HBV Integration Induces Complex Interactions between Host and Viral Genomic Functions at the Insertion Site

Dake Zhang1*, Ke Zhang2,3, Urlike Protzer3,4 and Changqing Zeng5,6

1Beijing Advanced Innovation Centre for Biomedical Engineering, Key Laboratory for Biomechanics and Mechanobiology of Ministry of Education, School of Biological Science and Medical Engineering, Beihang University, Beijing, China; 2SCG Cell Therapy Pte. Ltd, Singapore; 3Institute of Virology, Technical University of Munich/Helmholtz Zentrum München, Munich, Germany; 4German Center for Infection Research (DZIF), Munich, Germany; 5Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; 6University of Chinese Academy of Sciences, Beijing, China

Abstract

Hepatitis B virus (HBV), one of the well-known DNA oncogenic viruses, is the leading cause of hepatocellular carcinoma (HCC). In infected hepatocytes, HBV DNA can be integrated into the host genome through an insertional mutagenesis process inducing tumorigenesis. Dissection of the genomic features surrounding integration sites will deepen our understanding of mechanisms underlying integration. Moreover, the quantity and biological activity of integration sites may reflect the DNA damage within affected cells or the potential survival benefits they may confer. The well-known human genomic features include repeat elements, particular regions (such as telomeres), and frequently interrupted genes (e.g., telomerase reverse transcriptase [i.e., TERT], lysine methyltransferase 2B [i.e., KMT2B], cyclin E1 [CCNE1], and cyclin A2 [CCNA2]). Consequently, distinct genomic features within diverse integrations differentiate their biological functions. Meanwhile, accumulating evidence has shown that viral proteins produced by integrants may cause cell damage even after the suppression of HBV replication. The integration-derived gene products can also serve as tumor markers, promoting the development of novel therapeutic strategies for HCC. Viral integrants can be single copy or multiple copies of different fragments with complicated rearrangement, which warrants elucidation of the whole viral integrant arrangement in future studies. All of these considerations underlie an urgent need to develop novel methodology and technology for sequence characterization and function evaluation of integration events in chronic hepatitis B-associated disease progression by monitoring both host genomic features and viral integrants. This endeavor may also serve as a promising solution for evaluating the risk of tumorigenesis and as a companion diagnostic for designing therapeutic strategies targeting integration-related disease complications.

Citation of this article: Zhang D, Zhang K, Protzer U, Zeng C. HBV integration induces complex interactions between host and viral genomic functions at the insertion site. J Clin Transl Hepatol 2021;9(3):399–408. doi: 10.14218/JCTH.2021.00062.

Introduction

There are still more than 250 million people infected with chronic hepatitis B virus (HBV) worldwide. Hepatocellular carcinoma (HCC) imposes a heavy socioeconomic public health burden. Among patients with cirrhosis, HBV infection has a reported annual incidence of 3%7. In 1980, shortly after the discovery of HBV, its integration was reported into HCC tissue cells and hepatoma cell lines.2,3 There has been a long debate on the oncogenic roles of HBV integrations.4–6 It is considered to be involved in the oncogenic process due to its intrinsic insertional mutagenesis (Fig. 1). Although current treatment using nucleoside or nucleotide analogs can efficiently suppress HBV replication, it is still not able to eliminate the virus from infected hepatocytes, mainly due to the persistent existence of covalently closed circular DNA (commonly referred to as cccDNA).7

Meanwhile, double-stranded linear DNA (dsDNA) is the alleged dominant substrate that gets incorporated into the host genome via repair of DNA double-stranded breaks (DSBs).8 Indeed, HBV integrations tend to locate at either sites of host genomic instability,5,10 or sites of cellular DNA damage.11 Meanwhile, the HBV genome does not encode either integrase or other proteins recognizing specific genomic regions for integration.9,12,13 and it is unlikely that HBV could deliberately “target” certain genes or regions. By far, only repeat elements or short homologous sequences, mediating the DNA repairing via non-homologous end joining,

Keywords: Double-strand break; Chimeric reads; Junction reads; Fusion transcript; Virus cellular junction; Virus cell junction; Virus host junction.

Abbreviations: cccDNA, covalently closed circular DNA; CCN, cyclin; cDNA, cell-free DNA dsDNA, double-stranded linear DNA; DSB, DNA double-stranded breaks; FNI, fibronectin 1; HBV, hepatitis B virus; HBx, hepatitis B X; HCC, hepatocellular carcinoma; hTERT, human telomerase reverse transcriptase; KMT2B, lysine methyltransferase 2B; LINE1, long interspersed nuclear elements 1; mDNA, mitochondrial DNA; NGS, next-generation of sequencing; NTCP, Na+-taurocholate polypeptide cotransporter; NUMT, nuclear copies of mDNA; pgRNA, pregenomic RNA; SINE, short interspersed nuclear elements; TPS3, tumor protein p53; TSS, transcriptional start site.

*Correspondence to: Dake Zhang, Beijing Advanced Innovation Centre for Biomedical Engineering, Key Laboratory for Biomechanics and Mechanobiology of Ministry of Education, School of Biological Science and Medical Engineering, Beihang University, Beijing 100083, China. ORCID: https://orcid.org/0000-0001-9508-8209. E-mail: dakezhang@buaa.edu.cn

Publication details: Journal of Clinical and Translational Hepatology 2021 vol. 9(3) 399–408. doi: 10.14218/JCTH.2021.00062.
are the main genomic feature of integration sites. 14 It has been known for a long time that the ends of an integrant most likely terminate at a 11 bp repeat region (5′-TTCAC-CTCTGC-3′), namely the cohesive-end regions termed DR1 and DR2.15 About 37–40% of the reported viral breakpoints are mapped within the DR2-DR1 region, 16–18 where transcription and replication of the genome are initiated. During HBV replication, the reverse transcription from pregenomic RNA (commonly referred to as pgRNA) to cccDNA or dslDNA uses an 18 nucleotide RNA primer, which is the hydrolysate of pgRNA, to initiate the synthesis of the positive-sense DNA strand.12 In about 90% of nucleocapsids, the RNA primer translocates to the DR2, leading to the synthesis of relaxed circular DNA; while in the remaining 10%, it binds to the DR1, priming the synthesis of dslDNA as the main source of viral DNA incorporated into the host genome.12 The HCC risk due to viral integrations remains as long as viral replication continues in liver cells. However, the detailed processes underlying viral integration still need clarification by investigators attempting to identify novel options to cure HBV infection. Recently, much attention has been dedicated to reviewing biological processes and consequences related to viral integrations in the background of the HBV life cycle.9,12 The development of high-throughput sequencing solutions significantly facilitates resolving thousands of viral integration breakpoints.16,18–22 Our objective here is to summarize the features of host genomic sequences surrounding integration sites and viral fragments as integrants. These features of distinct integrations may be crucial to unraveling their contribution to liver disease progression. Pinpointing their involvement may aid efforts to identify novel diagnostic markers of disease progression and targets for improved therapeutic management of HBV infection.

HBV integration detection: Hybridization, cloning, amplification and next-generation sequencing (NGS) solutions

The short HBV DNA fragments are difficult to identify after being inserted into a large human genome (3Gb) in each affected cell. Meanwhile, these cells may also harbor limited integrations. In relative homogeneous populations of cell lines carrying HBV integrations, there are no more than 10 dominant integrations. For instance, the HepG2.2.15 cell line may contain at most five integrations,21 and the PLC/PRF/5 cell line had nine integration regions reported.23 Among the cell populations of the infected liver, it is estimated that there is about 1 integration per 10^3–10^4 cells according to observations in an in vitro infection model.8 Therefore, it
requires detection methods that have adequate sensitivity to identify such limited events in bulk tissue samples.  

Early in the 1980s, the hybridization method was first applied to prove the existence of HBV integrations in cell lines and tumor tissues.\(^2,24\) In Southern blot hybridization, digested DNA after a restriction enzyme treatment hybridized with \(^{32P}\)-labeled HBV DNA probes, and researchers took the autoradiograph of the bands of digested DNA in the separation gel to identify the fragment harboring HBV DNA. Target fragments could be ligated into plasmid as integrant clones, which can be further fragmented and radiolabeled as probes for \textit{in situ} hybridization. Prior to Sanger sequencing of integrant clones,\(^25\) \textit{in situ} hybridization was applied to see the cytogenetic locus of these probes binding to metaphase chromosomes and thereby determine the chromosome position of viral integrants.\(^25\) The hybridization strategy and clone sequencing without an NGS platform are inefficient and insensitive to profile integrations.

To improve the methodology of HBV integration profiling, Minami \textit{et al}.\(^27\) (1995) developed the Alu PCR strategy, which is more practical for integration screening. Briefly, Alu elements, which have over one million copies dispersed throughout the human genome,\(^28\) separate the entire genome into relatively small regions, the length of which is suitable for performing PCR amplification; the primer pair, one of which is Alu-specific and the other binds HBV sequencing (Fig. 1), may successfully amplify the virus-host chimeric regions. With this approach, Devrim \textit{et al}.\(^29\) identified 21 viral integrations from 18 patients in one study. Likewise, Mason \textit{et al}.\(^30–33\) first cleaved total liver cell DNA into fragments by NcoI, which cuts HBV DNA at nucleotide 1374, and then ligated these fragments into circles for nested PCR of virus-cellular junctions. This sensitive inverse-PCR assay is a quantitative method for integration detection. Nevertheless, since not all integrations are next to an Alu element or the cleavage site of NcoI, some may be omitted in subsequent analysis.

NGS solutions, either by direct sequencing or target sequencing after viral DNA enrichment (Fig. 1), can presumably achieve unbiased parallel analysis of tens to hundreds of samples. Extracted tissue DNA is randomly fragmented for sequencing library construction, the bioinformatic analysis aims to identify the so-called “junction read” or “chimeric read” in sequencing data, which is composed of both HBV and human DNA together and originates from the boundaries of integration events. Considering the relatively low frequencies of integrations, directly sequencing the whole nuclear genome for bulk tissues requires adequate sequencing coverage/depth to capture these chimeric fragments. Jiang \textit{et al}.\(^36\) adopted deep whole-genome sequencing (80X, 240G per sample; and 240X, 720G per sample) and identified 255 integrations in paired tumor and adjacent liver tissues from three HBV-positive HCC patients. The high cost and data analysis requirement will limit its application at the population scale. Target DNA enrichment has proven to be highly effective in exome sequencing of the human genome. It inspires the usage of viral DNA probes for HBV DNA enrichment prior to NGS analysis, which significantly reduces the sequencing volume to less than 2G per sample.\(^16,21\) With the increased throughput, Zhao \textit{et al}.\(^16\) obtained 4,225 integration breakpoints (host-virus boundaries at the integration site) from 426 patients to characterize a viral integration pattern in one study, followed by other teams.\(^17,34–37\) Most NGS studies only took the single breakpoint as the indicator for one integration event. But one integration has two breakpoints (Fig. 1), and they may overestimate the number of integrations in each sample.\(^21\) Fortunately, different integration events are likely to be far apart from each other in the cellular genome of hepatocytes. Therefore, we developed a strategy to pair adjacent breakpoints in the human genome for the same integration event.\(^21\) After the breakpoints in the human genome have been successfully paired, the corresponding viral breakpoints can then be paired. The boundaries of integration events are indicated by the sharp and consistent positions of the mapped reads, the particular distribution patterns of which have shown diverse modes of integrations (Fig. 2A). This pairing strategy can be used to predict the integrants and show their protein-coding ability or harbored regulatory elements. It can also illustrate insertion orientation and HBV fragments within the integration site. Besides, this strategy successfully uncovered the possibility of multiple fragments with different orientations in the same site (Fig. 2B), which have been validated in long-read sequencing studies.\(^38\) The orientation of inserted fragments will determine their upstream and downstream directions and hence influence their regulatory roles for involved genes in the aforementioned biological effects of integrations. Therefore, further complicating the regulative activity of the integration.

Viral integrations have been reported to locate at the boundaries of chromosome rearrangements, such as large deletions, translocations, and inversions (Fig. 2C), which make the start and end of the same integrant distant from each other even in different chromosomes respectively.\(^39,40\) Only when the structure variations of the host genome are dissected, can they have both breakpoints identified. Therefore, short-read sequencing will fail to predict the entire integrant for integration sites with complicated local sequence rearrangement. To meet the requirement for the characterization of entire integrants, long-read sequencing, with the longest single read over 2 Mbp, is able to directly read through the entire integration region.\(^23\) It can not only reveal integrations affected by large structure variations within the host genome but also characterize the organization of multiple copies of HBV fragments within the same integration site.\(^23\)

### Recurrent genes interrupted by integration events

From the 1980s to the early 1990s, a few studies began to report HBV integration that affected the expression levels of a series of genes related to tumor development. They include \textit{HSTL1},\(^41\) tumor protein p53 (i.e. TP53),\(^25,42\) cyclin A2 (\textit{CCNA2}),\(^43\) \textit{myc} family oncogenes,\(^44,45\) retinoblastoma beta or erb-B-like genes,\(^46,47\) and mevalonate kinase gene.\(^48\) With the accumulation of identified integration events, the frequencies of common recurrent genes affected by viral integrations become well characterized. The human telomerase reverse transcriptase (i.e. \textit{hTERT}), which prevents telomere shortening after cell division, is not only the first reported recurrent gene affected by HBV integrations but it is also the most common one in HCC samples, ranging from nearly 20\% (i.e. 17\%, 14/82, Fujimoto \textit{et al}.;\(^49\) 24\%, 101/426, Sun \textit{et al}.;\(^16\) 24\%, 101/426, Zhao \textit{et al}.;\(^16\) 27\%, 48/177, Péneau \textit{et al}.\(^50\)) to over 35\% (36\%, 34/95, Sze \textit{et al}.\(^37\)). The second most common recurrent gene was lysine methyltransferase 2B (\textit{KMT2B}, also known as \textit{MLL4} or \textit{MLL2}), encoding histone methyltransferases, with a frequency of about 10\% (2\%, Péneau \textit{et al}.;\(^50\) 7\%, Zhao \textit{et al}.;\(^16\) 7\%, Fujimoto \textit{et al}.;\(^49\) 12\%, Sun \textit{et al}.;\(^16\) 12\%, Sze \textit{et al}.\(^37\)). They were followed by CCNE1 (1\%, Fujimoto \textit{et al}.;\(^49\) 1.6\%, Zhao \textit{et al}.;\(^16\) 2\%, Péneau \textit{et al}.;\(^50\) 3\%, Sze \textit{et al}.;\(^37\) 5\% Sun \textit{et al}.\(^18\)) and \textit{CCNA2} (1\%, Fujimoto \textit{et al}.;\(^49\) 1.8\%, Zhao \textit{et al}.;\(^16\) 2\%, Sze \textit{et al}.\(^37\)).

Whereas, integrations seem to account for a relatively small proportion of chromic changes of total HCC.\(^54\) For instance, about 70\% of all HCC carry \textit{TERT} mutations, and HBV integrations contribute to about 7\% of them.\(^52\) In other words, 5\% of HCC are due to \textit{TERT} integration;
moreover, 58% of HCC have genomic changes affecting cell cycle control, and HBV integrations account for about 8% of those interrupting the CCNE1.\textsuperscript{52} Considering HBV infection contributing to 50% of HCC cases, the contribution of integrations should be doubled in HBV-related HCC.\textsuperscript{53}

HBV integration in gene promoter region: cis-activation effect

Integrations can influence the transcription of nearby genes by changing the promoter activity (Fig. 1; Gene Annotation), contributing to phenotype changes of affected cells. The integration site can be either only hundreds of base pairs (e.g., 257 bp upstream of \textit{hTERT})\textsuperscript{54} or over 100 kb (e.g., 2,582 kb upstream of \textit{CCNE1})\textsuperscript{55} from the genes in the vicinity along the human genome. In the first large-scale screening for HBV integrations in HCC, Sun \textit{et al.}\textsuperscript{18} explored integrations in the promoter region which were from 0 to −5 kb relative to the transcriptional start site (commonly referred to as TSS), while Fujimoto \textit{et al.}\textsuperscript{45} enlarged this region up to 10 kb upstream, which has become widely adopted. Nevertheless, the frequency and influence of integration in the promoter may be underestimated since some are located beyond the 10 kb region,\textsuperscript{56} and Ding \textit{et al.}\textsuperscript{35} further extended the annotation region to 150 kb. However, analysis of the effects of long-range integration on gene expression regulation may be challenging and need further extensive validation.

Meanwhile, both the insertion orientation and the integration-TSS distance may influence their effect on the promoter activity of viral integrations. Telomerase expression was absent in mature hepatocytes, but 90% of the HCC\textsuperscript{57} had \textit{TERT}-related integrations, of which 80% are located in the promoter region.\textsuperscript{18} In 2001, Horikawa \textit{et al.}\textsuperscript{58} first described the HBV integrations in the \textit{cis}-activated \textit{TERT} promoter in an orientation-independent manner. Recently, Sze \textit{et al.}\textsuperscript{37} found that when the orientation of an inserted HBV fragment (harboring enhancer I) was opposite to the \textit{TERT} transcription direction, the promoter activity was reduced by 40% in comparison to the same-orientation integration. In 2012, Toh \textit{et al.}\textsuperscript{34} found that for \textit{TERT} promoter integrations within 3 kb upstream in HCC tumor samples, the nearer it occurred to the TSS, the higher was the level of downstream \textit{TERT} transcription. In tumors, the transcriptional levels of \textit{TERT} could be over 10-fold higher, and this...
increase has been associated with poorer survival of affected patients.\textsuperscript{16,35,37,49}

**Intragenic and intergenic integrations: fusion transcripts and trans-activation**

Fusion transcript detection from RNA-Seq has revealed that integration sites harboring a DR1 fragment, neighbored by regulatory elements such as the enhancer II, the preC promoter and the hormone response element, may produce virus-host chimeric transcripts.\textsuperscript{19} This type of fusion transcript, commonly originating from interrupted genes, may have trans-acting effects on the regulation of gene expression (Fig. 1; Gene Annotation). In 1990, Takada et al.\textsuperscript{19} reported the 3’ truncated hepatitis B X gene (HBx)-cell fusion product; in 1995, Graef et al.\textsuperscript{50} showed that HBV-mevalonate kinase fusion protein may lead to abnormal phosphorylation of cellular proteins by the affecting metabolism of mevalonate. Saigo et al.\textsuperscript{51} described several HBV integrations located within a 300 bp region within intron 3 flanked by the Alu element of MLL4, resulting in HBx/MLL4 chimeric transcripts and fusion proteins that suppressed expression of certain genes. Then, Dong et al.\textsuperscript{62} found integrations within MLL4 could also occur in exon 3 and intron 5. In their study, the exonic integration resulted in an over 5-fold up-regulation of MLL4 transcription, while the intronic one did not lead to significant changes in gene expression.\textsuperscript{62} Meanwhile, Jiang et al.\textsuperscript{19} observed the one integration, inserting two copies of HBV fragments in the 3’ end of the exon 3 of MLL4 gene, causing an over 20-fold up-regulation of its overall expression level; five HBV-MLL4 integrations interrupting exon 3, 5 and 6 respectively in Furuta et al.’s\textsuperscript{63} study all led to increased MLL4 transcription in tumor samples. Viral integrations within fibronectin 1 (i.e. FN1) are most frequently observed in adjacent, non-tumor samples, and those at intronic ones are in the majority.\textsuperscript{18} However, the frequency of HBV-FNI integration differs greatly among studies (4%, 17/426, Zhao et al.;\textsuperscript{18} 8%, 14/170, Pénéau et al.;\textsuperscript{20} 12.5%, 5/40, Ding et al.;\textsuperscript{35} 19%, 8/42, Furuta et al.;\textsuperscript{63} 40%, 15/41, Furuta et al.).\textsuperscript{63} HBV-FNI integrations were unlikely to significantly influence the expression level of FNI revealed in all these studies, which indicates different transcriptional activity and subsequent biological effects of integrations. Furuta et al.\textsuperscript{63} speculated that HBV-FNI fusion transcripts might produce fusion protein, which can be involved in the pathogenesis of liver fibrosis, considering the regulatory roles of FNI in fibrosis. Nevertheless, Pénéau et al.\textsuperscript{20} argued that HBV-FNI may not have a functional effect since the FNI was not overexpressed due to integration. Future efforts should be made to answer why non-neoplastic hepatocytes carrying this common integration are not more likely to transform into tumor cells.

Intergenic integrations possibly account for functional chimeric noncoding transcripts. The HBx-long interspersed nuclear elements 1 (i.e. L1NEF1) is another type of fusion transcript, acting as a non-coding RNA.\textsuperscript{10,64} Lau et al.\textsuperscript{64} showed functions of HBx-LINE1 did not rely on the fusion protein and it may affect β-catenin trans-activity, which is suggestive of a role played by Wnt signaling activation. Subsequently, Liang et al.\textsuperscript{65} used the cell-line model to confirm that HBx-LINE1 can serve as a miR-122 sponge, which is a liver-specific miRNA and a key regulator in liver diseases.\textsuperscript{64} Lau et al.\textsuperscript{64} reported the incidence of HBx-LINE1 in HCC samples from the Chinese population reached 23.3% (21/90), but Trung et al.\textsuperscript{67} did not detect this characteristic among their 119 Vietnamese patients with HBV-associated HCC. Further efforts may need to clarify the underlying confounding factors, such as heterogeneity of tumor samples and diversity of inserted HBx fragments with different RNA production ability. LINE1-related genes have attracted more and more interest in their roles in tumorigenesis and have been proposed as novel targets in HCC treatment.\textsuperscript{10,68,69}

**Chimeric fragments of HBV DNA and mitochondrial DNA (mtDNA)**

HBV-mtDNA integrations were observed, and the most frequently inserted site reported by Furuta et al.\textsuperscript{63} is hypervariable region 1 (also known as HV 1 or HV1) in the control region of mtDNA, which contains the mitochondrial origin of replication and transcription. Although there is no evidence that HBV DNA can directly integrate into mtDNA, mtDNA can enter the cell nucleus via an uncharacterized process and is widely believed to integrate into the nuclear genome via non-homologous end joining at DSBs, forming nuclear copies of mtDNA (NUMTs).\textsuperscript{70} Therefore, either HBV DNA was incorporated into the NUMTs or mtDNA and HBV dsDNA could also be ligated together via the same mechanism, but this possibility requires further validation by long-read sequencing to read through the entire integration site to determine if HBV-mtDNA chimeric fragments were definitely incorporated into the nuclear genome. Similar to HBV integrations in hepatocytes, the NUMTs also have diverse orientations and multiple fragments within the same site.\textsuperscript{71} Multiple mtDNA fragment insertional mutagenesis at a single genomic site is unlikely. This is apparent since integration into a specific site is a random rather than a directed process. This consideration accounts for why DSB insertion is not likely to occur at a specific site (e.g., previous integration site) in a particular clone. They may originate from multiple copies of mtDNA fragments ligated or concatenated before being incorporated into the nuclear genome.\textsuperscript{71} Delineating the similarities and differences between NUMTs and HBV integrations may provide more clues for clarifying the complicated underlying biological process.

**Repeat elements**: Alu PCR was first applied in HBV integration detection due to its ability to narrow the breadth of regions for analysis,\textsuperscript{72} while accumulated evidence has shown that viral integrations (>50%) tend to locate at repeat regions within the human genome.\textsuperscript{35,72} They include short interspersed nuclear elements (SINE) (including Alu repeats), LINEs and simple repeats (also termed microsatellites). The top repeat element is LINE, followed by LTR, SINE and SINE-VNTR-Alu according to Budzinska et al.’s\textsuperscript{73} observation in chronic hepatitis patients and an in vitro cell model. Nevertheless, we found the most common repeat region was directly interrupted rather than abutting the cancer-enriched human alpha satellite (ALR/Alpha) identified in over 40% of the HCC cases.\textsuperscript{71} (TTAGGG), is the featured sequence of nucleotides in telomeres, which are commonly affected by integrations. It indicates that the insertion of viral fragments may change the telomere length. The telomeres are 8–10 kb long in young individuals, decreasing at a rate of 24.8–27.7 bp per year;\textsuperscript{74} viral integrations may range from hundreds of base pairs to over 3 kb, that increase the length significantly. The elongation of telomeres can prevent affected cells from replicative senescence and sustain cell division activity, leading to genomic instability.\textsuperscript{74} This potential effect may need further experimental validation.

Besides, we found the L1ME1 of family LINE-1 (L1) (~20% cases) is the most frequently affected LINE element.\textsuperscript{21} Interestingly, more evidence has been provided revealing that the activation of L1 contributes to hepatitis virus-related HCC.\textsuperscript{69,75} Particularly, HBV infection suppresses interferon signal transduction by disrupting either STAT1 nuclear im-
port or phosphorylation. The inhibition of interferon is believed to activate the L1 retrotransposon, which subsequently creates DSBs in host cells. LINEs and SINEs are subtypes of transposable elements, which are attracting increasing amounts of attention in recent studies. This interest is due to the realization that they play roles in shaping genome structure by their insertional activity and alteration of transcriptional networks.

**DNA methylation of HBV integration sites**

DNA methylation is one of the main epigenetic modifications regulating gene expression through chromosomal structural alteration, changes in both DNA conformation, and DNA stability. Aberrant methylation of a CpG island in the promoter and gene body are known to lead to transcriptional changes of genes. HBV infection is known to cause changes in the DNA methylation status of hepatocytes, which may not only regulate the viral replication but also contribute to host immune responses. Transcription activity at integration sites is likely to be also under epigenetic regulation of host cells, and most of “functional” viral integrations may have to remain “unsilenced” to continue their involvement in inducing phenotype changes in the infected cells. In 2015, Watanabe et al. described that the methylation status was consistent in the integrant and flanking regions in the host genome in PLC/PRF/5 cell lines and HBV-HCC tumor and adjacent tissues. In 2016, Wang et al. pointed out that HBV integrations in tumor tissues were significantly enriched in the CpG islands, which are crucial sites wherein changes in the DNA methylation status can alter gene expression regulation. In 2020, Zhang et al. found that, within 6,073 different previously identified integration regions, methylation levels of neighboring nucleotides reflected the global hypomethylation in the tumor genome well, regardless of whether integration occurred. This agreement is supportive of the notion that the integration regions tend to be hypomethylated during HCC development independent of integration occurrence. All of these findings point to the idea that the ability of integrations to induce altered biological functions may depend on the corresponding epigenetic status of inserted sites.

**HBV integrants: Latent danger after HBV clearance?**

Even after cccDNA clearance, viral integrants can still exist stably in the host nuclear genome and replicate along with the host genome, as well as most likely not being lost during cell divisions. To the best of our knowledge, there is so far no evidence for the spontaneous removal of viral integrants in affected cells. Genome editing technology is a feasible solution to eliminate the target fragment in the genome. In 2017, Li et al. excised a full-length 3.175-bp integrant in a stable HBV cell line HepG2.A646 using the CRISPR-Cas9 system and also disrupted HBV cccDNA. Assuming viral proteins produced by integrants act as neoantigens in tumor cells, this possibility forms the basis of novel immune therapeutic strategies for HBV-related HCC. In 2019, Tan et al. proposed using expression profiling of HBV integrants in HCC to select T cells for immunotherapy of HCC, and showed that even integrants, which encode not whole HBsAg but fragmented S genes, may produce hepatitis B surface antigen-derived epitopes for T cell receptor-engineered T cell therapy. In 2020, de Beijer et al. attempted to identify the HBx and polymerase-derived T cell epitopes for HBV-directed epitogen-specific immunotherapies. Most of the identified integrants only cover partial fragments and no more than one copy of the entire HBV genome. At least a 1.1-fold over-length HBV genome is required to achieve viral replication when being cloned into plasmids. Therefore, natural HBV integrants are “defective”, which means they are not able to produce viruses. Nevertheless, diverse integrants still have differential activities. In the aforementioned cis-activation effect of integrants within gene promoter regions, regulatory elements in the viral genome, particularly enhancer I, are believed to play important roles. Meanwhile, HBV integrations always harbor the complete open reading frames pre-S/S, and a 3′ truncated X gene. Without viral replication, they still can produce peptides of surface and X proteins, which play crucial roles in tumorigenesis during a long history of HBV infection. Oncogenic roles of HBx include its pleiotropic activities on DNA repair, cell cycle regulation, and diverse signaling pathways. Particularly, most of the integrated X genes are 3′ truncated and able to produce chimeric proteins. Meanwhile, COOH-terminally truncated HBx is known to regulate cell cycle and apoptosis, activate C-Jun/matrix metalloproteinase protein 10 to increase cell proliferation, and enhance tumor cell invasion and metastasis. Early in 1987, Nagaya et al. summarized integrants in 31 reported cases, among which 4 were preS partially deleted. These mutants produce truncated surface proteins that accumulate in liver cells and may cause endoplasmic reticulum stress, with the consequent induction of oxidative DNA damage and genomic instability. In 2017, Watanabe et al. revealed that viral integrations may be a non-negligible source of hepatitis B surface antigen, which may interfere with the endpoint expectations of antiviral therapies, depending on the levels of viral proteins.

Indeed, not all integrations are active and some may be silenced with epigenetic modifications, consistent with surrounding host genomic regions. Nevertheless, genome instability and aberrant methylation of infected hepatocytes may lead to variant accumulation or DNA hypomethylation during liver disease progression. There would be newly-established regulatory relations between integrations and affected genes due to structure variations, or production and accumulation of viral proteins coding by an integrant after abnormal activation, which may act as a latent risk factor in HBV infection that may cause diverse damage to infected liver cells.

**Reflection of integration events on clone expansion of affected hepatocytes**

HBV integration may occur frequently in the chronic hepatitis stage. As early as 1981, Bréchot et al. pointed out that HBV integrations may occur early during infection and other studies showed viral integrations are already present in acute or chronic hepatitis patients. By development and application of the inverted PCR, Summers and Mason were first able to comprehensively investigate the clonal expansion by qualitative analysis of hepatadnaviral DNA integrations in the woodchuck, chimpanzee, and human hepatocytes. Integration can occur immediately after infection, and large hepatocyte clones with viral integrations can be detected in all the stages of chronic hepatitis. For spontaneous DSBs required for viral integrations, the majority appear in the context of DNA replication and about 10 to 50 DSBs occur per cell per day in any given cell, depending on cell cycle and tissue. Considering the sparse and random feature of DSBs, it may require adequate dsDNA as substrates or generate more DSBs in the host genome. In the chronic hepatitis stage, both hepatocyte clone expansion and high HBV replication levels are common, and studies have observed more integration events in relatively normal liver tissues.
that the existence of these kinds of hepatocyte clones and their high level of integration events reflect a substantial risk of hepatocarcinogenesis in the near future. After tumor formation, HCC may not be able to accumulate integrations with the same frequency as in normal hepatocytes. First, HCC has undergone down-regulation of HBV receptor expression, the Na+/taurocholate polypeptide cotransporter (i.e. NTCP), is not susceptible to infection, during which 10% of the viral particles introduce dsDNA into infected cells. Second, efficient HBV replication requires infected cells to maintain a state of cell differentiation, which is missing in HCC. Therefore, the HBV replication level in HCC tissues is likely to be incompatible with that in normal tissues. Halgand et al. observed that HBV pgRNA was detectable in most (90%) HCC non-tumor tissues but in only 67% of the HCC in tumor tissues. Fewer new infections and reduced HBV replication reduce the integration possibility. Regardless of more clone expansion with increased accumulation of DSBs, fewer integration events were detected in HCC tissues.

Integrations do not necessarily occur as driver mutations in tumorigenesis, but they can be found in 80–90% HBV-related HCC. This makes HBV integration a novel marker to trace the clone evolution during tumorigenesis (Fig. 3). Some studies have compared the integrations in multiple lesions from the same patient to determine if they were intrahepatic metastasis or independent multicentric tumors. Sequencing of cell-free DNA (commonly referred to as cfDNA), released from dead cells, will find the chimeric fragments from the integration sites. Interestingly, only integrants originating from tumor clones have been successfully identified, despite the coexistence of large clones carrying integrations in paired normal tissues, and thereby circulating chimeric fragments can aid the early detection of HBV-related HCC. Chen et al. claimed complete resection of the tumor was manifested by the disappearance of integrated fragments in cfDNA. All of these results indicate the limited cell death of relatively normal clones carrying viral integrations. It not only provides a non-invasive way to monitor the tumor clone expansion but also reminds us of the importance of further efforts to explore the interaction between host immune cells and relatively normal hepatocyte clones harboring viral integrations.

**Conclusions**

HBV DNA integration within cell nuclei impedes the process of DNA repair to correct damage due to diverse pressures. This occurrence may directly lead to tumorigenesis or reflect the clone evolution history of affected hepatocytes during chronic infection. Most of the HBV integrations behave like gain-of-function mutations, being responsible for a variable phenotype of affected cells. Biological functions of HBV integrations within hepatocytes associate with their genomic positions within the host nuclear genome, which associate with interrupted genes, newly added promoters, and alterations of the methylation status mediating control of regions in close proximity to regions modulating gene promoter activity. HBV integrants possess gene promoter regulatory activity based on their ability to modify the coding of viral proteins, which also contributes to disease progression. The future profiling of viral integrations requires not only comprehensive efforts to combine multilevel factors in the host genome but also to develop solutions, such as long-read sequencing to read through the entire integration sites to map the entire profile of viral integrants. Taken together, these undertakings will make integration profiling a powerful tool to provide a more accurate evaluation of liver disease progression during HBV chronic infection and design personalized treatments targeting viral integrants.

**Acknowledgments**

We would like to thank Dr. Peter S. Reinach from Wenzhou
Dake Zhang has an authorized patent for the probe-based HBV DNA capture in plasma as a liquid biopsy to monitor HCC development. The other authors have no conflict of interests related to this publication.

Author contributions

Substantial contributions to conception (DZ, KZ), manuscript writing and organizing, collection of data and data interpretation (DZ, KZ, CZ, UP).

References

[1] Zamor PJ, deLemos AS, Russo MW. Viral hepatitis and hepatocellular carcinoma: etiology and management. J Gastrointest Oncol 2017;8(2):229-242. doi: 10.21037/jgo.2017.03.13.

[2] Brechot C, Pourel C, Louise A, Rain B, Tollière P. Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. Nature 1980;286(5772):533-535. doi: 10.1038/286533a0.

[3] Edman JC, Gray P, Valenzuela P, Randall RB, Rutter WJ. Integration of hepatitis B virus sequences and their expression in a human hepatoma cell. Nature 1980;286(5772):533-538. doi: 10.1038/286533a0.

[4] Jiang S, Yang Z, Li W, Li X, Yang J, Zhang J, et al. Re-evaluation of the carcinogenic significance of hepatitis B virus integration in hepatocarcinogenesis. PLoS One 2012;7(9):e40363. doi: 10.1371/journal.pone.0040363.

[5] Davison FD, Fegan EA, Portmann B, Williams R. HBV-DNA sequences in tumor and nontumor tissue in a patient with the fibrolamellar variant of hepatocellular carcinoma. Hepatology 1990;12(4 Pt 1):676-679. doi: 10.1002/hep.1840120410.

[6] Shafritz DA. Integration of HBV-DNA into liver and hepatocellular carcinoma cells during persistent HBV infection. J Cell Biochem 1982;20(3):303-316. doi: 10.1002/jcb.240200310.

[7] Xia Y, Liang TJ. Development of direct-acting antiviral and host-targeting agents for hepatitis B virus infection. Gastroenterology 2019;156(2):311-324. doi: 10.1053/j.gastro.2018.07.057.

[8] Tu T, Budzinska MA, Vondran FWR, Shackel NA, Urban S, Tu T. Cellular genomic sites of hepatitis B virus integration. Proc Natl Acad Sci U S A 2004;101(30):11315-11316. doi: 10.1073/pnas.0304925101.

[9] Xia Y, Liang TJ. Development of direct-acting antiviral and host-targeting agents for hepatitis B virus infection. Gastroenterology 2019;156(2):311-324. doi: 10.1053/j.gastro.2018.07.057.

[10] Tu T, Budzinska MA, Shackel NA, Urban S, Tu T. Cellular genomic sites of hepatitis B virus integration. Proc Natl Acad Sci U S A 2004;101(30):11315-11316. doi: 10.1073/pnas.0304925101.

[11] Bill CA, Summers J. Genomic DNA double-strand breaks are targets for hepatitis B virus integration in human hepatoma cells. J Virol 2000;74(22):11397-11400. doi: 10.1128/JVI.07959-00.

[12] Zhao K, Liu A, Xia Y. Insights into hepatitis B virus DNA integration-55 years of enveloped virus particles. J Virol 2018;92(11):e02007–17. doi: 10.1128/JVI.02007-17.

[13] Heikenwalder M, Protzer U. LINE(1)s of evidence in HBV-driven liver cancer. Cell Host Microbe 2014;15(3):249–250. doi: 10.1016/j.chom.2014.02.015.

[14] Toh ST, Jin Y, Liu L, Wang J, Babrzadeh F, Gharizadeh B, et al. Recurrent targeted genes reveals enriched integration in structural alterations and sequence variations. Carcinogenesis 2013;34(4):787–798. doi: 10.1093/carcin/bgs406.

[15] Ding D, Lou X, Hua D, Yu W, Li L, Wang J, et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. PLoS Genet 2012;8(12):e1003065. doi: 10.1371/journal.pgen.1003065.

[16] Matzinger P, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape of viral associations in human cancers. Nat Genet 2020;52(3):320–330. doi: 10.1038/s41438-019-0559-9.

[17] Sze KM, Ho DW, Chiu YT, Tsui YM, Chan LK, Lee JM, et al. Hepatitis B virus telomerase reverse transcriptase promoter integration harnesses host ELF4, resulting in telomerase reverse transcriptase gene transcription in hepatocellular carcinoma. Hepatology 2021;73(1):23–40. doi: 10.1002/hep.31231.

[18] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Goil U. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020;21(1):30. doi: 10.1186/s13059-020-02195-v.

[19] Chen XP, Long X, Jia WL, Wu HJ, Zhao J, Liang HF, et al. Viral integration drives multifocal HCC during the occult HBV infection. J Exp Clin Cancer Res 2019;38(1):261. doi: 10.1186/s13046-019-1267-y.

[20] Takada S, Gotoh Y, Hayashi S, Yoshida M, Koike K. Structural rearrangement of integrated hepatitis B virus DNA as well as telomere flanking DNA is present in chronically infected tissues. J Virol 1990;64(2):822–828. doi: 10.1128/JVI.64.2.822-828.1990.

[21] Hatada I, Tokino T, Ochiya T, Matsubara K. Co- amplification of integrated hepatitis B virus DNA and transforming gene hst-1 in a hepatocellular carcinoma. Oncogene 1988;3(5):537–540.

[22] Slagle BL, Zhou YZ, Bu X, Heppell K, Yang Y. Hepatitis B virus integration event in human chromosome 17p near the pS3 gene identifies the region of the chromosomal breakpoints.
Zhang D. et al: Genomic functions at HBV integration sites. Journal of Clinical and Translational Hepatology 2021. Vol: 399–408. 407
impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. Cancer Res 2001;61(21):7803–7810.

[97] Ma NF, Lau SH, Hu L, Xie D, Wu J, Yang J, et al. COOH-terminal truncated HBV X protein plays key role in hepatocarcinogenesis. Clin Cancer Res 2008;14(16):5061–5068. doi:10.1158/1078-0432.CCR-07-5082.

[98] Sze KM, Chu GK, Lee JM, Ng IO. C-terminal truncated hepatitis B virus x protein is associated with metastasis and enhances invasiveness by C-Jun/ matrix metalloproteinase protein 10 activation in hepatocellular carcinoma. Hepatology 2013;57(1):131–139. doi:10.1002/hep.25979.

[99] Li W, Li M, Liao D, Lu X, Gu X, Zhang Q, et al. Carboxyl-terminal truncated HBx contributes to invasion and metastasis via deregulating metastasis suppressors in hepatocellular carcinoma. Oncotarget 2016;7(34):55110–55127. doi:10.18632/oncotarget.10399.

[100] Pollicino T, Casciola J, Saffioti F, Raimondo G. Hepatitis B virus PreS/S gene variants: pathobiology and clinical implications. J Hepatol 2014;61(2):408–417. doi:10.1016/j.jhep.2014.04.041.

[101] Wooddell CJ, Yuen MF, Chan HL, Gish RG, Locarnini SA, et al. RNAi-based treatment of chronically infected patients and chimpanzees reveals that integrated hepatitis B virus DNA is a source of HBsAg. Sci Transl Med 2017;9(409):eaan0241. doi:10.1126/scitranslmed.aan0241.

[102] Bréchot C, Hadchouel M, Scotto J, Fonck M, Potet F, et al. State of hepatitis B virus DNA in hepatocytes of patients with hepatitis B surface antigen-positive and -negative liver diseases. Proc Natl Acad Sci U S A 1981;78(6):3906–3910. doi:10.1073/pnas.78.6.3906.

[103] Kimbi GC, Kramvis A, Kew MC. Integration of hepatitis B virus DNA into chromosomal DNA during acute hepatitis B. World J Gastroenterol 2005;11(41):6416–6421. doi:10.3748/wjg.v11.i41.6416.

[104] Tu T, Mason WS, Clouston AD, Shackell NA, McCaughan GW, Yeh MM, et al. Clonal expansion of hepatocytes with a selective advantage occurs during all stages of chronic hepatitis B virus infection. J Viral Hepat 2015;22(9):737–753. doi:10.1111/jvh.12380.

[105] Syeda AH, Hawkins M, McGlynn P. Recombination and replication. Cold Spring Harb Perspect Biol 2014;6(11):a016550. doi:10.1101/cshperspect.a016550.

Zhang D. et al: Genomic functions at HBV integration sites