantial evidence suggests that dslDNA inserts into the host chromosome through non-homologous end-joining (NHEJ) or microhomology-mediated end-joining (MMEJ) to produce an integrated chimeric form (Yang and Summers, 1995; Gong et al., 1999). Although HBV integration is not essential for viral replication, it is believed to induce chromosome instability (Sung et al., 2012), insertional mutagenesis (Li et al., 2019), or activation of cancer-related genes (Schlüter et al., 1994). Therefore, viral integration is considered the main contributor to the progression of liver diseases such as HCC (Yeh et al., 2023).

In addition, the serum hepatitis B surface antigen (HBsAg) has been associated with persistent viral infection due to immunological exhaustion (Bertoletti and Ferrari, 2016). Therefore, HBsAg clearance and sustained loss of HBV DNA (functional cure) are recommended as the

* Corresponding authors.
E-mail addresses: huyuan@cqmu.edu.cn (Y. Hu), pepper027@163.com (H. Deng), ahuang@cqmu.edu.cn (A.-L. Huang).
[1] Zerui Yang, Jingyan Zeng, and Yueyue Chen contributed equally to this work.

https://doi.org/10.1016/j.virs.2024.06.003
Received 24 January 2024; Accepted 4 June 2024
Available online 7 June 2024

ideal endpoints for antiviral treatment of CHB (Song et al., 2021). HBV integration also involves an intact region encoding HBsAg, contributing to HBsAg production, indicating that the serum HBsAg is transcribed not only from intrahepatic ccDNA but also from integrated HBV DNA (Meier et al., 2021; Erken et al., 2022). A recent report suggested that most HBsAg originates from integrated HBV DNA during HBeAg-negative infection (Wooddell et al., 2017). Therefore, it is important to detect and compare the characteristics of HBV integration among different stages of chronic liver disease (CHB, LC or HCC), especially in HBeAg-negative patients, to distinguish the source of HBsAg, which would be beneficial for the optimization of antiviral therapy for CHB (Gao et al., 2023).

Recent studies have shown that integrated HBV DNA can be detected in liver biopsies of CHB patients using the PacBio long-read platform (van Buuren et al., 2022) or probe-based capture strategies (Zhang et al., 2022). However, liver biopsy handling may not be suitable due to its invasive nature. Circulating cell-free DNA (cfDNA), a blood-derived biomarker, has been widely used for the noninvasive diagnosis of tumors (Marzese et al., 2013; Luo et al., 2021). Due to its genetic information, cfDNA can be utilized for early cancer detection and monitoring of cancer development (Cai et al., 2019). To this point, cfDNA can serve as a noninvasive method for the quantification of HBV integration in HBV-infected patients.

Recently, a high-throughput viral integration detection (HIVID) method, based on viral probe-based capture and next-generation sequencing (NGS) has been utilized to enrich integrated HBV DNA (Li et al., 2013). Using this technology, HBV-host chimeric DNA fragments were successfully detected in the cfDNA of HCC samples (Chen W. et al., 2020; Li et al., 2020). However, whether HBV integration breakpoints can be detected in the cfDNA of other liver disease progression, such as CHB or LC samples, has not been reported yet.

In this study, we developed a workflow that combined an HIVID strategy and improved bioinformatics analysis to detect HBV integration in HBeAg-negative HBV-infected samples at different stages of liver disease progression, compared the characteristics of HBV integration among these groups, and conducted specific virus-host specific PCR to confirm these integration patterns. Our results demonstrated that HBV integration events could be detected in cfDNA derived from CHB patients, LC patients, or HCC patients by the developed HBV DNA detection strategy; thus, this method is a promising noninvasive tool for evaluating antiviral therapy efficacy.

2. Materials and methods

2.1. Patients and samples

A total of 25 patients were enrolled in this study at The First Affiliated Hospital of Chongqing Medical University between 2022 and 2023, including 9 patients with CHB, 10 with LC, and 6 with HCC (Supplementary Table S1). All of these patients were HBsAg positive and HBeAg negative. HCV- or HIV-positive patients were excluded. HCC diagnosis was based on typical imaging analyses consistent with the criteria of the Guidelines for the Diagnosis and Treatment of Primary Liver Cancer (Zhou et al., 2023). LC diagnosis was made through pathological examination or clinical evidence, including nodular or splenomegaly on liver imaging and/or thrombocytopenia (Yoshiji et al., 2021). CHB diagnosis followed the guidelines for the prevention and treatment of chronic hepatitis B, version 2022 (You et al., 2023). Serum from a healthy, HBV-uninfected individual was used as the negative control.

2.2. Virologic, serological, and biochemical assessments of clinical samples

The serum HBV DNA levels were quantified using the automated real-time PCR analyzer Cobas TaqMan 48 (Roche, Switzerland) with a lower limit of detection of 20 IU/mL. Quantitative serum HBsAg levels and the presence of HBeAg and anti-HBe were measured using the Abbott Architect immunoassay system (Abbott Laboratories, Chicago, USA). Blood biochemical parameters, including alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels, were measured using the automatic biochemical analyzer Cobas c701 (Roche, Switzerland).

2.3. Cell culture, virus concentration, and HBV infection

HepG2-NTCP cells were cultivated in Dulbecco's modified Eagle's medium (DMEM) (Gibco, Beijing, China) supplemented with 10% fetal bovine serum, 100 U/mL streptomycin, and 100 μg/mL penicillin at 37 °C in a 5% CO_2 incubator. HBV inoculum production and infection were conducted as previously reported (Chen Y. et al., 2020). Briefly, the HBV inoculum was concentrated from the supernatants of HepAD38 cells with Beckman Optima™ XPN-100 and stored at −80 °C until use. HepG2-NTCP cells were infected with 1000 virus genome equivalents (VGEs) of HBV particles per cell. For drug treatments, HepG2-NTCP cells were pretreated with 200 nM Myrcludex B (MedChemexpress, NJ, USA) for one day prior to virus infection. The virus parameters and HBV-DNA integrations were measured at 3 days post-infection (dpi).

2.4. Measurement of viral proteins (HBsAg and HBeAg), HBV DNA, and 3.5 kb HBV RNA

The cultured supernatants were collected for HBeAg or HBsAg measurement by using an enzyme-linked immunosorbent assay following the manufacturer's instructions (Kehua, Shanghai, China). HBV core-associated DNAs were extracted and detected as described previously (Chen Y. et al., 2020). HBV 3.5 kb RNA was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and reverse transcribed into cDNA with the PrimeScript RT Reagent Kit with gDNA Eraser (RR047A, Takara, Dalian, China) according to the manufacturer's instructions. HBV 3.5 kb RNA was quantified by real-time quantitative PCR (qPCR) as previously reported (Chen Y. et al., 2020).

2.5. Extraction of Cell Genomic DNA and Plasma Cell-Free DNA

Cellular genomic DNA was extracted with a QIAamp DNA Mini Kit (51304, Qiagen, Germany) according to the manufacturer's instructions. For plasma cfDNA extraction from clinical patients, 10 mL of nonhemolytic whole blood was drawn into an EDTA tube. The plasma was isolated by centrifugation at 16000 ×*g* for 10 min, after which the cfDNA was extracted using the MagMAX Cell-Free DNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions.

The cell number was calculated by quantifying the amount of β-globin based on an estimated 6.667 pg/hgDNA per cell, as previously reported (Luo et al., 2016). The digital drop PCR (ddPCR) assay was performed according to the manufacturer's instructions (Stilla Technologies, Paris, PAR, France). Briefly, 40 cycles of denaturation at 95 °C for 5 s, annealing at 58 °C for 15 s, and elongation at 72 °C for 10 s were performed at 95 °C for 30 s.

2.6. HBV Probe Capture and Next-generation Sequencing

The procedure of DNA capture-NGS for HBV integration site identification was described in a previous report (Chen W. et al., 2020). Viral probes for capture were designed according to different HBV genomes (A, B, C, D, E, F, G, and H) and produced by iGeneTech (iGeneTech, Beijing, China). Two hundred nanograms of genomic DNA from HepG2-NTCP cells was sheared into fragments at 150–200 bp, and the library was constructed with an IGT® Enzyme Plus Library Prep Kit V3 (iGeneTech, Beijing, China). Total cfDNA (ranging from 32.16 ng to 112.01 ng) from each clinical sample was directly used for library construction (IGT™ Fast Library Prep Kit v2.0, iGeneTech, Beijing, China) according to the manufacturer's instructions. Then, 200 ng of the HBV probe was added to the 750 ng sample library to target the entire HBV genome for sequence

capture and enrichment (TargetSeq One® Hyb & Wash Kit v2.0, iGeneTech, Beijing, China). High-throughput and high-depth DNA sequencing was performed on the NovaSeq 6000 platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads.

2.7. Data processing, alignment, and breakpoint detection

The strategy to analyze the sequencing data were established as following steps (Fig. 1): Pair-end reads were trimmed using Trimmomatic (v0.39) after quality control processing to remove adapter and low-quality sequences with parameters as follows: (1) adapter sequences were removed; (2) leading or tailing bases were trimmed if the base quality was below 15; (3) read ends were clipped off if the average base quality in a 5-bp window slid along the read fell below 15; and (4) the reads with a length less than 30 bp or an average quality score less than 15 were filtered. High-quality reads were mapped to the human genome with the *BWA-MEM* algorithm. After removing the perfectly matched reads, the unmapped reads were assembled into unique sequences by using BBAP software with default parameters (Lin et al., 2017). A modified reference that merged the human (UCSC hg19 assembly) and HBV (NCBI HBV genotyping reference) genomes was constructed locally. The assembled unique sequences were mapped to the modified reference by *BWA-MEM* with default parameters. The Picard MarkDuplicates toolkit (https://broadinstitute.github.io/picard/) was used for PCR duplicate removal, and the reads with identical starting and ending points were marked as PCR duplicates and merged into one unique sequence. To improve the sequence alignments, the GATK toolkit (McKenna et al., 2010) was applied to realign the assembled unique sequences for mapping error correction. The chimeric reads with breakpoints were identified if a part of the reads aligned to the HBV genome and the other part aligned to the human genome.

2.8. HBV DNA integration detection and annotation

The breakpoints of chimeric reads that aligned to both the human and virus genomes were considered candidate HBV DNA integration sites. The improved HBV DNA detection strategy was established with several optimization steps as follows: briefly, HBV DNA integration breakpoints located in the tandem repeat regions were discarded. The neighboring HBV DNA integration events with a distance of less than 200 bp were merged, and integration breakpoints located in homologous regions were also merged according to sequence similarity. Homologous regions were determined as follows: 100 bp 5′- and 3′- flanking region sequences of the integration breakpoint were extracted from the human reference genome, and BLAT was used to evaluate the similarity between all flanking sequences and sequences with identities $\geq$ 95% were considered homologous regions. After that, the merged integration breakpoints were considered to be HBV-integrated events. To evaluate the integration levels, we calculated the normalized number of support unique sequences (*nnsus*) as follows (Li et al., 2013):

$$nnsus = \frac{nsus}{teus} * 10^6$$

Although *nsus* is the number of support unique sequences, *teus* is the total of effective unique sequences.

Each HBV DNA integration breakpoint was annotated based on the transcription start site (TSS) from the UCSC genome annotation. The classification of genomic regions was determined based on the genome annotation (version 31, build hg19), 2500 bp upstream from TSS was defined as promoter region. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the R package clusterProfiler (Wu et al., 2021), and the waterfall map of the recurrent integration events located in the genes was generated with the package GenVisR (Skidmore et al., 2016).

2.9. The evaluation accuracy of the strategy

To control the background noise and determine the accuracy of this experiment, we synthesized 30 human-HBV DNA chimeric artificial sequences of 242 bp from previous constructed library. The detailed information of these chimeric artificial sequences was listed in Supplementary Table S2. Briefly, the samples were divided into the following groups: (1) about total 1,000,000 copies of 30 artificial chimeric HBV sequences (each one was 33,000 copies) were mixed into 3300 ng of human genomic DNA as the positive group; (2) about 3300 ng of human genomic DNA and 1,000,000 copies of HBV DNA were mixed; (3) 3300 ng of human genomic DNA was only input; and (4) 1,000,000 copies of HBV DNA was only input. Then we determined the value of limit of detection (LoD) by mixing about 1000 ng of human genomic DNA with 5 equal chimeric HBV DNA (total copies range from 30 to 300,000). All test samples were enriched by HBV probe capture, and libraries were constructed and sequenced on Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) sequencers using a 150-cycle paired-end sequencing kit. Approximately 40 million reads were obtained for the simulated sample, and we detected the breakpoints using an established private pipeline as described above. After that, we determined the background noise and accuracy of this strategy. HBV integrations with *nsus* $\geq$ 3 and *nnsus* $\geq$ 0.7 were defined as true signals for the following data analysis.

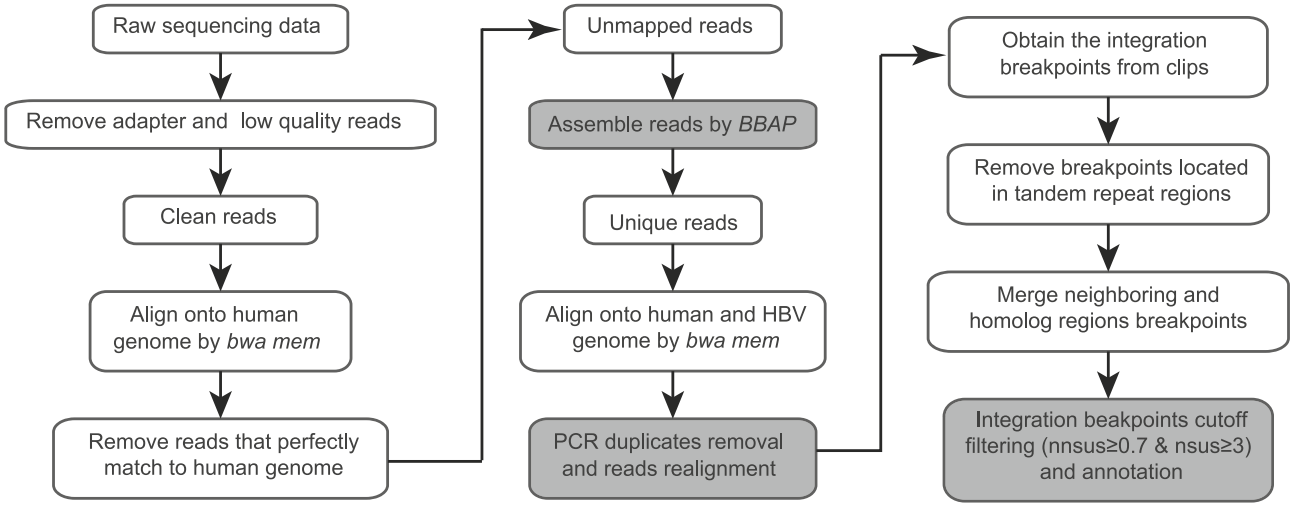

Fig. 1. Bioinformatic workflow for the identification of HBV DNA integration sequences.

2.10. Integration events validation by virus–host specific PCR and sanger sequencing

Virus-host specific PCR and Sanger sequencing were employed to verify the HBV integration breakpoints of the selected samples. The specific PCR primers used were designed based on virus-host chimeric sequences, in which one primer targeted the HBV genome and the other targeted the human genome. Generally, the PCR program consisted of 95 °C for 3 min, followed by 30 cycles of denaturation at 95 °C for 10 s, annealing for 30 s, and extension at 72 °C for 1 min. The PCR amplification products were subjected to Sanger sequencing to verify the presence of the virus-host chimeric sequence. The primer sequences are shown in Supplementary Table S3.

2.11. Statistical analysis

Variables are expressed as the median (range) or mean ± standard deviation. The statistical analyses were performed with GraphPad Prism 8.0. Statistical analysis was performed using one-way ANOVA with Tukey's multiple comparisons test. A two-tailed Pearson correlation test was used to calculate the correlation coefficient of the number of integration events and clinical variables. A P value < 0.05 was considered to indicate statistical significance for all the statistical tests.

3. Results

3.1. Developing HBV DNA integration breakpoint analysis to detect HBV integration events in vitro

Previous reports have used HIVID for enriching integrated HBV DNA and analyzing HBV integration events (Li et al., 2013). In this research, to develop a robust HBV DNA integration detection workflow, we utilized the DNA capture-NGS method and optimized the HIVID strategy with modifications in reads assembly, the merging of neighboring and homolog regions, cut-off filtering, and integration annotation steps (Fig. 1).

The background noise was determined (*rnnsus* ≥ 0.7 were defined as true signals, Supplementary Fig. S1A) and the value of LoD was 26.39 copies/μL (Supplementary Fig. S1B). Then we compared the sensitivity and accuracy of our established detection strategy with those of HIVID by using chimeric artificial DNA samples. The accuracy was 96.8% *vs* 70.00%, and the positive predictive value (PPV) was 93.3% *vs* 73.7% (Supplementary Figs. S1C and 1D), which indicated that the optimized detection strategy was superior to that of HIVID.

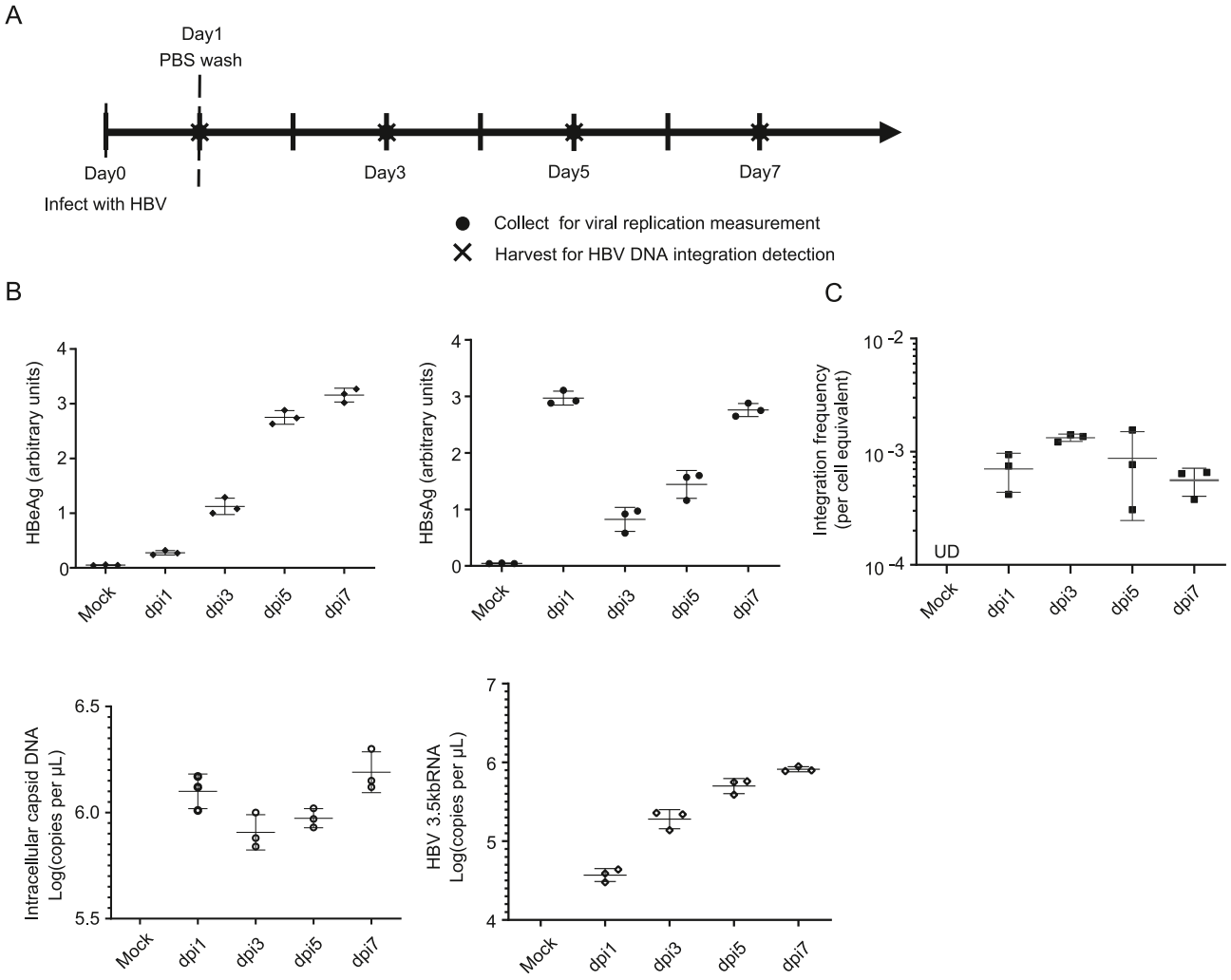

Fig. 2. HBV integration occurs at an early step in the HBV infection cell model. **A** HepG2-NTCP cells were infected on day 0 with an HBV inoculum of 1000 virus genome equivalents (VGE) per cell and washed off at 1-day post-infection (dpi). The cells and supernatants were harvested at 1, 3, 5, and 7 dpi. **B** Kinetics of HBV replication parameters. Cell supernatants were collected and analyzed by ELISA for secreted HBsAg and HBeAg, and cell lysates were collected to measure intracellular 3.5 kb HBV RNA and core-associated DNA. **C** HBV DNA integration detection at different time points in the HepG2-NTCP infection model. Mock, negative control represented the HepG2-NTCP cells with no HBV particle infection. UD, undetected. The data are presented as the means ± SDs of three experiments.

We first investigated whether this strategy could detect HBV integration events in HepG2-NTCP cells. Previous reports have demonstrated that HBV integration events occur in the early stages of infection, as detected by inverse nested PCR (Tu et al., 2018). Briefly, HepG2-NTCP cells were infected with HBV particles at 1000 virus genome equivalents (VGE) per cell, and then the cells were washed and cultured for seven days. The intracellular viral DNA, RNA, or viral proteins in the supernatant were collected and measured at 1, 3, 5, or 7 dpi, as depicted in Fig. 2A. The levels of intracellular 3.5 kb HBV RNA and secreted HBeAg were much lower at 1 dpi and increased markedly at 3 dpi, indicating that the levels of ccDNA or transcription of ccDNA may be lower at 1 dpi. In contrast, the high amounts of intracellular HBV DNA and secreted HBsAg at 1 dpi were likely derived from incoming viral particles. After 3 dpi, the levels of intracellular HBV DNA continued to increase (Fig. 2B).

The HBV integration breakpoints were quantified by the optimized strategy described above. In three independent experiments, a total of 112 integration breakpoints were detected in 12 samples, with

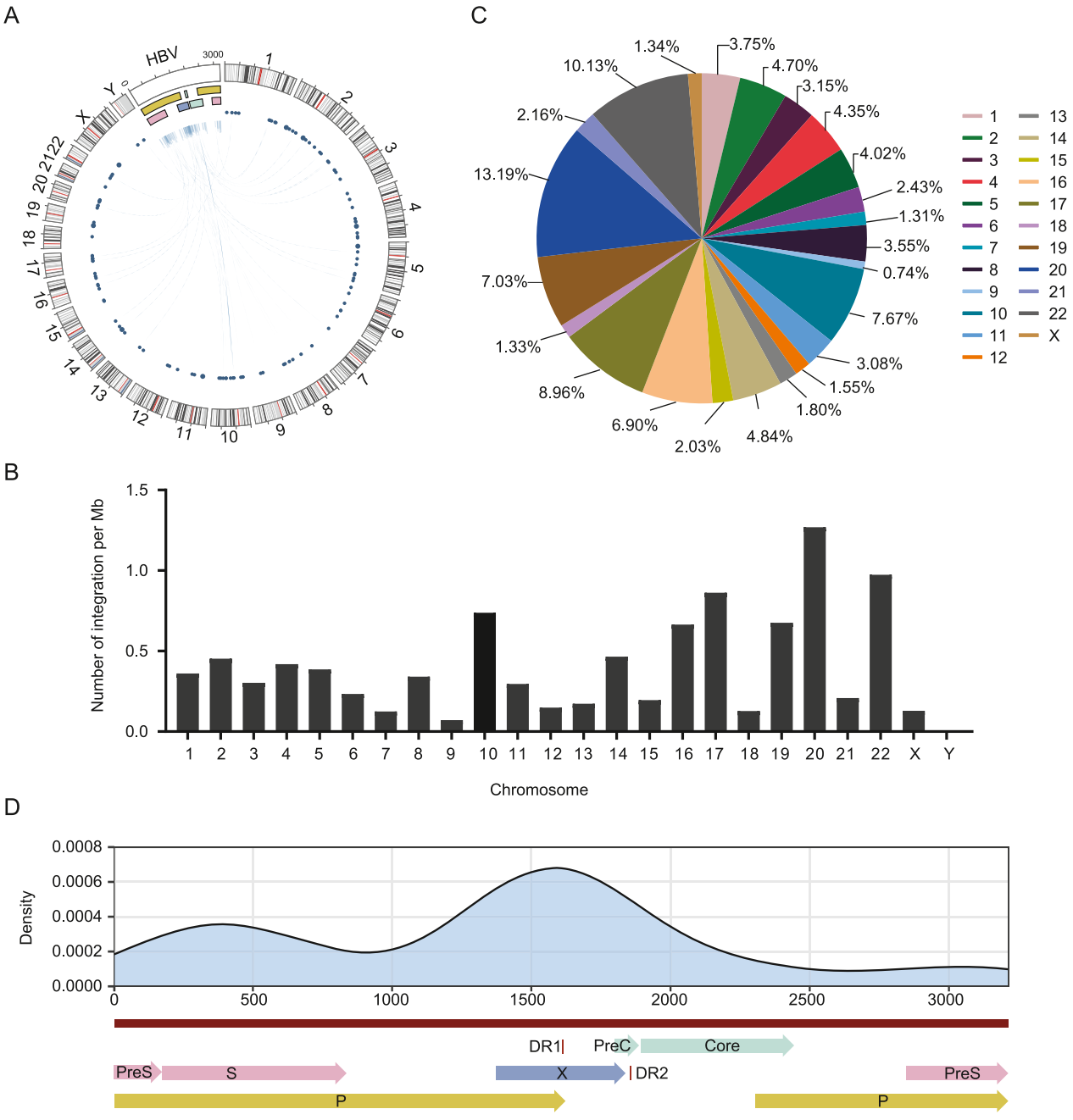

Fig. 3. Distribution characteristics of HBV DNA integration breakpoints detected in HepG2-NTCP cells. HepG2-NTCP cells were infected with an HBV inoculum of 1000 VGE per cell and washed off at 1-day post-infection (dpi). Cells were collected at 1, 3, 5, and 7 dpi, respectively. HBV integration breakpoints were analyzed by the optimized strategy. The total 112 breakpoints detected from above timepoints from three independent experiments are summarized and mapped here. **A** The distribution of 112 integration breakpoints in the HBV genome and human chromosomes is shown in circos plot. Each line represents an HBV DNA integration event, with one end displaying breakpoints in the HBV genome and the other end displaying breakpoints in the human chromosome. The size of the dots indicates the frequency of HBV DNA integration events, and human chromosome numbers are shown on the outer rim. **B, C** The distribution (**B**) and proportion (**C**) of HBV DNA integration events occurring on human chromosomes. **D** The distribution density of HBV DNA integration breakpoints observed in HepG2-NTCP cells is mapped in the HBV genome.

integration rates of approximately 7.58 integrations/10^4 cells (ranging from 3.07 to 15.5/10^4 cells). Integration events were detected at 1 dpi, with an average integration rate of 7 integrations/10^4 cells, which increased significantly to 13.3 integrations/10^4 cells at 3 dpi (Fig. 2C). Although the expression of viral RNA, DNA, and proteins continuously increased remarkably from 3 dpi, the integration events remained at a stable level. In another experiment, HepG2-NTCP cells were pretreated with HBV entry inhibitor Myrcludex. The treatment effectively suppressed HBeAg levels (Supplementary Fig. S2), and the integration rate in Myrcludex-treated HepG2-NTCP cells was approximately 0.08 integrations/10^4 cells at 3 dpi. Those results indicated that the optimized strategy could detect the HBV integration at the early stage of HBV infection in the HepG2-NTCP cell.

3.2. Analysis of HBV integration characteristics in vitro

Next, the distribution of HBV DNA integration breakpoints *in vitro* was mapped to human chromosomes and depicted in circos plot. As shown in Fig. 3A, the 112 HBV integration breakpoints were randomly distributed on the whole human chromosome. After normalization of the number of integrations per length of each chromosome, we observed that HBV integration was dispersed across the whole cellular genome (Fig. 3B). In total, the distribution frequency for each chromosome was no more than 15%. The highest frequency was distributed on chromosome 20 (13.19%), followed by chromosome 22 (10.13%) and chromosome 17 (8.96%) (Fig. 3C). The genome annotation of host breakpoints revealed that all breakpoints are located in different host genes (Supplementary Table S4), suggesting that HBV integration events in the early stages of infection were not the specific chromosomal hot spots. Furthermore, we analyzed the distribution of 112 HBV integration breakpoint sites in the HBV genome. Obviously, the distribution density of integration breakpoint sites around nt1600 was relatively greater across the whole HBV genome. More than one-third of the integration events occurred between nt1500–1900 (42/112, 37.5%) in the HBV genome, possibly due to the presence of DR1 and DR2 regions near nt1500–1900. In addition, we observed that a small peak of integration events occurred between nt200–400 (17/112, 15.2%), which is within the coding region of HBsAg (Fig. 3D).

To confirm the presence of HBV integration breakpoints detected by the improved quantification strategy, we designed a pair of primers (Supplementary Table S3) and conducted virus-host specific PCR to amplify the NGS-sequenced cDNA of *FNDC4* at 1 dpi. Sanger sequencing analysis revealed that human chromosome 2 harbors a partial 35 bp fragment of HBV DNA (Supplementary Fig. S3). Two-nt homologous sequences were observed at the virus-host junction sites, which indicated that this integration belongs to the MMEJ pathway.

3.3. Detecting HBV integration breakpoints in cfDNA of clinical samples

Next, to investigate the occurrence of HBV integration in the plasma cfDNA of HBV-infected individuals, 25 patients with different disease stages (9 patients with CHB, 10 patients with LC, and 6 patients with HCC) were included in this research. The clinical characteristics of the patients are shown in Table 1 and Supplementary Table S1. Among the 25 patients, there were 12 males and 13 females, with a median age of 51 (ranging from 35 to 71) years. All patients were HBsAg positive (ranging from 2.05 to 4.40 log IU/mL), and HBeAg negative. The HBV DNA viral loads ranged from undetectable (patients 58#, 63#, and 71#) to 6.2 log IU/mL.

The cfDNA of these 25 patients was extracted, enriched, sequenced, and analyzed using the optimized integration detection strategy as described above. The average number of raw reads for each sample was 30.4 million, and the average sequencing depth of the target area reached 25 × . There was no significant difference in the sequencing depth of samples among the three disease stages (Fig. 4A). Similarly, after filtering reads that perfectly matched the human or HBV genomes, a total of 333 HBV integration events in 22 samples (88%) were identified, with an average of 15.14 breakpoints per sample (range, 0–79/sample). There were 74, 217, and 42 integration events detected in CHB, LC, and HCC patients, respectively. The details are listed in Supplementary Table S5. CHB patients had a median of 4 (range, 0–26), LC patients had a median of 9 (range, 0–79), and HCC patients had a median of 5 (range, 3–15). Although there was no significant difference in the number of integrated breakpoints between these different disease stages (Fig. 4B), the number of *nnsus* in HCC patients was much greater than that in CHB or LC patients (P values < 0.05, Fig. 4C).

The associations of integration numbers with clinical parameters were also analyzed. Among the samples with HBeAg-negative status, the level of HBV DNA in the serum was not associated with HBsAg. The presence of integrations was not associated with the serum levels of HBV DNA, HBsAg, ALT, AST, or AFP (Supplementary Fig. S4).

3.4. Profile of HBV integrated breakpoints on HBV and the human genome detected in cfDNA

The successful detection of HBV integration events in CHB, LC and HCC patients makes it possible to compare the characteristics of HBV integration across these different liver disease stages. Again, we first mapped the entire HBV genome location of the 333 breakpoints detected in CHB, LC, and HCC patients. As shown in Fig. 5A, all 333 integration sites were randomly distributed in the HBV genome. Among these integration events, 27.93% (93/333) were significantly enriched between nt1500 and 1900 of the HBV genome, particularly between nt1700 and 1850 (Fig. 5B). We also investigated the frequency of integration breakpoints around HBV-encoding genes, and found that most of the HBV integration breakpoints in HBV-infected samples were more prevalent in PreC, with 10.2, 35.6, and 15.3 events per kb in CHB, LC, and HCC, respectively. The second most frequent region was HBx, accounting for 6.0, 14.6, and 4.7 events per kb in CHB, LC, and HCC, respectively (Fig. 5C). In terms of the distribution in the human genome, although chromosome 17 accounted for the largest proportion of genes associated with HCC (Fig. 5D), a significant difference analysis revealed that these

Table 1
Characteristics of HBV-infected patients.

Characteristics	ALL (n = 25)	CHB (n = 9)	LC (n = 10)	HCC (n = 6)	*P* value
Gender, Female/Male (%)	13/12 (52%)	4/5 (44.4%)	4/6 (40%)	5/1 (83.3%)	$P > 0.05$
Age (years)	51 (35–71)	49 (35–71)	52 (45–66)	53 (42–56)	$P > 0.05$
Serum HBV DNA (Log_{10} IU/mL)	4.08 (UD–6.2)	4.25 (UD–6.2)	2.23 (UD–6.06)	3.41 (UD–4.48)	$P > 0.05$
Serum HBsAg (Log_{10} IU/mL)	3.23 (2.05–4.40)	3.37 (2.40–4.40)	3.14 (2.20–4.32)	3.31 (2.05–3.53)	$P > 0.05$
AST(U/L)	115 (17–1028)	127 (17–1028)	137 (30–970)	69.5 (34–236)	$P > 0.05$
ALT(U/L)	182 (15–2192)	335 (16–2192)	128 (15–1102)	64.5 (35–229)	$P > 0.05$

The data are expressed in median of each group (range). N% in brackets is the percentage of female patients in each group. AST, aspartate aminotransferase; ALT, alanine aminotransferase, UD, undetectable.

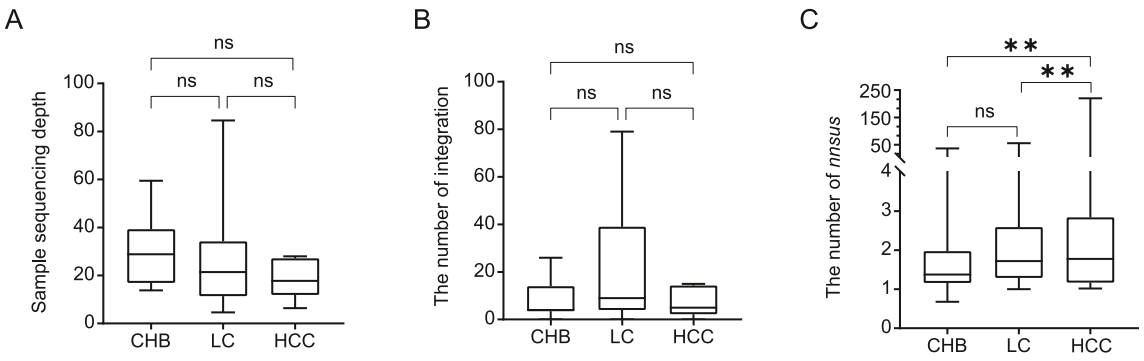

Fig. 4. Comparisons of the capture sequencing data across different stages of liver diseases. Patients with different disease stages (9 patients with CHB, 10 patients with LC, and 6 patients with HCC) were included in this research. The cfDNA was extracted and analyzed using the optimized integration detection strategy. Comparisons of the sequencing depth (**A**), number of integrations (**B**), and number of normalized number of support unique sequences (*nnsus*) (**C**) at CHB, LC, or HCC. Statistical analysis was performed using one-way ANOVA with Tukey's multiple comparisons test. A *P*-value <0.05 was considered to indicate statistical significance; **, *P*-values <0.01; ns, not significant.

integration breakpoints were randomly distributed across each chromosome (*P*-values >0.05).

3.5. Functional enrichment analysis of HBV-integrated genes

There were 55, 161, and 36 genes associated with HBV DNA integration sites within 15 kb in CHB, LC, and HCC, respectively (Supplementary Table S5). Functional enrichment analysis of the KEGG pathway revealed that the genes associated with the three disease stages were enriched in the mTOR signaling pathway, regulation of autophagy, and adipocytokine signaling pathway (Fig. 6A). GO enrichment analysis demonstrated that the integrated genes were enriched in processes such as oxygen carrier activity, regulation of cell maturation, and regulation of autophagy (Fig. 6B).

Next, we investigated the gene location of the integration breakpoints in the cellular genome. The results of the genomic annotation analysis revealed that *AC012494.1* was the most abundant gene (Fig. 6C), which was detected in 8 out of 25 patients. Lysine-specific methyltransferase 2B (*KMT2B*, known as *MLL4*) was detected in HCC samples (#72), and was reported to be related to the occurrence and development of HCC (Saigo et al., 2008; Dong et al., 2015; Zheng et al., 2021). Functional region analysis on the proportion of HBV integration breakpoints in the coding and noncoding regions revealed that the integration breakpoints were preferentially located in the gene-coding region (61.86%) (Fig. 6D). A comparable distribution was observed in the CHB, LC, and HCC groups (Fig. 6E). These results indicated that HBV DNA integration into host genome structural regions was preferentially located in the gene-coding region, which was independent of HBV-associated liver diseases.

3.6. Validation of the integration events in cfDNA identified by Capture-NGS

Finally, the sequences of the HBV and human genomes in each read of cfDNA samples from HBV-infected patients were analyzed. The representative results for HBV-infected patients with CHB, LC or HCC are shown in Fig. 7. The mechanism of HBV integration through the NHEJ pathway can be referred from the HBV-host junction sites in different liver stages, such as 67# in CHB, 12 # in LC, and 72# in HCC, as a previous report indicated that HBV integration occurs through the NHEJ pathway with limited homology sequences (<2 nt) in the virus-host junction sites (Tu et al., 2018). Another type of integration mechanism, the MMEJ pathway, could also be found in these clinical samples as HBV integration breakpoints harbor several homologous bases. For example, the insertion region of *TRIM56* on human chromosome 7 harbors a 3-bp homolog at the site of integration with the HBV genome, which indicates that this is due to the MMEJ pathway (Fig. 7B, 55# in CHB). These data suggested that both the NHEJ and MMEJ pathways are involved in HBV integration *in vivo*. Second, the majority of chimeric sequences harboring single breakpoints (284/333, 98.5%) were directly detected by this improved strategy, while complete integration events (51/333, 15.3%) could be inferred by combination with two single breakpoints (Supplementary Table S5). Sequence analysis revealed that human chromosome 19 harbors a 2-bp deletion with the insertion of a partial 1575-bp HBV DNA sequence (Fig. 7C). The insertion point of human chromosome 19 is the coding region of *KMT2B*, which is reported to be one of the target genes with a high frequency of integration in HCC patients (Saigo et al., 2008; Dong et al., 2015).

4. Discussion

Although HBV integration has been known for a long time (Yang and Summers, 1995, 1999), the molecular mechanisms, functions, and clinical implications remain largely unknown. An increasing number of studies suggests that HBV integration may serve as an oncogenic factor in HCC (Yeh et al., 2023). In addition, the impact of HBV integration on functional outcomes also need to be clarified. Due to the dual resources of serum HBsAg, integration might maintain the function of serum HBsAg in the absence of actively transcribed ccDNA during the HBeAg-negative stage. In individuals where integration DNA serves as the primary source of HBsAg transcripts, nucleos(t)ide analog therapy failed to reduce the HBsAg serum levels (Grudda et al., 2022; Hu and Huang, 2024). Therefore, accurate HBV integration detection methods are important for evaluating the antiviral efficacy in HBV-infected diseases, such as CHB, LC, and HCC. Quantification methods encompass PCR-based methods, such as inverse nested PCR (inv PCR) (Budzinska et al., 2018b), as well as high-throughput sequencing technologies (Zhao et al., 2016; Jang et al., 2021; van Buuren et al., 2022). Each method has its advantages or limitations (Budzinska et al., 2018a). Given their high sensitivity, cost-effectiveness, and minimal labor requirements, we chose probe-based capture and NGS strategies to detect HBV integration in clinical samples. Moreover, several software programs, such as VIcaller (Chen et al., 2019) and VirusFinder 2.0 (Wang et al., 2015), have also been developed to enhance the sensitivity of detecting viral insertion sites within host genomes. In this study, we developed a robust approach for HBV integration detection, focusing on the following improvement: (1) NGS sequences contain massive numbers of duplicate, repetitive sequences, potentially leading to assembly errors (Treangen and Salzberg, 2011). However, BBAP assembly has greater assembly efficiency and accuracy than other assembly methods (Lin et al., 2017). Therefore, we integrated the BBAP into our HBV integration detection strategy. (2) Multimapped alignments occurred during read mapping, due to the abundance of duplicated sequences in human genome. And the supporting read clip sites of one breakpoint may be skewed by sequencing error, assembly issues, mapping inaccuracies, or DNA repair processes.

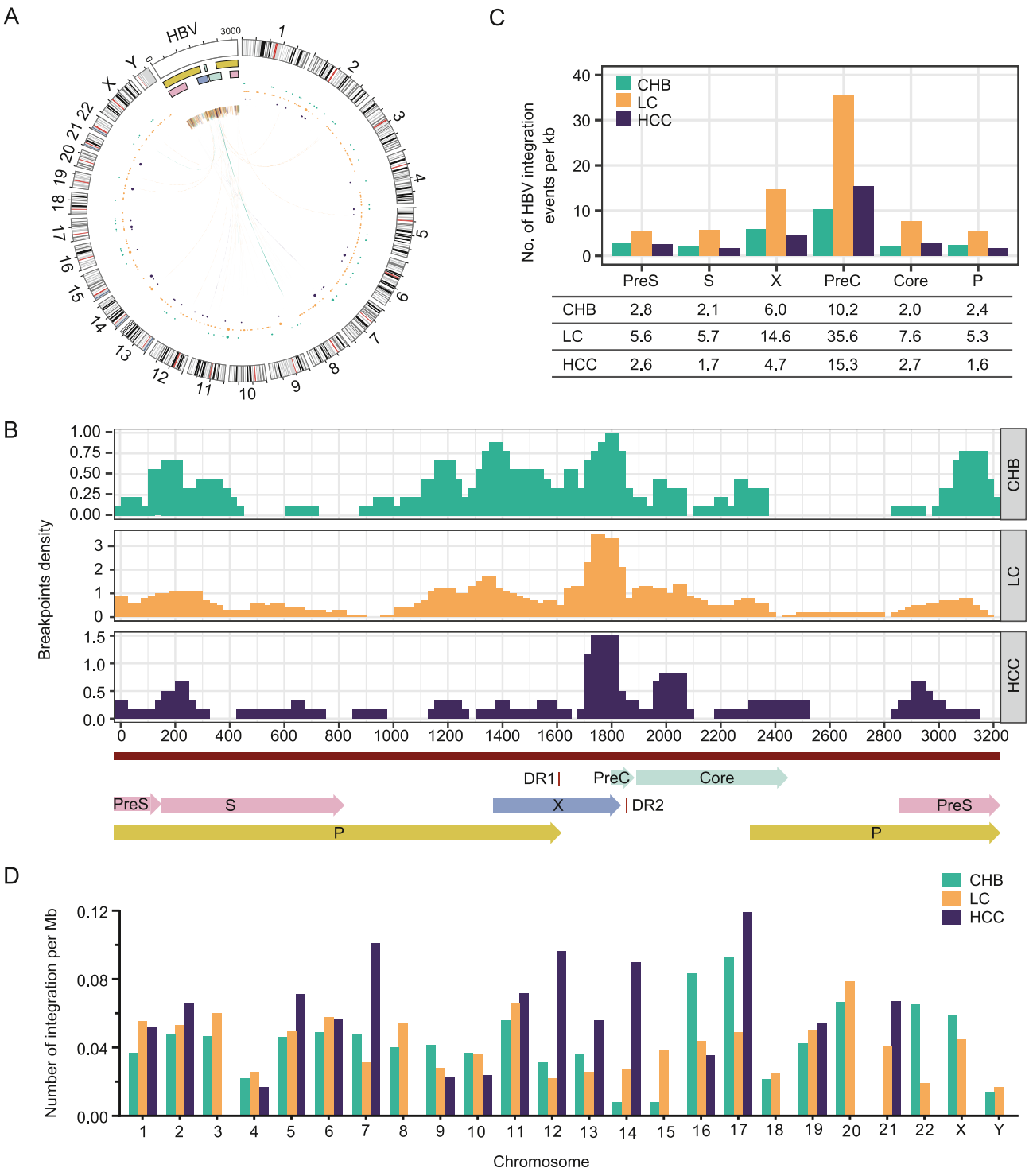

Fig. 5. Distribution characteristics of HBV integration identified in HBV-infected patients with different stages of liver disease. **A** The distribution of 333 integration breakpoints in the HBV genome and human chromosome is shown in circos plot. Each line represents an HBV integration breakpoint, connecting one side to the HBV genome and the other side to the human genome. Different colored dots represent different stages of the disease (CHB: green, LC: orange, and HCC: purple), and the size of the dots represents the frequency of the breakpoint. **B, C** Distribution of HBV DNA integration breakpoints on the HBV genome (**B**) and HBV-encoding genes (**C**). **D** The relative ratio of the distribution of HBV DNA integration events on human chromosomes.

Therefore, it's necessary to merge these breakpoints into a single integration. (3) During the process of library construction and sequencing, technical errors may occur due to the low abundance of cfDNA in the serum. To estimate background noise and assess accuracy, we constructed 30 artificial HBV-human chimeric sequences. Our data demonstrated that this improved strategy had good efficiency and accuracy and thus has potential for use in detecting HBV DNA integration both *in vitro and in vivo*.

Before applying this improved HBV integration detection strategy to clinical samples, we evaluated its efficacy in detecting HBV integration events in HBV-infected HepG2-NTCP cells. We detected an integration event at 1 dpi by this method (Fig. 2), and this integration breakpoint was

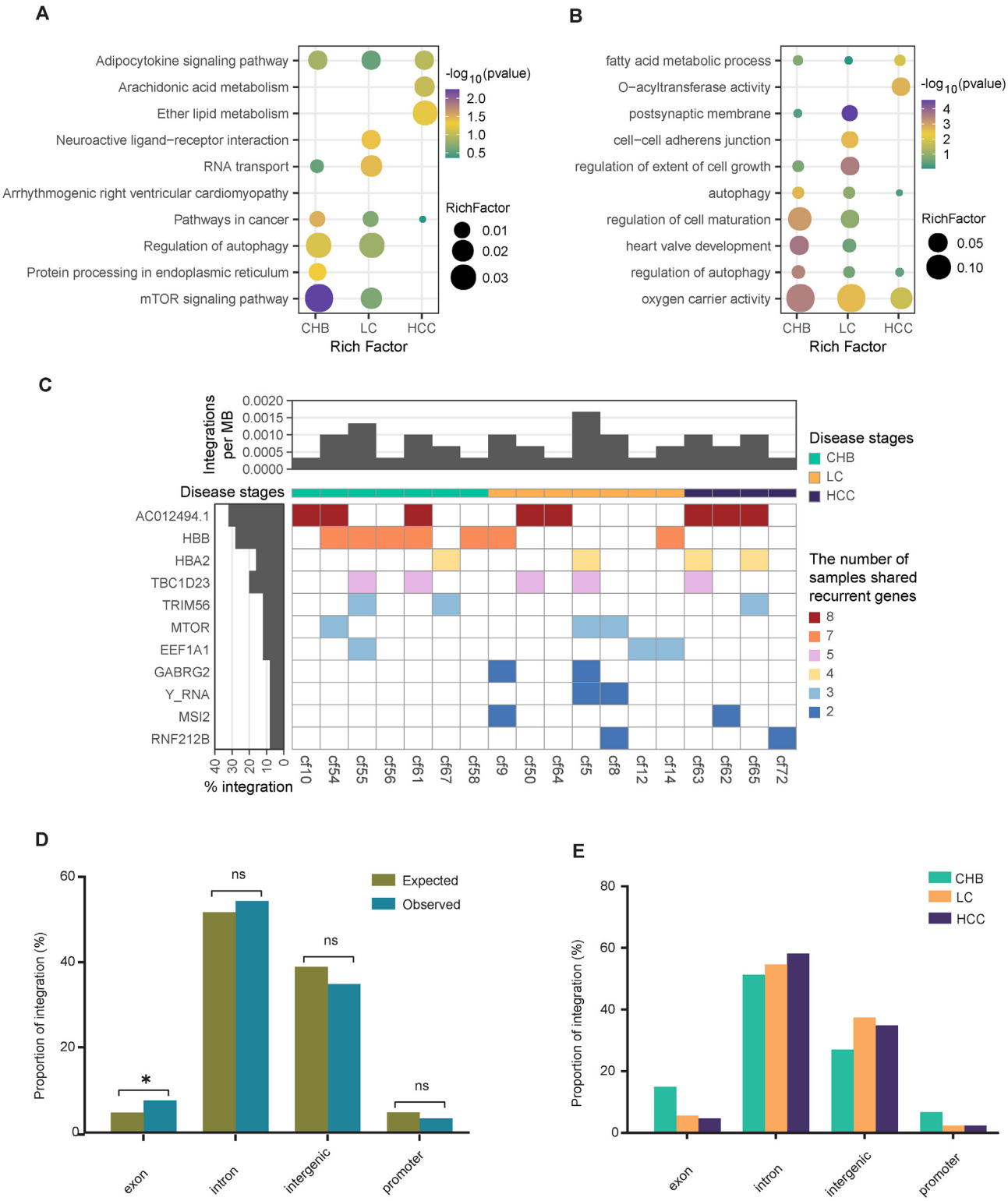

Fig. 6. Annotation analysis of HBV integration breakpoints identified in HBV-infected patients with different stages of liver disease. **A, B** KEGG (**A**) and Gene Ontology (**B**) analyses on genes around breakpoints or interrupted by integrations. **C** Relatively high-frequency genes on human chromosomes with HBV DNA integration breakpoints. The top rectangle represents the frequency of integration breakpoints occurring in patients at different stages, and the small rectangles of different colors in the bottom rectangle represent the frequency of gene occurrence. **D, E** The proportions of HBV DNA integration breakpoints in functional regions. **D** The proportions between expected (assuming uniform, random distribution) and observed (actual numbers) regions are compared. **E** The proportions of HBV DNA integration breakpoint detected in patients with different liver disease are compared. Statistical analysis was performed with chi-square tests. A *P*-value <0.05 was considered to indicate statistical significance. *, *P*-values <0.05; ns, not significant.

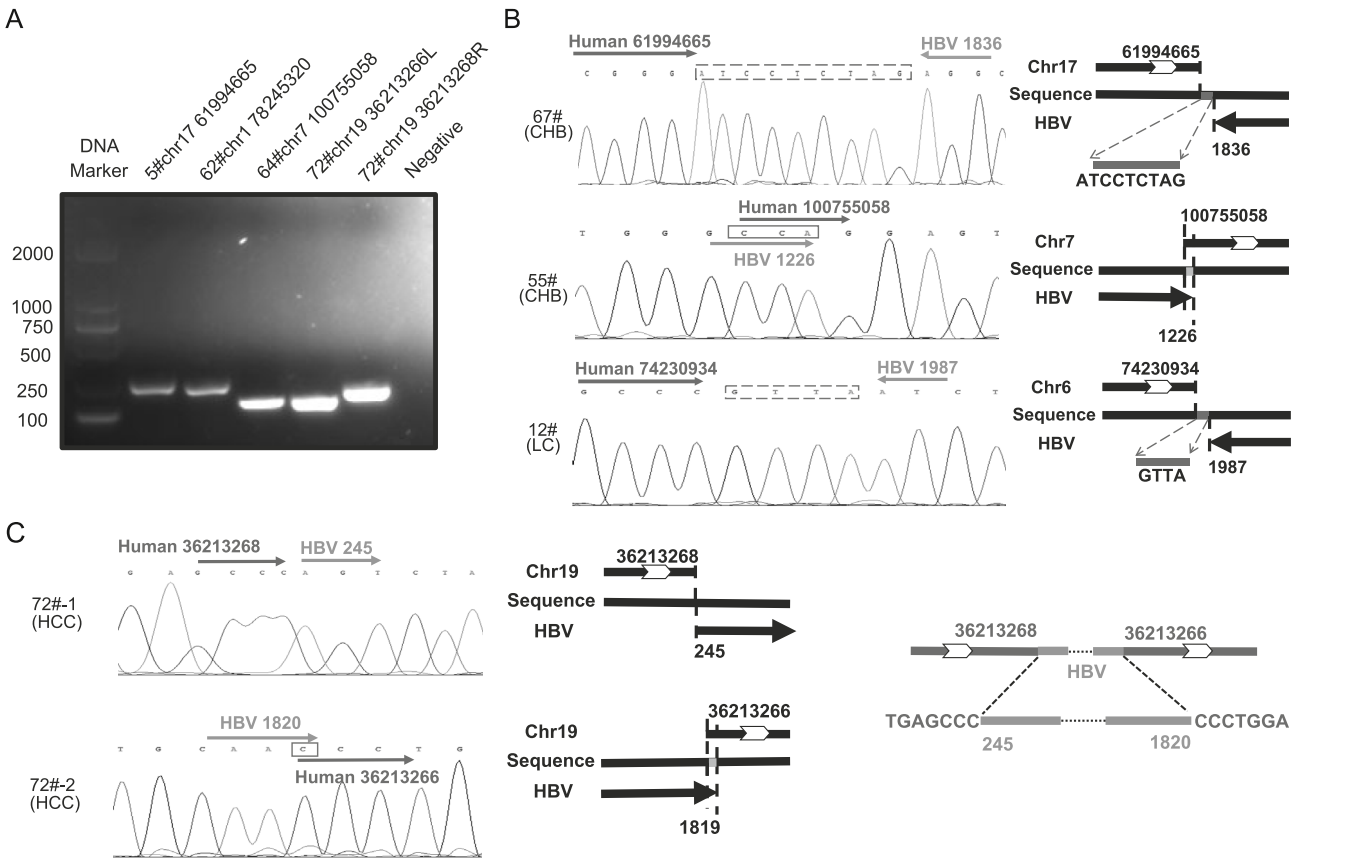

Fig. 7. Identification and validation of HBV DNA integration in cfDNA. **A** Agarose gel electrophoresis of chimeric DNA fragments from clinical samples. **B** Sequence alignment results and integration pattern of 12#, 55#, and 67# samples. **C** Sequence alignment results and integration pattern of 72# samples. The complete integration event in the 72# sample was inferred by the algorithm and confirmed by Sanger sequencing. The dashed line or solid line indicates nonhomologous or homologous sequences between the virus and host, respectively.

confirmed by a virus-host-specific PCR (Supplementary Fig. S3). This finding corroborated previous reports (Tu et al., 2018; Chauhan et al., 2019), indicating that HBV integration occurs shortly after viral infection. Interestingly, although the amount of viral-associated DNA increased after 3 dpi, the number of integration events was stable (Fig. 2), supporting the hypothesis that the dslDNA from the input virion serves as a substrate for HBV integration during early infection (Tu et al., 2018). However, the viral factors involved in this process is still unclear. Further research employing the HBV viral protein-deficient virions would be helpful to clarify this.

The successful detection of HBV integration *in vitro* prompted us to apply this strategy to clinical HBV-infected samples. Liver biopsy detection is an invasive treatment and may not be widely accepted by patients in clinical settings; thus, liquid biopsy for the detection of integrated HBV DNA presents a potential noninvasive strategy. Although Alu-PCR has revealed HBV integration in peripheral blood mononuclear cells (PBMCs) (Murakami et al., 2004), the HBV DNA patterns in PBMCs differ from those in the liver (Bouffard et al., 1990). Recent studies have demonstrated the detection and validation of HBV integration in plasma or urine cfDNA paired with HCC tissues, indicating that the cfDNA could be a promising liquid biomarker for the detection of HBV integration (Li et al., 2020; Lin et al., 2021). Other researches have reported the presence of detectable HBV-integrated breakpoints in plasma cfDNA samples from patients with LC or HCC (Zheng et al., 2021), as well as single-junction sites in patients with CHB (Chen W. et al., 2020). We hypothesize that HBV-integrated breakpoints may be detectable in plasma cfDNA samples from patients with different stages of liver disease. In this study, we enrolled 25 HBV infected individuals at CHB, LC, or HCC stage with HBeAg-negative, and tested our hypothesis using their plasma cfDNA samples. Our findings provide evidence that HBV-integrated breakpoints can be detected in the cfDNA of patients with CHB, LC or HCC (Fig. 4). Thus, we can further compare the characteristics of HBV integration among these different stages of liver disease.

Firstly, HBV integration breakpoints were predominantly enriched at nt1700–1850, regardless of liver disease status. This region is a feature of dslDNA ends (Mason et al., 2021), supporting the hypothesis that dslDNA serves as the primary substrate for HBV integration *in vivo*. Secondly, HBV DNA integration into the host genome involves the occurrence of a double-strand break, including NHEJ or MMEJ (Hino et al., 1989; Mladenov et al., 2016; Pannunzio et al., 2018). Sequence analysis of specific HBV-host PCR junctions indicated that both NHEJ and MMEJ were involved in HBV integration among the three different stages of liver disease (Fig. 7). Thirdly, although HBV integration seems to be random in the human genome in CHB, LC, or HCC patients, these viral integrations are preferentially located in the gene-coding region, and some of them may interfere with host gene function by forming chimeric proteins. For example, a partial 1575-bp HBV DNA sequence inserted at the exon region of *KTM2B* at site 72 in HCC (Fig. 7) may lead to abnormal expression and potentially influence cancer development.

This study had several limitations. First, the small sample size necessitates further testing of clinical samples to validate the efficacy of this improved HBV DNA detection strategy. Second, we only detected the plasma cfDNA of patients with CHB, LC, or HCC. Further studies incorporating paired liver tissue and urine cfDNA samples would greatly enhance our understanding of HBV integration patterns across different stages of liver diseases.

5. Conclusion

In this research, we developed an optimized HBV DNA detection strategy based on a DNA probe capture strategy and applied it to measure HBV integration events in plasma cfDNA derived from HBeAg-negative CHB, LC, or HCC patients. This strategy could be a promising noninvasive tool for evaluating the efficacy of antiviral therapy in HBV-infected individuals. Furthermore, we compared the profiles of HBV integration across CHB, LC, and HCC patients, providing valuable insights into the mechanism of HBV integration in *vivo*.

Data availability

The raw sequencing data of this study have been deposited into the Sequence Read Archive with accession number PRJNA1062444 and the Science Data Bank (https://doi.org/10.57760/sciencedb.07541). The scripts used to process and analyze the data will be available from the corresponding author upon reasonable request.

Ethics statement

The study was approved by the Ethics Committee of The First Affiliated Hospital of Chongqing Medical University (2023–043). Written informed consent was obtained from all participants.

Author contributions

Zerui Yang: methodology, investigation, formal analysis, writing and original draft preparation. Jingyan Zeng: methodology, investigation, formal analysis, and validation. Yueyue Chen: methodology, investigation, formal analysis. Mengchun Wang: investigation, formal analysis. Hongchun Luo: funding acquisition, investigation, formal analysis. Ailong Huang: funding acquisition, formal analysis, conceptualization. Haijun Deng: funding acquisition, formal analysis, validation, conceptualization, writing-reviewing and editing. Yuan Hu: funding acquisition, conceptualization, supervision, writing-reviewing and editing.

Conflict of interest

The authors declare no potential conflicts of interest.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (2018YFE0107500) and the CQMU Program for Youth Innovation in Future Medicine (W0160) to Yuan Hu; Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJQN202100422), Scientific Research Projects in Yuzhong District of Chongqing (20210116) to HaiJun Deng; the 111 Project (D20028) to AiLong Huang; and the Natural Science Foundation of Chongqing (cstc2021jcyjmsxmX0202) to Hongchun Luo.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virs.2024.06.003.

References

Bertoletti, A., Ferrari, C., 2016. Adaptive immunity in HBV infection. J. Hepatol. 64, S71–s83.
Bouffard, P., Lamelin, J.P., Zoulim, F., Pichoud, C., Trepo, C., 1990. Different forms of hepatitis B virus DNA and expression of HBV antigens in peripheral blood mononuclear cells in chronic hepatitis B. J. Med. Virol. 31, 312–317.
Budzinska, M.A., Shackel, N.A., Urban, S., Tu, T., 2018a. Cellular genomic sites of hepatitis B virus DNA integration. Genes 9, 365.
Budzinska, M.A., Shackel, N.A., Urban, S., Tu, T., 2018b. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. Emerg. Microb. Infect. 7, 142.
Cai, J., Chen, L., Zhang, Z., Zhang, X., Lu, X., Liu, W., Shi, G., Ge, Y., Gao, P., Yang, Y., Ke, A., Xiao, L., Dong, R., Zhu, Y., Yang, X., Wang, J., Zhu, T., Yang, D., Huang, X., Sui, C., Qiu, S., Shen, F., Sun, H., Zhou, W., Zhou, J., Nie, J., Zeng, C., Stroup, E.K., Zhang, X., Chiu, B.C., Lau, W.Y., He, C., Wang, H., Zhang, W., Fan, J., 2019. Genome-wide mapping of 5-hydroxymethylcytosines in circulating cell-free DNA as a non-invasive approach for early detection of hepatocellular carcinoma. Gut 68, 2195–2205.
Chauhan, R., Churchill, N.D., Mulrooney-Cousins, P.M., Michalak, T.I., 2017. Initial sites of hepadnavirus integration into host genome in human hepatocytes and in the woodchuck model of hepatitis B-associated hepatocellular carcinoma. Oncogenesis 6, e317.
Chauhan, R., Shimizu, Y., Watashi, K., Wakita, T., Fukasawa, M., Michalak, T.I., 2019. Retrotransposon elements among initial sites of hepatitis B virus integration into human genome in the HepG2-NTCP cell infection model. Cancer Genet 235–236, 39–56.
Chen, W., Zhang, K., Dong, P., Fanning, G., Tao, C., Zhang, H., Guo, S., Wang, Z., Hong, Y., Yang, X., Lai, S., Ding, H., Zhao, H., Zeng, C., Protzer, U., Zhang, D., 2020. Noninvasive chimeric DNA profiling identifies tumor-originated HBV integrants contributing to viral antigen expression in liver cancer. Hepatol Int 14, 326–337.
Chen, X., Kost, J., Sulovari, A., Wong, N., Liang, W.S., Cao, J., Li, D., 2019. A virome-wide clonal integration analysis platform for discovering cancer viral etiology. Genome Res. 29, 819–830.
Chen, Y., Shen, B., Zheng, X., Long, Q., Xia, J., Huang, Y., Cai, X., Wang, D., Chen, J., Tang, N., Huang, A., Hu, Y., 2020. DHX9 interacts with APOBEC3B and attenuates the anti-HBV effect of APOBEC3B. Emerg. Microb. Infect. 9, 366–377.
Dong, H., Zhang, L., Qian, Z., Zhu, X., Zhu, G., Chen, Y., Xie, X., Ye, Q., Zang, J., Ren, Z., Ji, Q., 2015. Identification of HBV-MLL4 integration and its molecular basis in Chinese hepatocellular carcinoma. PLoS One 10, e0123175.
Erken, R., Loukachov, V., van Dort, K., van den Hurk, A., Takkenberg, R.B., de Niet, A., Jansen, L., Willemse, S., Reesink, H., Kootstra, N., 2022. Quantified integrated hepatitis B virus is related to viral activity in patients with chronic hepatitis B. Hepatology 76, 196–206.
Gao, N., Guan, G., Xu, G., Wu, H., Xie, C., Mo, Z., Deng, H., Xiao, S., Deng, Z., Peng, L., Lu, F., Zhao, Q., Gao, Z., 2023. Integrated HBV DNA and cccDNA maintain transcriptional activity in intrahepatic HBsAg-positive patients with functional cure following PEG-IFN-based therapy. Aliment. Pharmacol. Ther. 58, 1086–1098.
Gong, S.S., Jensen, A.D., Chang, C.J., Rogler, C.E., 1999. Double-stranded linear duck hepatitis B virus (DHBV) stably integrates at a higher frequency than wild-type DHBV in LMH chicken hepatoma cells. J. Virol. 73, 1492–1502.
Grudda, T., Hwang, H.S., Taddese, M., Quinn, J., Sulkowski, M.S., Sterling, R.K., Balagopal, A., Thio, C.L., 2022. Integrated hepatitis B virus DNA maintains surface antigen production during antiviral treatment. J. Clin. Invest. 132, e161818.
Hino, O., Ohtake, K., Rogler, C.E., 1989. Features of two hepatitis B virus (HBV) DNA integrations suggest mechanisms of HBV integration. J. Virol. 63, 2638–2643.
Hu, J.L., Huang, A.L., 2024. Classifying hepatitis B therapies with insights from covalently closed circular DNA dynamics. Virol. Sin. 39, 9–23.
Jang, J.W., Kim, J.S., Kim, H.S., Tak, K.Y., Nam, H., Sung, P.S., Bae, S.H., Choi, J.Y., Yoon, S.K., Roberts, L.R., 2021. Persistence of intrahepatic hepatitis B virus DNA integration in patients developing hepatocellular carcinoma after hepatitis B surface antigen seroclearance. Clin. Mol. Hepatol. 27, 207–218.
Li, C.L., Li, C.Y., Lin, Y.Y., Ho, M.C., Chen, D.S., Chen, P.J., Yeh, S.H., 2019. Androgen receptor enhances hepatic telomerase reverse transcriptase gene transcription after hepatitis B virus integration or point mutation in promoter region. Hepatology 69, 498–512.
Li, C.L., Ho, M.C., Lin, Y.Y., Tzeng, S.T., Chen, Y.J., Pai, H.Y., Wang, Y.C., Chen, C.L., Lee, Y.H., Chen, D.S., Yeh, S.H., Chen, P.J., 2020. Cell-free virus-host chimera DNA from hepatitis B virus integration sites as a circulating biomarker of hepatocellular cancer. Hepatology 72, 2063–2076.
Li, W., Zeng, X., Lee, N.P., Liu, X., Chen, S., Guo, B., Yi, S., Zhuang, X., Chen, F., Wang, G., Poon, R.T., Fan, S.T., Mao, M., Li, Y., Li, S., Wang, J., Jianwang, Xu X., Jiang, H., Zhang, X., 2013. HIVID: an efficient method to detect HBV integration using low coverage sequencing. Genomics 102, 338–344.
Lin, S.Y., Su, Y.P., Trauger, E.R., Song, B.P., Thompson, E.G.C., Hoffman, M.C., Chang, T.T., Lin, Y.J., Kao, Y.L., Cui, Y., Hann, H.W., Park, G., Shieh, F.S., Song, W., Su, Y.H., 2021. Detection of hepatitis B virus-host junction sequences in urine of infected patients. Hepatol Commun 5, 1649–1659.
Lin, Y.Y., Hsieh, C.H., Chen, J.H., Lu, X., Kao, J.H., Chen, P.J., Chen, D.S., Wang, H.Y., 2017. De novo assembly of highly polymorphic metagenomic data using in situ generated reference sequences and a novel blast-based assembly pipeline. BMC Bioinf. 18, 223.
Luo, H., Wei, W., Ye, Z., Zheng, J., Xu, R.H., 2021. Liquid biopsy of methylation biomarkers in cell-free DNA. Trends Mol. Med. 27, 482–500.
Luo, X., Huang, Y., Chen, Y., Tu, Z., Hu, J., Tavis, J.E., Huang, A., Hu, Y., 2016. Association of hepatitis B virus covalently closed circular DNA and human APOBEC3B in hepatitis B virus-related hepatocellular carcinoma. PLoS One 11, e0157708.
Marzese, D.M., Hirose, H., Hoon, D.S., 2013. Diagnostic and prognostic value of circulating tumor-related DNA in cancer patients. Expert Rev. Mol. Diagn 13, 827–844.
Mason, W.S., Jilbert, A.R., Litwin, S., 2021. Hepatitis B virus DNA integration and clonal expansion of hepatocytes in the chronically infected liver. Viruses 13, 210.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
Meier, M.A., Calabrese, D., Suslov, A., Terracciano, L.M., Heim, M.H., Wieland, S., 2021. Ubiquitous expression of HBsAg from integrated HBV DNA in patients with low viral load. J. Hepatol. 75, 840–847.

Mladenov, E., Magin, S., Soni, A., Iliakis, G., 2016. DNA double-strand-break repair in higher eukaryotes and its role in genomic instability and cancer: cell cycle and proliferation-dependent regulation. Semin. Cancer Biol. 37–38, 51–64.
Murakami, Y., Minami, M., Daimon, Y., Okanoue, T., 2004. Hepatitis B virus DNA in liver, serum, and peripheral blood mononuclear cells after the clearance of serum hepatitis B virus surface antigen. J. Med. Virol. 72, 203–214.
Nguyen, M.H., Wong, G., Gane, E., Kao, J.H., Dusheiko, G., 2020. Hepatitis B virus: advances in prevention, diagnosis, and therapy. Clin. Microbiol. Rev. 33.
Pannunzio, N.R., Watanabe, G., Lieber, M.R., 2018. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J. Biol. Chem. 293, 10512–10523.
Pollicino, T., Caminiti, G., 2021. HBV-integration studies in the clinic: role in the natural history of infection. Viruses 13, 368.
Revill, P., Testoni, B., Locarnini, S., Zoulim, F., 2016. Global strategies are required to cure and eliminate HBV infection. Nat. Rev. Gastroenterol. Hepatol. 13, 239–248.
Saigo, K., Yoshida, K., Ikeda, R., Sakamoto, Y., Murakami, Y., Urashima, T., Asano, T., Kenmochi, T., Inoue, I., 2008. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. Hum. Mutat. 29, 703–708.
Schlüter, V., Meyer, M., Hofschneider, P.H., Koshy, R., Caselmann, W.H., 1994. Integrated hepatitis B virus x and 3' truncated preS/S sequences derived from human hepatomas encode functionally active transactivators. Oncogene 9, 3335–3344.
Skidmore, Z.L., Wagner, A.H., Lesurf, R., Campbell, K.M., Kunisaki, J., Griffith, O.L., Griffith, M., 2016. GenVisR: genomic visualizations in R. Bioinformatics 32, 3012–3014.
Song, A., Lin, X., Chen, X., 2021. Functional cure for chronic hepatitis B: accessibility, durability, and prognosis. Virol. J. 18, 114.
Sung, W.K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N.P., Lee, W.H., Ariyaratne, P.N., Tennakoon, C., Mulawadi, F.H., Wong, K.F., Liu, A.M., Poon, R.T., Fan, S.T., Chan, K.L., Gong, Z., Hu, Y., Lin, Z., Wang, G., Zhang, Q., Barber, T.D., Chou, W.C., Aggarwal, A., Hao, K., Zhou, W., Zhang, C., Hardwick, J., Buser, C., Xu, J., Kan, Z., Dai, H., Mao, M., Reinhard, C., Wang, J., Luk, J.M., 2012. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat. Genet. 44, 765–769.
Treangen, T.J., Salzberg, S.L., 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. 13, 36–46.
Tu, T., Budzinska, M.A., Shackel, N.A., Urban, S., 2017. HBV DNA integration: molecular mechanisms and clinical implications. Viruses 9, 75.
Tu, T., Budzinska, M.A., Vondran, F.W.R., Shackel, N.A., Urban, S., 2018. 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. 92, e02007-17.
van Buuren, N., Ramirez, R., Soulette, C., Suri, V., Han, D., May, L., Turner, S., Parvangada, P.C., Martin, R., Chan, H.L.Y., Marcellin, P., Buti, M., Bui, N., Bhardwaj, N., Gaggar, A., Li, L., Mo, H., Feierbach, B., 2022. Targeted long-read sequencing reveals clonally expanded HBV-associated chromosomal translocations in patients with chronic hepatitis B. JHEP Rep 4, 100449.
Wang, Q., Jia, P., Zhao, Z., 2015. VERSE: a novel approach to detect virus integration in host genomes through reference genome customization. Genome Med. 7, 2.
Wooddell, C.I., Yuen, M.F., Chan, H.L., Gish, R.G., Locarnini, S.A., Chavez, D., Ferrari, C., Given, B.D., Hamilton, J., Kanner, S.B., Lai, C.L., Lau, J.Y.N., Schluep, T., Xu, Z., Lanford, R.E., Lewis, D.L., 2017. RNAi-based treatment of chronically infected patients and chimpanzees reveals that integrated hepatitis B virus DNA is a source of HBsAg. Sci. Transl. Med. 9, eaan0241.
Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2, 100141.
Yang, W., Summers, J., 1995. Illegitimate replication of linear hepadnavirus DNA through nonhomologous recombination. J. Virol. 69, 4029–4036.
Yang, W., Summers, J., 1999. Integration of hepadnavirus DNA in infected liver: evidence for a linear precursor. J. Virol. 73, 9710–9717.
Yeh, S.H., Li, C.L., Lin, Y.Y., Ho, M.C., Wang, Y.C., Tseng, S.T., Chen, P.J., 2023. Hepatitis B virus DNA integration drives carcinogenesis and provides a new biomarker for HBV-related HCC. Cell Mol Gastroenterol Hepatol 15, 921–929.
Yoshiji, H., Nagoshi, S., Akahane, T., Asaoka, Y., Ueno, Y., Ogawa, K., Kawaguchi, T., Kurosaki, M., Sakaida, I., Shimizu, M., Taniai, M., Terai, S., Nishikawa, H., Hiasa, Y., Hidaka, H., Miwa, H., Chayama, K., Enomoto, N., Shimosegawa, T., Takehara, T., Koike, K., 2021. Evidence-based clinical practice guidelines for liver cirrhosis 2020. J. Gastroenterol. 56, 593–619.
You, H., Wang, F., Li, T., Xu, X., Sun, Y., Nan, Y., Wang, G., Hou, J., Duan, Z., Wei, L., Jia, J., Zhuang, H., Chinese Society of Hepatology CMA, Chinese Society of Infectious Diseases CMA, 2023. Guidelines for the prevention and treatment of chronic hepatitis B (version 2022). J Clin Transl Hepatol 11, 1425–1442.
Zhang, M., Zhang, H., Cheng, X., Wang, X., Xu, H., Gao, X., Wu, R., Zhang, D., Xia, Y., Niu, J., 2022. Liver biopsy of chronic hepatitis B patients indicates HBV integration profile may complicate the endpoint and effect of entecavir treatment. Antivir. Res. 204, 105363.
Zhao, L.H., Liu, X., Yan, H.X., Li, W.Y., Zeng, X., Yang, Y., Zhao, J., Liu, S.P., Zhuang, X.H., Lin, C., Qin, C.J., Zhao, Y., Pan, Z.Y., Huang, G., Liu, H., Zhang, J., Wang, R.Y., Yang, Y., Wen, W., Lv, G.S., Zhang, H.L., Wu, H., Huang, S., Wang, M.D., Tang, L., Cao, H.Z., Wang, L., Lee, T.L., Jiang, H., Tan, Y.X., Yuan, S.X., Hou, G.J., Tao, Q.F., Xu, Q.G., Zhang, X.Q., Wu, M.C., Xu, X., Wang, J., Yang, H.M., Zhou, W.P., Wang, H.Y., 2016. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. Nat. Commun. 7, 12992.
Zheng, B., Liu, X.L., Fan, R., Bai, J., Wen, H., Du, L.T., Jiang, G.Q., Wang, C.Y., Fan, X.T., Ye, Y.N., Qian, Y.S., Wang, Y.C., Liu, G.J., Deng, G.H., Shen, F., Hu, H.P., Wang, H., Zhang, Q.Z., Ru, L.L., Zhang, J., Gao, Y.H., Xia, J., Yan, H.D., Liang, M.F., Yu, Y.L., Sun, F.M., Gao, Y.J., Sun, J., Zhong, C.X., Wang, Y., Kong, F., Chen, J.M., Zheng, D., Yang, Y., Wang, C.X., Wu, L., Hou, J.L., Liu, J.F., Wang, H.Y., Chen, L., 2021. The landscape of cell-free HBV integrations and mutations in cirrhosis and hepatocellular carcinoma patients. Clin. Cancer Res. 27, 3772–3783.
Zhou, J., Sun, H., Wang, Z., Cong, W., Zeng, M., Zhou, W., Bie, P., Liu, L., Wen, T., Kuang, M., Han, G., Yan, Z., Wang, M., Liu, R., Lu, L., Ren, Z., Zeng, Z., Liang, P., Liang, C., Chen, M., Yan, F., Wang, W., Hou, J., Ji, Y., Yun, J., Bai, X., Cai, D., Chen, W., Chen, Y., Cheng, W., Cheng, S., Dai, C., Guo, W., Guo, Y., Hua, B., Huang, X., Jia, W., Li, Q., Li, T., Li, X., Li, Y., Li, Y., Liang, J., Ling, C., Liu, T., Liu, X., Lu, S., Lv, G., Mao, Y., Meng, Z., Peng, T., Ren, W., Shi, H., Shi, G., Shi, M., Song, T., Tao, K., Wang, J., Wang, K., Wang, L., Wang, W., Wang, X., Wang, Z., Xiang, B., Xing, B., Xu, J., Yang, J., Yang, J., Yang, Y., Yang, Y., Ye, S., Yin, Z., Zeng, Y., Zhang, B., Zhang, B., Zhang, L., Zhang, S., Zhang, T., Zhang, Y., Zhao, M., Zhao, Y., Zheng, H., Zhou, L., Zhu, J., Zhu, K., Liu, R., Shi, Y., Xiao, Y., Zhang, L., Yang, C., Wu, Z., Dai, Z., Chen, M., Cai, J., Wang, W., Cai, X., Li, Q., Shen, F., Qin, S., Teng, G., Dong, J., Fan, J., 2023. Guidelines for the diagnosis and treatment of primary liver cancer (2022 edition). Liver Cancer 12, 405–444.