



Review Article

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

Dake Zhang<sup>1\*</sup> , Ke Zhang<sup>2,3</sup>, Urlike Protzer<sup>3,4</sup> and Changqing Zeng<sup>5,6</sup>

<sup>1</sup>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, China; <sup>2</sup>SCG Cell Therapy Pte. Ltd, Singapore; <sup>3</sup>Institute of Virology, Technical University of Munich/Helmholtz Zentrum München, Munich, Germany; <sup>4</sup>German Center for Infection Research (DZIF), Munich, Germany; <sup>5</sup>Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; <sup>6</sup>University of Chinese Academy of Sciences, Beijing, China

Received: 7 February 2021 | Revised: 29 March 2021 | Accepted: 31 March 2021 | Published: 25 April 2021

## 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 char-

acterization 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

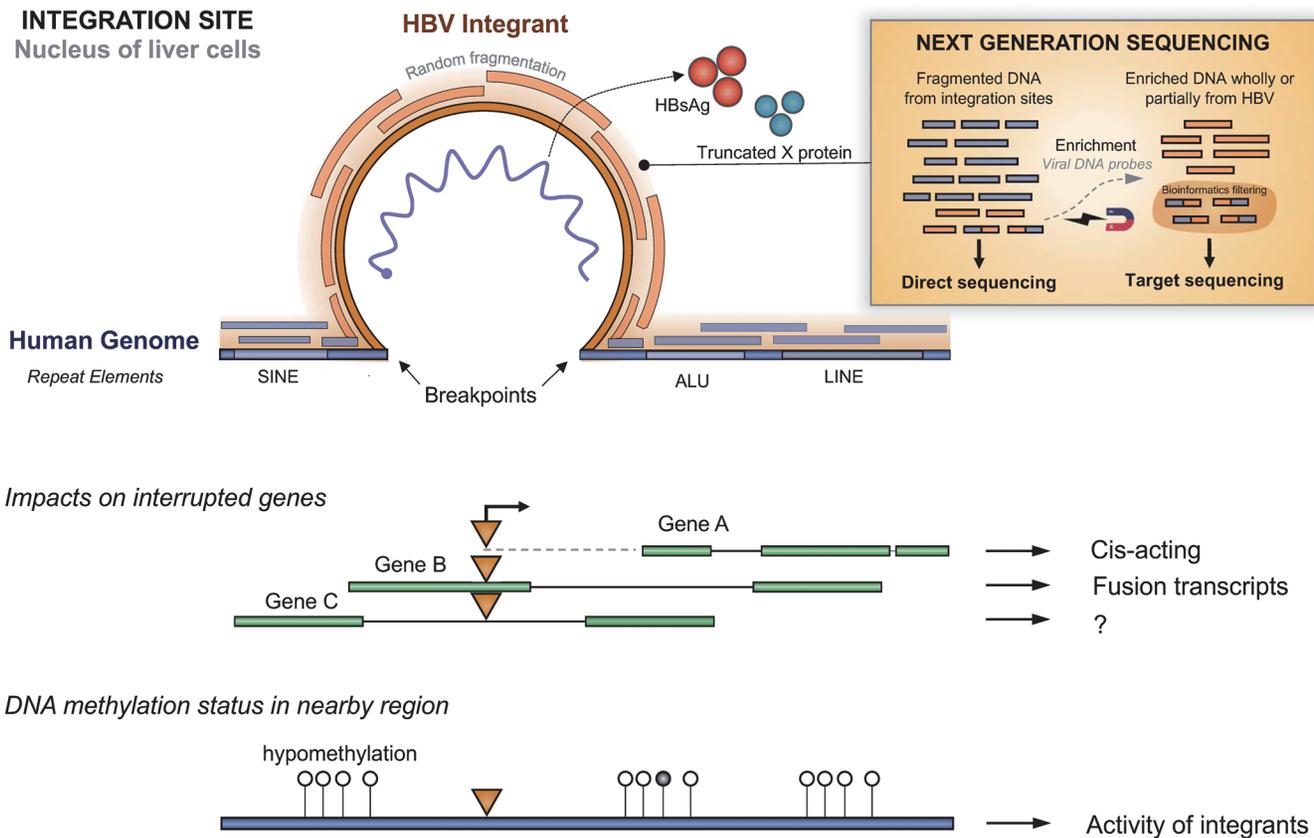
There are still more than 250 million people infected with chronic hepatitis B virus (HBV) worldwide. Hepatocellular carcinoma (HCC) imposes a heavy socioeconomic public health burden. Among patients with cirrhosis, HBV infection has a reported annual incidence of 3%.<sup>1</sup> In 1980, shortly after the discovery of HBV, its integration was reported into HCC tissue cells and hepatoma cell lines.<sup>2,3</sup> There has been a long debate on the oncogenic roles of HBV integrations.<sup>4–6</sup> 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).<sup>7</sup>

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).<sup>8</sup> Indeed, HBV integrations tend to locate at either sites of host genomic instability,<sup>9,10</sup> or sites of cellular DNA damage.<sup>11</sup> Meanwhile, the HBV genome does not encode either integrase or other proteins recognizing specific genomic regions for integration,<sup>9,12,13</sup> 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; cfDNA, cell-free DNA dsDNA, double-stranded linear DNA; DSB, DNA double-stranded breaks; *FN1*, 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; mtDNA, mitochondrial DNA; NGS, next-generation of sequencing; *NTCP*, Na<sup>+</sup>/taurocholate polypeptide cotransporter; NUMT, nuclear copies of mtDNA; pgRNA, pregenomic RNA; SINE, short interspersed nuclear elements; *TP53*, 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](mailto:dakezhang@buaa.edu.cn)



**Fig. 1. HBV integrations in the host nuclear genome.** HBV DNA fragments are incorporated into sites in the host nuclear genome, which may include double-stranded breaks due to diverse factors. Viral integrants can also be transcribed and translated to express viral proteins such as hepatitis B surface antigen and truncated X protein, which are involved in liver disease progression. It is well known that repeat elements are characteristic of the integration region, such as ALUs, SINE and LINES. Recently, solutions based on the NGS platform directly determine all of the DNA fragments from the integration sites. Viral-host chimeric fragments are evidence of the integration events. Particularly, these chimeric fragments can be enriched by using viral probes to capture those containing viral DNA, which increases the sensitivity of integration detection and extensively reduce the sequencing volume. The “breakpoints” of integrations refer to the position on the chimeric fragments where viral and human DNA fragments ligate with each other. They are also the boundaries of the integrations. Integrations may generate either a new promoter region to *cis*-activate the transcription of downstream genes, or directly interrupt the genes and potentially produce fusion transcripts. DNA methylation status in a nearby region may regulate the transcriptional activity of integrants.

are the main genomic feature of integration sites.<sup>14</sup> 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.<sup>15</sup> About 37–40% of the reported viral breakpoints are mapped within the DR2-DR1 region,<sup>16–18</sup> 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 dsIDNA uses an 18 nucleotide RNA primer, which is the hydrolysate of pgRNA, to initiate the synthesis of the positive-sense DNA strand.<sup>12</sup> 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 dsIDNA as the main source of viral DNA incorporated into the host genome.<sup>12</sup>

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.<sup>9,12</sup> The development of high-throughput sequencing solutions significantly facilitates resolving thousands of viral

integration breakpoints.<sup>16,18–22</sup> 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,<sup>21</sup> and the PLC/PRF/5 cell line had nine integration regions reported.<sup>23</sup> Among the cell populations of the infected liver, it is estimated that there is about 1 integration per 10<sup>3</sup>–10<sup>4</sup> cells according to observations in an *in vitro* infection model.<sup>8</sup> 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.<sup>2,24</sup> In Southern blot hybridization, digested DNA after a restriction enzyme treatment hybridized with <sup>32</sup>P-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 *in situ* hybridization. Prior to Sanger sequencing of integrant clones,<sup>25</sup> *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.<sup>26</sup> 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 *et al.*<sup>27</sup> (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,<sup>28</sup> 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 sequences (X gene), may successfully amplify the virus-host chimeric regions. With this approach, Devrim *et al.*<sup>29</sup> identified 21 viral integrations from 18 patients in one study. Likewise, Mason *et al.*<sup>30–33</sup> 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 *et al.*<sup>19</sup> 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.<sup>16,21</sup> With the increased throughput, Zhao *et al.*<sup>16</sup> 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.<sup>17,34–37</sup>

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.<sup>21</sup> 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

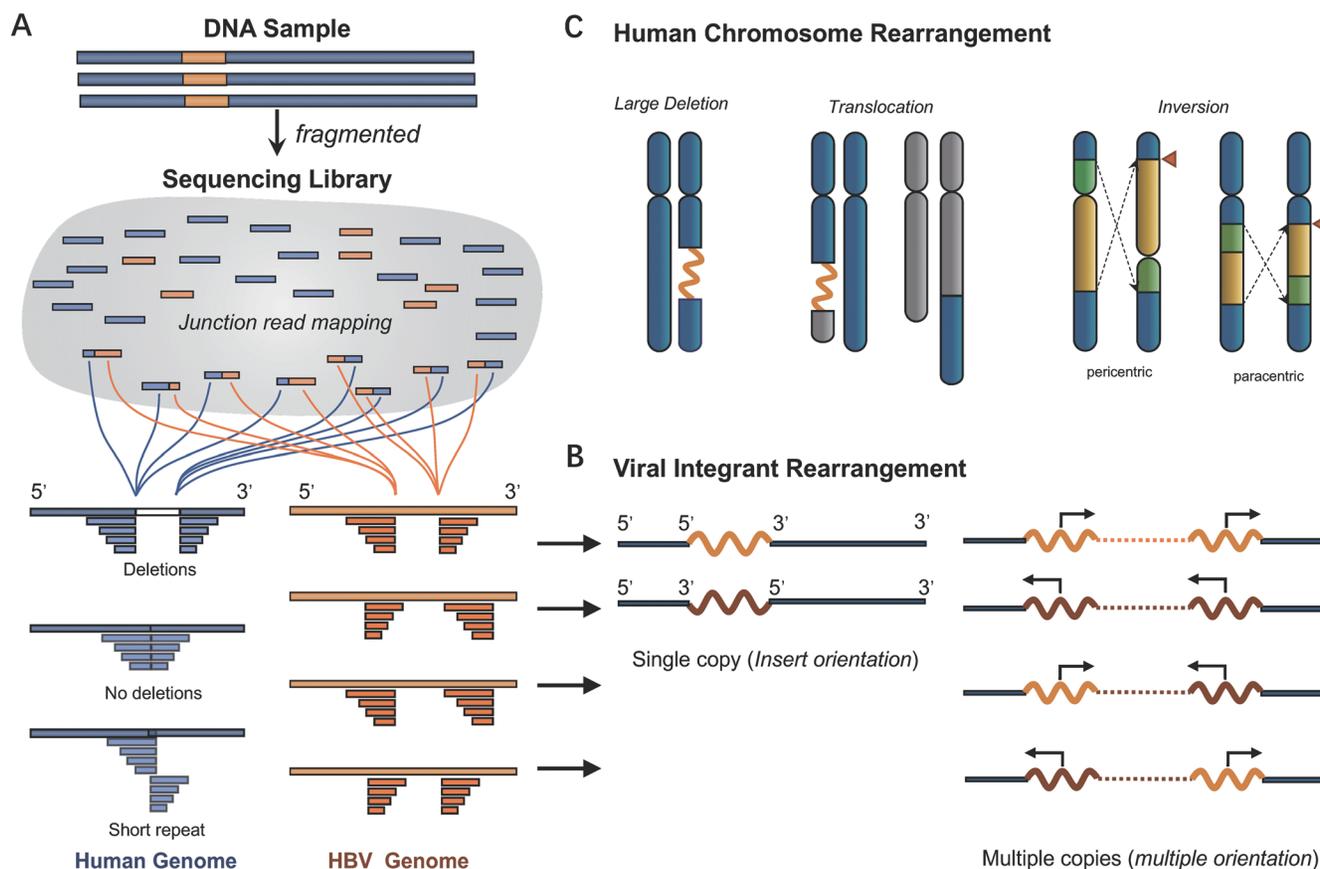
human genome for the same integration event.<sup>21</sup> After the breakpoints in the human genome have been successfully paired, the corresponding viral breakpoints can then be paired. The boundaries of integration events are indicated by the sharp and consistent positions of the mapped reads, the 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.<sup>38</sup> 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 regulatory activity of the integration.

Viral integrations have been reported to locate at the boundaries of chromosome arrangements, 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.<sup>39,40</sup> 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 Mb, is able to directly read through the entire integration region.<sup>23</sup> 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.<sup>23</sup>

### 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 *HST-1*,<sup>41</sup> tumor protein p53 (i.e. *TP53*),<sup>25,42</sup> cyclin A2 (*CCNA2*),<sup>43</sup> *myc* family oncogenes,<sup>44,45</sup> retinoic acid receptor beta or erb-B-like genes,<sup>46,47</sup> and mevalonate kinase gene.<sup>48</sup> 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. *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 *et al.*;<sup>49</sup> 24%, 101/426, Sun *et al.*;<sup>18</sup> 24%, 101/426, Zhao *et al.*;<sup>16</sup> 27%, 48/177, Péneau *et al.*<sup>50</sup>) to over 35% (36%, 34/95, Sze *et al.*<sup>37</sup>). The second most common recurrent gene was lysine methyltransferase 2B (*KMT2B*, also known as *MLL4* or *MLL2*), encoding histone methyltransferases, with a frequency of about 10% (2%, Péneau *et al.*;<sup>50</sup> 7%, Zhao *et al.*;<sup>16</sup> 7%, Fujimoto *et al.*;<sup>49</sup> 12%, Sun *et al.*;<sup>18</sup> 12%, Sze *et al.*<sup>37</sup>). They were followed by *CCNE1* (1%, Fujimoto *et al.*;<sup>49</sup> 1.6%, Zhao *et al.*;<sup>16</sup> 2%, Péneau *et al.*;<sup>50</sup> 3%, Sze *et al.*;<sup>37</sup> 5% Sun *et al.*<sup>18</sup>) and *CCNA2* (1%, Fujimoto *et al.*;<sup>49</sup> 1.8%, Zhao *et al.*;<sup>16</sup> 2%, Sze *et al.*<sup>37</sup>).

Whereas, integrations seem to account for a relatively small proportion of crucial genomic changes of total HCC.<sup>51</sup> For instance, about 70% of all HCC carry *TERT* mutations, and HBV integrations contribute to about 7% of them.<sup>52</sup> In other words, 5% of HCC are due to *TERT* integration;



**Fig. 2. Integration identification based on short-read sequencing and confounding factors.** (A) In current solutions based on short-read sequencing, DNA samples will be fragmented into ideal size for subsequent library construction. Sequencing experiments will generate millions of reads, generally 150 bp in length, which need further bioinformatic analysis to map the reference host and viral genomes. Junction or chimeric reads, originating from an integration boundary, will be split into two pieces that are mapped to the host or viral reference genomes, respectively. Each integration event has two breakpoints, which can be paired according to their positions in the human genome. After successful pairing, read mapping patterns of paired breakpoints indicate the complicated rearrangement of viral integrants within the integration sites. (B) A single copy of an integrant may have either forward or reverse 5'-3' orientations with the host genome. The multiple copies of viral fragments are joined at distinct orientations before being incorporated into the integration site. Their orientation will determine which of them are downstream and affected by the integrant due to the *cis*-activation effect. (C) Large deletions, translocations and inversions are capable of joining two distant regions together and integration events tend to occur at the boundary of these structural variations. Such interaction will prevent pairing two breakpoints far apart from one another in the same integration event without any information about structural variations.

moreover, 58% of HCC have genomic changes affecting cell cycle control, and HBV integrations account for about 8% of those interrupting the *CCNE1*.<sup>52</sup> Considering HBV infection contributing to 50% of HCC cases, the contribution of integrations should be doubled in HBV-related HCC.<sup>53</sup>

#### **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 *hTERT*<sup>54</sup>) or over 100 kb (e.g., 2,582 kb upstream of *CCNE1*<sup>55</sup>) from the genes in the vicinity along the human genome. In the first large-scale screening for HBV integrations in HCC, Sun *et al.*<sup>18</sup> 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 *et al.*<sup>49</sup> enlarged this region up to 10 kb upstream, which has become widely adopted. Nevertheless, the frequency and influence of inte-

gration in the promoter may be underestimated since some are located beyond the 10 kb region,<sup>56</sup> and Ding *et al.*<sup>35</sup> 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<sup>57</sup> had *TERT*-related integrations, of which 80% are located in the promoter region.<sup>18</sup> In 2001, Horikawa *et al.*<sup>58</sup> first described the HBV integrations in the *cis*-activated *TERT* promoter in an orientation-independent manner. Recently, Sze *et al.*<sup>37</sup> found that when the orientation of an inserted HBV fragment (harboring enhancer I) was opposite to the *TERT* transcription direction, the promoter activity was reduced by 40% in comparison to the same-orientation integration. In 2012, Toh *et al.*<sup>34</sup> found that for *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 *TERT* transcription. In tumors, the transcriptional levels of *TERT* could be over 10-fold higher, and this

increase has been associated with poorer survival of affected patients.<sup>16,35,37,49</sup>

### **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.<sup>19</sup> 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.*<sup>59</sup> reported the 3' truncated hepatitis B X gene (HBx)-cell fusion product; in 1995, Graef *et al.*<sup>60</sup> showed that HBV-mevalonate kinase fusion protein may lead to abnormal phosphorylation of cellular proteins by the affecting metabolism of mevalonate. Saigo *et al.*<sup>61</sup> described several HBV integrations located within a 300bp 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.*<sup>62</sup> 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.<sup>62</sup> Meanwhile, Jiang *et al.*<sup>19</sup> 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<sup>63</sup> 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.<sup>18</sup> However, the frequency of HBV-*FN1* integration differs greatly among studies (4%, 17/426, Zhao *et al.*;<sup>16</sup> 8%, 14/170, Péneau *et al.*;<sup>50</sup> 12.5%, 5/40, Ding *et al.*;<sup>35</sup> 19%, 8/42, Furuta *et al.*;<sup>63</sup> 40%, 15/41, Furuta *et al.*<sup>63</sup>). HBV-*FN1* integrations were unlikely to significantly influence the expression level of *FN1* revealed in all these studies, which indicates different transcriptional activity and subsequent biological effects of integrations. Furuta *et al.*<sup>63</sup> speculated that HBV-*FN1* fusion transcripts might produce fusion protein, which can be involved in the pathogenesis of liver fibrosis, considering the regulatory roles of FN1 in fibrosis. Nevertheless, Péneau *et al.*<sup>50</sup> argued that HBV-*FN1* may not have a functional effect since the *FN1* 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. LINE1) is another type of fusion transcript, acting as a non-coding RNA.<sup>10,64</sup> Lau *et al.*<sup>64</sup> showed functions of HBx-*LINE1* did not rely on the fusion protein and it may affect  $\beta$ -catenin trans-activity, which is suggestive of a role played by Wnt signaling activation. Subsequently, Liang *et al.*<sup>65</sup> 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.<sup>66</sup> Lau *et al.*<sup>64</sup> reported the incidence of HBx-*LINE1* in HCC samples from the Chinese population reached 23.3% (21/90), but Trung *et al.*<sup>67</sup> did not detect this transcript in 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.<sup>10,68,69</sup>

### **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.*<sup>63</sup> is hypervariable region I (also known as HV I 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).<sup>70</sup> 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.<sup>71</sup> 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.<sup>71</sup> 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,<sup>27</sup> while accumulated evidence has shown that viral integrations (>50%) tend to locate at repeat regions within the human genome.<sup>35,72</sup> 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<sup>73</sup> 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.<sup>21</sup>

(TTAGGG)<sub>n</sub> 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;<sup>74</sup> 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.<sup>74</sup> 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.<sup>21</sup> Interestingly, more evidence has been provided revealing that the activation of L1 contributes to hepatitis virus-related HCC.<sup>69,75</sup> Particularly, HBV infection suppresses interferon signal transduction by disrupting either STAT1 nuclear im-

port or phosphorylation.<sup>76</sup> The inhibition of interferon is believed to activate the L1 retrotransposon,<sup>77</sup> which subsequently creates DSBs in host cells.<sup>78</sup> 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.<sup>79</sup>

### 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.<sup>80</sup> Aberrant methylations of a CpG island in the promoter and gene body are known to lead to transcriptional changes of genes.<sup>81</sup> 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.<sup>82,83</sup> 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.*<sup>84</sup> 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.*<sup>16</sup> 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.*<sup>85</sup> 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.<sup>86</sup> 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.<sup>87</sup> In 2017, Li *et al.*<sup>88</sup> excised a full-length 3,175-bp integrant in a stable HBV cell line HepG2.A64 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.<sup>89,90</sup> In 2019, Tan *et al.*<sup>91</sup> 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.*<sup>92</sup> attempted to identify the HBx and polymerase-derived T cell epitopes for effective HBV antigen-specific immunotherapies.

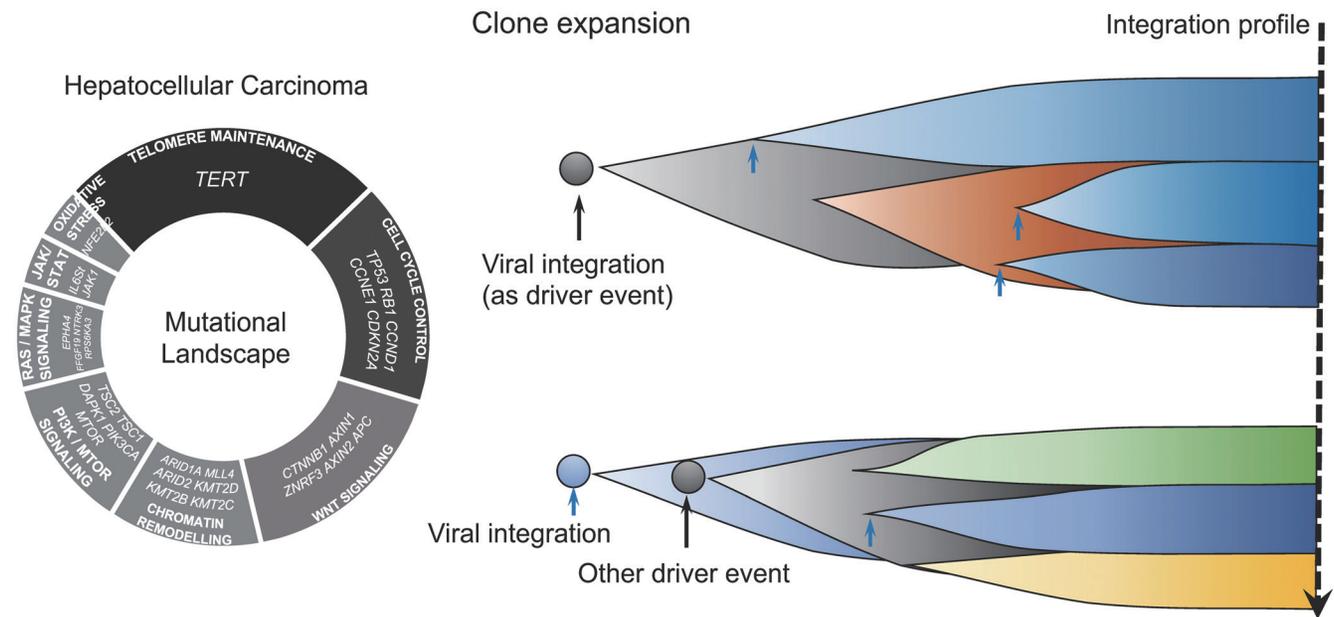
Most of the identified integrants only cover partial fragments and no more than one copy of the entire HBV ge-

nome. At least a 1.1-fold over-length HBV genome is required to achieve viral replication when being cloned into plasmids.<sup>93</sup> Therefore, natural HBV integrants are "defective", which means they are not able to produce viruses.<sup>94</sup> Nevertheless, diverse integrants still have differential activities. In the aforementioned *cis*-activation effect of integrations 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.<sup>12</sup> 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.<sup>95</sup> Particularly, most of the integrated X genes are 3' truncated and able to produce chimeric proteins.<sup>34</sup> Meanwhile, COOH-terminally truncated HBx is known to regulate cell cycle and apoptosis,<sup>96,97</sup> activate C-Jun/matrix metalloproteinase protein 10 to increase cell proliferation,<sup>98</sup> and enhance tumor cell invasion and metastasis.<sup>99</sup> Early in 1987, Nagaya *et al.*<sup>94</sup> 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.<sup>100</sup> In 2017, Wooddell *et al.*<sup>101</sup> 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.<sup>84</sup> 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, Bréchet *et al.*<sup>102</sup> pointed out that HBV integrations may occur early during infection,<sup>103</sup> 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<sup>31-33</sup> were first able to comprehensively investigate the clonal expansion by quantitative analysis of hepadnaviral 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.<sup>8,21,104</sup> For spontaneous DSBs required for viral integrations, the majority appear in the context of DNA replication,<sup>105</sup> and about 10 to 50 DSBs occur per cell per day in any given cell, depending on cell cycle and tissue.<sup>106</sup> Considering the sparse and random feature of DSBs, it may require adequate dsDNA as substrates or generate more DSBs in the host genome.<sup>107</sup> 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.<sup>35,63</sup> Mason *et al.*<sup>108</sup> believe



**Fig. 3. Oncogenic roles of viral integrations in clone evolution of tumorigenesis.** Activity changes in diverse cell signaling pathways due to genomic mutations are known to induce HCC. Viral integrations are not the only mutagenic factor responsible for these changes; they only account for a small proportion. The integration profile of a particular tumor sample contains both integration events providing both driver and passenger events. They accumulate during distinct stages of tumor clone evolution.

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<sup>+</sup>/taurocholate polypeptide cotransporter (i.e. *NTCP*),<sup>109</sup> 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,<sup>110</sup> 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.*<sup>111</sup> 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.<sup>35,63</sup>

Integrations do not necessarily occur as driver mutations in tumorigenesis, but they can be found in 80–90% HBV-related HCC.<sup>18</sup> 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.<sup>112</sup> 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.<sup>21,113</sup> Chen *et al.*<sup>21</sup> reported no viral integration in cfDNA from chronic hepatitis patients, and Wang *et al.*<sup>114</sup> 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 nor-

mal 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

Medical University for language editing and proofreading of the manuscript.

## Funding

This work was supported by the 111Project (Project No.: B13003), Innovation Promotion Association CAS (2016098), and National Natural Science Foundation of China (81201700) to D.Z.

## Conflict of interest

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

- 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.2017.03.14.
- Brechot C, Pourcel C, Louise A, Rain B, Tiollais 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.
- Edman JC, Gray P, Valenzuela P, Rall LB, Rutter WJ. Integration of hepatitis B virus sequences and their expression in a human hepatoma cell. *Nature* 1980;286(5772):535–538. doi:10.1038/286535a0.
- Jiang S, Yang Z, Li W, Li X, Wang Y, 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.
- Davidson FD, Fagan 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.
- 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.
- Xia Y, Liang TJ. Development of direct-acting antiviral and host-targeting agents for treatment of hepatitis B virus infection. *Gastroenterology* 2019;156(2):311–324. doi:10.1053/j.gastro.2018.07.057.
- Tu T, Budzinska MA, Vondran FWR, Shackel NA, Urban S. Hepatitis B virus DNA integration occurs early in the viral life cycle in an *in vitro* infection model via sodium taurocholate cotransporting polypeptide-dependent uptake of enveloped virus particles. *J Virol* 2018;92(11):e02007–17. doi:10.1128/JVI.02007-17.
- Tu T, Budzinska MA, Shackel NA, Urban S. HBV DNA integration: Molecular mechanisms and clinical implications. *Viruses* 2017;9(4):75. doi:10.3390/v9040075.
- 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.
- Bill CA, Summers J. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc Natl Acad Sci U S A* 2004;101(30):11135–11140. doi:10.1073/pnas.0403925101.
- Zhao K, Liu A, Xia Y. Insights into hepatitis B virus DNA integration-55 years after virus discovery. *The Innovation* 2020;1(2):100034. doi:10.1016/j.xinn.2020.100034.
- Tsukuda S, Watashi K. Hepatitis B virus biology and life cycle. *Antiviral Res* 2020;182:104925. doi:10.1016/j.antiviral.2020.104925.
- Budzinska MA, Shackel NA, Urban S, Tu T. Cellular genomic sites of hepatitis B virus DNA integration. *Genes (Basel)* 2018;9(7):365. doi:10.3390/genes9070365.
- Dejean A, Sonigo P, Wain-Hobson S, Tiollais P. Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11-base-pair direct repeat. *Proc Natl Acad Sci U S A* 1984;81(17):5350–5354. doi:10.1073/pnas.81.17.5350.
- Zhao LH, Liu X, Yan HX, Li WY, Zeng X, Yang Y, *et al*. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun* 2016;7:12992. doi:10.1038/ncomms12992.
- Li X, Zhang J, Yang Z, Kang J, Jiang S, Zhang T, *et al*. The function of targeted host genes determines the oncogenicity of HBV integration in hepa-

- tocellular carcinoma. *J Hepatol* 2014;60(5):975–984. doi:10.1016/j.jhep.2013.12.014.
- Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, *et al*. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012;44(7):765–769. doi:10.1038/ng.2295.
- Jiang Z, Jhunjunwala S, Liu J, Haverty PM, Kennemer MI, Guan Y, *et al*. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Res* 2012;22(4):593–601. doi:10.1101/gr.133926.111.
- Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, *et al*. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 2012;44(7):760–764. doi:10.1038/ng.2291.
- Chen W, Zhang K, Dong P, Fanning G, Tao C, Zhang H, *et al*. Noninvasive chimeric DNA profiling identifies tumor-originated HBV integrants contributing to viral antigen expression in liver cancer. *Hepatology* 2020;72(3):326–337. doi:10.1002/hep.24202.
- Li CL, Ho MC, Lin YY, Tzeng ST, Chen YJ, Pai HY, *et al*. Cell-free virus-host chimera DNA from hepatitis B virus integration sites as a circulating biomarker of hepatocellular cancer. *Hepatology* 2020;72(6):2063–2076. doi:10.1002/hep.31230.
- Meng G, Tan Y, Fan Y, Wang Y, Yang G, Fanning G, *et al*. TSD: A computational tool to study the complex structural variants using PacBio targeted sequencing data. *G3 (Bethesda)* 2019;9(5):1371–1376. doi:10.1534/g3.118.200900.
- Chakraborty PR, Ruiz-Opazo N, Shouval D, Shafritz DA. Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg-producing human hepatocellular carcinoma cell line. *Nature* 1980;286(5772):531–533. doi:10.1038/286531a0.
- Zhou YZ, Slagle BL, Donehower LA, vanTuinen P, Ledbetter DH, Butel JS. Structural analysis of a hepatitis B virus genome integrated into chromosome 17p of a human hepatocellular carcinoma. *J Virol* 1988;62(11):4224–4231. doi:10.1128/JVI.62.11.4224-4231.1988.
- Tokino T, Matsubara K. Chromosomal sites for hepatitis B virus integration in human hepatocellular carcinoma. *J Virol* 1991;65(12):6761–6764. doi:10.1128/JVI.65.12.6761-6764.1991.
- Minami M, Poussin K, Bréchet C, Paterlini P. A novel PCR technique using Alu-specific primers to identify unknown flanking sequences from the human genome. *Genomics* 1995;29(2):403–408. doi:10.1006/geno.1995.9004.
- Deininger P. Alu elements: know the SINEs. *Genome Biol* 2011;12(12):236. doi:10.1186/gb-2011-12-12-236.
- Gozuacik D, Murakami Y, Saigo K, Chami M, Mugnier C, Lagorce D, *et al*. Identification of human cancer-related genes by naturally occurring Hepatitis B Virus DNA tagging. *Oncogene* 2001;20(43):6233–6240. doi:10.1038/sj.onc.1204835.
- Mason WS, Liu C, Aldrich CE, Litwin S, Yeh MM. Clonal expansion of normal-appearing human hepatocytes during chronic hepatitis B virus infection. *J Virol* 2010;84(16):8308–8315. doi:10.1128/JVI.00833-10.
- Mason WS, Jilbert AR, Summers J. Clonal expansion of hepatocytes during chronic woodchuck hepatitis virus infection. *Proc Natl Acad Sci U S A* 2005;102(4):1139–1144. doi:10.1073/pnas.0409332102.
- Summers J, Jilbert AR, Yang W, Aldrich CE, Saputelli J, Litwin S, *et al*. Hepatocyte turnover during resolution of a transient hepadnaviral infection. *Proc Natl Acad Sci U S A* 2003;100(20):11652–11659. doi:10.1073/pnas.1204835.
- Mason WS, Low HC, Xu C, Aldrich CE, Scougall CA, Grosse A, *et al*. Detection of clonally expanded hepatocytes in chimpanzees with chronic hepatitis B virus infection. *J Virol* 2009;83(17):8396–8408. doi:10.1128/JVI.00700-09.
- Toh ST, Jin Y, Liu L, Wang J, Babrzadeh F, Garizadeh B, *et al*. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. *Carcinogenesis* 2013;34(4):787–798. doi:10.1093/carcin/bgs406.
- 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.
- Zapatka M, 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/s41588-019-0558-9.
- 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.
- Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouli Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020;21(1):30. doi:10.1186/s13059-020-1935-5.
- 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/s13046-019-1273-1.
- Takada S, Gotoh Y, Hayashi S, Yoshida M, Koike K. Structural rearrangement of integrated hepatitis B virus DNA as well as cellular flanking DNA is present in chronically infected hepatic tissues. *J Virol* 1990;64(2):822–828. doi:10.1128/JVI.64.2.822-828.1990.
- 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.
- Slagle BL, Zhou YZ, Butel JS. Hepatitis B virus integration event in human chromosome 17p near the p53 gene identifies the region of the chromo-

- some commonly deleted in virus-positive hepatocellular carcinomas. *Cancer Res* 1991;51(1):49–54.
- [43] Wang J, Zindy F, Chenivisse X, Lamas E, Henglein B, Br  chot C. Modification of cyclin A expression by hepatitis B virus DNA integration in a hepatocellular carcinoma. *Oncogene* 1992;7(8):1653–1656.
- [44] Wei Y, Ponzetto A, Tiollais P, Bendia MA. Multiple rearrangements and activated expression of c-myc induced by woodchuck hepatitis virus integration in a primary liver tumour. *Res Virol* 1992;143(2):89–96. doi:10.1016/s0923-2516(06)80086-5.
- [45] Baba M, Hasegawa H, Nakayabu M, Tamaki S, Watanabe S, Shima T, *et al*. Characteristics of human hepatocellular carcinoma cell lines (Hep-KANO) derived from a non-hepatic, non-cirrhotic hepatitis B virus carrier. *Jpn J Cancer Res* 1994;85(11):1105–1111. doi:10.1111/j.1349-7006.1994.tb02914.x.
- [46] Zhang XK, Egan JO, Huang D, Sun ZL, Chien VK, Chiu JF. Hepatitis B virus DNA integration and expression of an erb B-like gene in human hepatocellular carcinoma. *Biochem Biophys Res Commun* 1992;188(1):344–351. doi:10.1016/0006-291x(92)92391-a.
- [47] Dejean A, Bougueleret L, Grzeschik KH, Tiollais P. Hepatitis B virus DNA integration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma. *Nature* 1986;322(6074):70–72. doi:10.1038/322070a0.
- [48] Graef E, Caselmann WH, Wells J, Koshy R. Insertional activation of mevalonate kinase by hepatitis B virus DNA in a human hepatoma cell line. *Oncogene* 1994;9(1):81–87.
- [49] Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraiishi Y, *et al*. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet* 2016;48(5):500–509. doi:10.1038/ng.3547.
- [50] P  neau C, Imbeaud S, La Bella T, Hirsch TZ, Caruso S, Calderaro J, *et al*. Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma. *Gut* 2021. doi:10.1136/gutjnl-2020-323153.
- [51] Villanueva A. Hepatocellular carcinoma. *N Engl J Med* 2019;380(15):1450–1462. doi:10.1056/NEJMra1713263.
- [52] Schulze K, Nault JC, Villanueva A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *J Hepatol* 2016;65(5):1031–1042. doi:10.1016/j.jhep.2016.05.035.
- [53] Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, *et al*. Hepatocellular carcinoma. *Nat Rev Dis Primers* 2021;7(1):6. doi:10.1038/s41572-020-00240-3.
- [54] Kekul   AS, Lauer U, Meyer M, Caselmann WH, Hofschneider PH, Koshy R. The preS2/S region of integrated hepatitis B virus DNA encodes a transcriptional transactivator. *Nature* 1990;343(6257):457–461. doi:10.1038/343457a0.
- [55] Tsuei DJ, Hsu TY, Chen JY, Chang MH, Hsu HC, Yang CS. Analysis of integrated hepatitis B virus DNA and flanking cellular sequences in a childhood hepatocellular carcinoma. *J Med Virol* 1994;42(3):287–293. doi:10.1002/jmv.1890420316.
- [56] Murakami Y, Saigo K, Takashima H, Minami M, Okanoue T, Br  chot C, *et al*. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. *Gut* 2005;54(8):1162–1168. doi:10.1136/gut.2004.054452.
- [57] Nault JC, Ningarhari M, Rebouissou S, Zucman-Rossi J. The role of telomeres and telomerase in cirrhosis and liver cancer. *Nat Rev Gastroenterol Hepatol* 2019;16(9):544–558. doi:10.1038/s41575-019-0165-3.
- [58] Horikawa I, Barrett JC. cis-Activation of the human telomerase gene (hTERT) by the hepatitis B virus genome. *J Natl Cancer Inst* 2001;93(15):1171–1173. doi:10.1093/jnci/93.15.1171.
- [59] Takada S, Koike K. Trans-activation function of a 3' truncated X gene-cell fusion product from integrated hepatitis B virus DNA in chronic hepatitis tissues. *Proc Natl Acad Sci U S A* 1990;87(15):5628–5632. doi:10.1073/pnas.87.15.5628.
- [60] Graef E, Caselmann WH, Hofschneider PH, Koshy R. Enzymatic properties of overexpressed HBV-mevalonate kinase fusion proteins and mevalonate kinase proteins in the human hepatoma cell line PLC/PRF/5. *Virology* 1995;208(2):696–703. doi:10.1006/viro.1995.1201.
- [61] Saigo K, Yoshida K, Ikeda R, Sakamoto Y, Murakami Y, Urashima T, *et al*. 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* 2008;29(5):703–708. doi:10.1002/humu.20701.
- [62] Dong H, Zhang L, Qian Z, Zhu X, Zhu G, Chen Y, *et al*. Identification of HBV-MLL4 integration and its molecular basis in Chinese hepatocellular carcinoma. *PLoS One* 2015;10(4):e0123175. doi:10.1371/journal.pone.0123175.
- [63] Furuta M, Tanaka H, Shiraiishi Y, Unida T, Imamura M, Fujimoto A, *et al*. Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. *Oncotarget* 2018;9(38):25075–25088. doi:10.18632/oncotarget.25308.
- [64] Lau CC, Sun T, Ching AK, He M, Li JW, Wong AM, *et al*. Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell* 2014;25(3):335–349. doi:10.1016/j.ccr.2014.01.030.
- [65] Liang HW, Wang N, Wang Y, Wang F, Fu Z, Yan X, *et al*. Hepatitis B virus-human chimeric transcript HBx-LINE1 promotes hepatic injury via sequestering cellular microRNA-122. *J Hepatol* 2016;64(2):278–291. doi:10.1016/j.jhep.2015.09.013.
- [66] Fu X, Calin GA. miR-122 and hepatocellular carcinoma: from molecular biology to therapeutics. *EBioMedicine* 2018;37:17–18. doi:10.1016/j.ebiom.2018.10.032.
- [67] Trung NT, Hai LT, Giang DP, Hoan PQ, Binh MT, Hoan NX, *et al*. No expression of HBV-human chimeric fusion transcript (HBx-LINE1) among Vietnamese patients with HBV-associated hepatocellular carcinoma. *Ann Hepatol* 2019;18(2):404–405. doi:10.1016/j.aohpe.2019.02.002.
- [68] Ardeljan D, Wang X, Oghbaie M, Taylor MS, Husband D, Deshpande V, *et al*. LINE-1 ORF2p expression is nearly imperceptible in human cancers. *Mob DNA* 2019;11:1. doi:10.1186/s13100-019-0191-2.
- [69] Honda T. Links between human LINE-1 retrotransposons and hepatitis virus-related hepatocellular carcinoma. *Front Chem* 2016;4:21. doi:10.3389/fchem.2016.00021.
- [70] Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet* 2010;6(2):e1000834. doi:10.1371/journal.pgen.1000834.
- [71] Wei W, Pagnamenta AT, Gleadall N, Sanchis-Juan A, Stephens J, Broxholme J, *et al*. Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans. *Nat Commun* 2020;11(1):1740. doi:10.1038/s41467-020-15336-3.
- [72] Yan H, Yang Y, Zhang L, Tang G, Wang Y, Xue G, *et al*. Characterization of the genotype and integration patterns of hepatitis B virus in early- and late-onset hepatocellular carcinoma. *Hepatology* 2015;61(6):1821–1831. doi:10.1002/hep.27722.
- [73] Budzinska MA, Shackel NA, Urban S, Tu T. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg Microbes Infect* 2018;7(1):142. doi:10.1038/s41426-018-0145-7.
- [74] Shammass MA. Telomeres, lifestyle, cancer, and aging. *Curr Opin Clin Nutr Metab Care* 2011;14(1):28–34. doi:10.1097/MCO.0b013e32834121b1.
- [75] Schauer SN, Carreira PE, Shukla R, Gerhardt DJ, Gerdes P, Sanchez-Luque FJ, *et al*. L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res* 2018;28(5):639–653. doi:10.1101/gr.226693.117.
- [76] L  tgehetmann M, Bornscheuer T, Volz T, Allweiss L, Bockmann JH, Pollok JM, *et al*. Hepatitis B virus limits response of human hepatocytes to interferon-  in chimeric mice. *Gastroenterology* 2011;140(7):2074–2083, 2083.e1-2. doi:10.1053/j.gastro.2011.02.057.
- [77] Yu Q, Carbone CJ, Katlinskaya YV, Zheng H, Zheng K, Luo M, *et al*. Type I interferon controls propagation of long interspersed element-1. *J Biol Chem* 2015;290(16):10191–10199. doi:10.1074/jbc.M114.612374.
- [78] Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* 2006;357(5):1383–1393. doi:10.1016/j.jmb.2006.01.089.
- [79] Payer LM, Burns KH. Transposable elements in human genetic disease. *Nat Rev Genet* 2019;20(12):760–772. doi:10.1038/s41576-019-0165-8.
- [80] Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* 2019;20(10):590–607. doi:10.1038/s41580-019-0159-6.
- [81] Baylin SB. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* 2005;2(Suppl 1):S4–S11. doi:10.1038/npcnnc0354.
- [82] Hong X, Kim ES, Guo H. Epigenetic regulation of hepatitis B virus covalently closed circular DNA: Implications for epigenetic therapy against chronic hepatitis B. *Hepatology* 2017;66(6):2066–2077. doi:10.1002/hep.29479.
- [83] Dandri M. Epigenetic modulation in chronic hepatitis B virus infection. *Semin Immunopathol* 2020;42(2):173–185. doi:10.1007/s00281-020-00780-6.
- [84] Watanabe Y, Yamamoto H, Oikawa R, Toyota M, Yamamoto M, Kokudo N, *et al*. DNA methylation at hepatitis B viral integrants is associated with methylation at flanking human genomic sequences. *Genome Res* 2015;25(3):328–337. doi:10.1101/gr.175240.114.
- [85] Zhang H, Dong P, Guo S, Tao C, Chen W, Zhao W, *et al*. Hypomethylation in HBV integration regions aids non-invasive surveillance to hepatocellular carcinoma by low-pass genome-wide bisulfite sequencing. *BMC Med* 2020;18(1):200. doi:10.1186/s12916-020-01667-x.
- [86] Summers J, Mason WS. Residual integrated viral DNA after hepadnavirus clearance by nucleoside analog therapy. *Proc Natl Acad Sci U S A* 2004;101(2):638–640. doi:10.1073/pnas.0307421100.
- [87] Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 2013;8(11):2281–2308. doi:10.1038/nprot.2013.143.
- [88] Li H, Sheng C, Wang S, Yang L, Liang Y, Huang Y, *et al*. Removal of integrated hepatitis B virus DNA using CRISPR-Cas9. *Front Cell Infect Microbiol* 2017;7:91. doi:10.3389/fcimb.2017.00091.
- [89] Ni L, Feng Y, Dong C. The advancement of immunotherapy in hepatocellular carcinoma. *Hepatoma Res* 2020;6:25. doi:10.20517/2394-5079.2020.14.
- [90] Lim CJ, Chew V. Impact of viral etiologies on the development of novel immunotherapy for hepatocellular carcinoma. *Semin Liver Dis* 2020;40(2):131–142. doi:10.1055/s-0039-3399534.
- [91] Tan AT, Yang N, Lee Krishnamoorthy T, Oei V, Chua A, Zhao X, *et al*. Use of expression profiles of HBV-DNA integrated into genomes of hepatocellular carcinoma cells to select T cells for immunotherapy. *Gastroenterology* 2019;156(6):1862–1876.e9. doi:10.1053/j.gastro.2019.01.251.
- [92] de Beijer MTA, Jansen DTSL, Dou Y, van Esch WJE, Mok JY, Maas MJP, *et al*. Discovery and selection of hepatitis B virus-derived T cell epitopes for global immunotherapy based on viral indispensability, conservation, and HLA-binding strength. *J Virol* 2020;94(7):e01663–19. doi:10.1128/JVI.01663-19.
- [93] Guidotti LG, Matzke B, Schaller H, Chisari FV. High-level hepatitis B virus replication in transgenic mice. *J Virol* 1995;69(10):6158–6169. doi:10.1128/JVI.69.10.6158-6169.1995.
- [94] Nagaya T, Nakamura T, Tokino T, Tsurimoto T, Imai M, Mayumi T, *et al*. The mode of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. *Genes Dev* 1987;1(8):773–782. doi:10.1101/gad.1.8.773.
- [95] Geng M, Xin X, Bi LQ, Zhou LT, Liu XH. Molecular mechanism of hepatitis B virus X protein function in hepatocarcinogenesis. *World J Gastroenterol* 2015;21(38):10732–10738. doi:10.3748/wjg.v21.i38.10732.
- [96] Tu H, Bonura C, Giannini C, Mouly H, Soussan P, Kew M, *et al*. Biological

- 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, Cacciola I, Saffiotti 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 CI, Yuen MF, Chan HL, Gish RG, Locarnini SA, Chavez D, *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] Bréchet C, Hadchouel M, Scotto J, Fonck M, Potet F, Vyas GN, *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, Shackel 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.
- [106] Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A* 2003;100(22):12871-12876. doi:10.1073/pnas.2135498100.
- [107] Weitzman MD, Fradet-Turcotte A. Virus DNA replication and the host DNA damage response. *Annu Rev Virol* 2018;5(1):141-164. doi:10.1146/annurev-virology-092917-043534.
- [108] Mason WS, Gill US, Litwin S, Zhou Y, Peri S, Pop O, *et al*. HBV DNA integration and clonal hepatocyte expansion in chronic hepatitis B patients considered immune tolerant. *Gastroenterology* 2016;151(5):986-998.e4. doi:10.1053/j.gastro.2016.07.012.
- [109] Yan H, Zhong G, Xu G, He W, Jing Z, Gao Z, *et al*. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife* 2012;1:e00049. doi:10.7554/eLife.00049.
- [110] Ozer A, Khaoustov VI, Mearns M, Lewis DE, Genta RM, Darlington GJ, *et al*. Effect of hepatocyte proliferation and cellular DNA synthesis on hepatitis B virus replication. *Gastroenterology* 1996;110(5):1519-1528. doi:10.1053/gast.1996.v110.pm8613059.
- [111] Halgand B, Desterke C, Rivière L, Fallot G, Sebah M, Calderaro J, *et al*. Hepatitis B virus pregenomic RNA in hepatocellular carcinoma: A nosological and prognostic determinant. *Hepatology* 2018;67(1):86-96. doi:10.1002/hep.29463.
- [112] Wang A, Wu L, Lin J, Han L, Bian J, Wu Y, *et al*. Whole-exome sequencing reveals the origin and evolution of hepato-cholangiocarcinoma. *Nat Commun* 2018;9(1):894. doi:10.1038/s41467-018-03276-y.
- [113] Qu C, Wang Y, Wang P, Chen K, Wang M, Zeng H, *et al*. Detection of early-stage hepatocellular carcinoma in asymptomatic HBsAg-seropositive individuals by liquid biopsy. *Proc Natl Acad Sci U S A* 2019;116(13):6308-6312. doi:10.1073/pnas.1819799116.
- [114] Wang YC, Li CL, Ho MC, Lin YY, Tzeng ST, Chen YJ, *et al*. Cell-free junctional DNA fragment from hepatitis B virus integration in HCC for monitoring postresection recurrence and clonality. *J Clin Oncol* 2019;37(15\_suppl):4090. doi:10.1200/JCO.2019.37.15\_suppl.4090.