

Review

Open Access



# Earliest hepatitis B virus-hepatocyte genome integration: sites, mechanism, and significance in carcinogenesis

Ranjit Chauhan, Tomasz I. Michalak

Molecular Virology and Hepatology Research Group, Division of BioMedical Science, Faculty of Medicine, Memorial University, St. John's, Newfoundland and Labrador A1B 3V6, Canada

**Correspondence to:** Dr. Ranjit Chauhan, Molecular Virology and Hepatology Research Group, Division of BioMedical Science, Faculty of Medicine, Health Science Center, Memorial University, 300 Prince Philip Dr., St. John's, Newfoundland and Labrador A1B 3V6, Canada. E-mail: chauhanranjit@gmail.com

**How to cite this article:** Chauhan R, Michalak TI. Earliest hepatitis B virus-hepatocyte genomic integration: sites, mechanism, and significance in carcinogenesis. *Hepatoma Res* 2021;7:20. <http://dx.doi.org/10.20517/2394-5079.2020.136>

**Received:** 30 Oct 2020 **First Decision:** 9 Dec 2020 **Revised:** 15 Dec 2020 **Accepted:** 15 Dec 2020 **Published:** 9 Mar 2021

**Academic Editor:** Antonio Bertoletti **Copy Editor:** Monica Wang **Production Editor:** Jing Yu

## Abstract

Hepatocellular carcinoma (HCC) is the fifth most widespread cancer responsible for one fourth of cancer-related deaths globally. Persistent infection with hepatitis B virus (HBV) remains the main cause of HCC summing up to 50% of its causative etiology. Our recent studies, supported by findings from others, uncovered that HBV and its close relative woodchuck hepatitis virus (WHV) integrate into hepatocyte genome almost immediately, hence in minutes after infection. Retrotransposons and genes with translocation potential were found to be frequent sites of HBV insertions, suggesting a mechanism of HBV DNA spread across liver genome from the earliest stages after virus invasion. Many other genes were identified as the sites of early hepadnavirus merges in human hepatocyte-like lines infected *de novo* with HBV and in natural woodchuck WHV infection model. It was uncovered that head-to-tail joins (HTJs) prevail among the earliest virus-host fusions, implying their formation via the non-homologous-end-joining (NHEJ) pathway. Overlapping homologous junctions resulting from the micro-homology-mediated-overlapping-joining (MHMOJ) were rarely detected. Formation of the initial HTJs coincided with strong induction of reactive oxygen species (ROS) and transient appearance of inducible nitric oxide (iNOS). This was accompanied by cell DNA damage and activation of the poly(ADP-ribose) polymerase 1 (PARP1)-mediated host DNA repair machinery, which may explain predominant HTJ format of the first virus-host fusions. Identification of initial integration sites and resulting alterations in hepatocyte phenotype may pave a way to discovery of reliable markers of HBV-triggered HCC, including HCC resulting from occult HBV infection. Our research strongly argues that HBV is an ultimate human carcinogen capable of initiation of a pro-oncogenic process immediately after first contact with a susceptible host.



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



**Keywords:** Hepatitis B virus, woodchuck hepatitis virus, virus-host genomic integration, virus initial integration sites, retrotransposons, virus-induced oxidative DNA damage, kinetics of DNA repair response, oncogenesis of hepatocellular carcinoma

## INTRODUCTION

Hepatitis B virus (HBV) is a pro-oncogenic DNA virus that has been identified as a leading risk factor for primary hepatocellular carcinoma (HCC) [1-3]. Once chronic, i.e., serum HBV surface antigen (HBsAg)-positive infection coinciding with chronic hepatitis type B (CHB) is established, the risk of developing liver cancer increases by many folds [4,5]. In recent years, the incidence rates of HCC have increased and new factors leading to HCC development were uncovered, including protracted non-inflammatory liver disease (NFLD). Nonetheless, CHB and, in more general terms, persistent HBV infection, remains the main cause of HCC despite availability of highly effective vaccines protecting against this virus [6-8]. On the other hand, there is no therapy capable of ultimate elimination of HBV from either symptomatically or silently infected patients [9,10]. The direct oncogenic properties of HBV are most explicitly evident in HCC developing in the absence of CHB and cirrhosis, as in cases of clinically silent HBV persistence [11,12]. This unapparent form of infection, named as occult HBV infection (OBI), is a consequence of enduring residual virus replication that is accompanied by traces of circulating HBV DNA in the absence of serum HBsAg detectable by currently available clinical tests [13]. This also is clearly apparent in woodchucks experimentally infected with woodchuck hepatitis virus (WHV), which represent overall an excellent model of molecular and immunological events and pathological outcomes encountered in HBV-infected humans [14,15]. This is well exemplified in primary occult infection (POI) induced by an intravenous (*i.v.*) injection with WHV at doses lower than 1000 virions, which can trigger HCC in the setting of seemingly entirely normal liver function and morphology, while a low level WHV replication and integrated viral DNA are detectable in the immune system and the liver [16]. With a similar frequency of about 20%, HCC develops in woodchucks recovered from an acute episode of hepatitis in which traces of infectious WHV persist for life [15]. This infection form is termed as secondary occult infection (SOI) and remains serum WHV DNA reactive at levels below 100-200 copies or virus genome equivalents (vge) per mL. It also is serum WHV surface antigen (WHsAg)-negative when evaluated by immunoassays with sensitivity compatible to that of clinical tests currently applied for HBsAg detection, although singular 22-nm WHsAg (envelope) particles and WHsAg short tubular forms can be detected in some animals by electron microscopy after ultracentrifugation [17]. In addition, antibodies to WHV core antigen (anti-WHc) and WHsAg (anti-WHs) are detectable in SOI, while POI is devoid of the WHV-specific humoral response [16,18]. In this context, the incidence of HCC in the course of serum WHsAg-positive chronic hepatitis is much greater and reaches 80%-90%. This suggests that WHV persisting in the liver at a high replication rate and in the context of prolonged hepatic inflammation significantly augments progression to HCC [19,20].

Integration of HBV DNA into human genome from the beginning of studies on molecular biology of hepadnaviruses was thought to be intimately associated with virus life cycle [21-24]. In reality, all stages of HBV infection and all *in vitro* and natural models of the infection examined so far demonstrated evidence of virus-host genomic fusions [25,26]. Hence, the integration of HBV X (HBx) and S gene sequences have been frequently identified in HBV-related HCC and associated non-tumorous hepatic tissue, and the prevailing opinion assumed that HBV-host DNA junctions are spontaneous and that they randomly occur throughout the liver genome [27-29]. Nonetheless, recent high-throughput studies indicated that certain genes might be more frequent targets for HBV integration than others, at least in hepatic tissue of patients with advanced CHB and HCC [30-34]. The oncogenic potency of HBV integrations appears to be mainly due to induction of genomic instability in hepatocyte and altered expression of individual genes with tumor suppressive or pro-oncogenic amplifying functions [35-43].

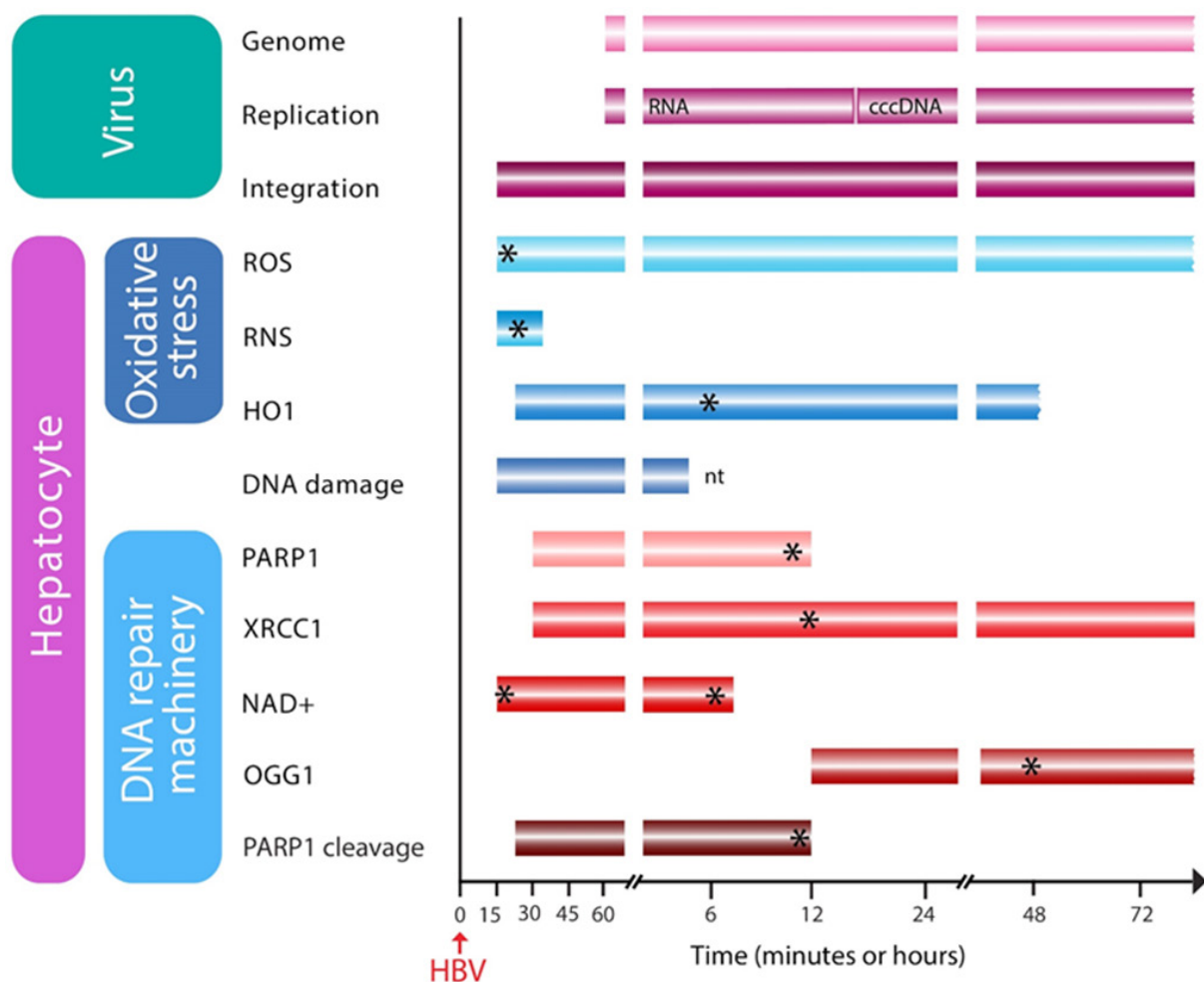
Although it became generally acknowledged that integration of HBV DNA into hepatocyte genome is an invariable consequence of HBV infection, the question, when the first virus-host DNA fusions are formed, which had been asked for decades by others and us, remained unanswered<sup>[16,44,45]</sup>. However, with recent availability of HBV-susceptible human hepatocyte-compatible cultures, including cells overexpressing sodium taurocholate co-transporting polypeptide (NTCP) serving as a HBV receptor and with access to enhanced approaches capable of detecting and sequencing genomic junctions with a high sensitivity and consistency, answering such a question became more realistic<sup>[46-50]</sup>. To identify the time kinetics of formation of the first (also called initial or earliest) HBV-hepatocyte genomic fusions, mechanisms of their creation and the nature of the host's sites involved, we explored human hepatocyte-like cells and cultured woodchuck hepatocytes susceptible to authentic (also termed as native, wild-type, or naturally occurring) HBV or WHV, as well as HepG2 cells overexpressing NTCP infected with recombinant HBV<sup>[27,51,52]</sup>. We also analyzed WHV-host integrations in the woodchuck model by examining liver biopsies collected at 1 h or 3 h after infection with WHV<sup>[51]</sup>. The principal approach to identify virus-host genomic fusions in our studies was virus genome-specific inverse-polymerase chain reaction (inv-PCR), the sensitivity and specificity of which to detect viral integrants was enhanced by nuclei acid hybridization (NAH) via Southern blot analysis with probes containing complete HBV or WHV sequences. Only amplicons displaying virus-specific signals were subsequently cloned and sequenced, and the virus-host fusions and the host's genes involved were identified with help of specialized software. The current review summarizes these studies, as well as relevant works from other groups, to provide an overall perspective on the earliest time of the appearance and the nature of the initial HBV-host genome integrations, currently known mechanisms of their formation, and their potential biological significance.

## THE EARLIEST MOLECULAR MARKERS OF HBV AND ITS REPLICATION IN HEPATOCYTES

The earliest molecular markers of HBV replication can broadly be divided into two categories, direct and indirect. HBV covalently closed circular DNA (cccDNA) and virus transcriptional templates (mRNAs) are direct indicators of active replication, while detection of HBV DNA is a sign of virus presence and its possible propagation. The appearance of viral protein in *de novo* infected cells is also considered as a sign of initiation of hepadnaviral replication.

Based on the recent *in vitro* experiments, the first appearance of HBV cccDNA was shown as early as 24 h p.i. in HepG2-NTCP-K7 cell clone and its level peaked at Day 3 and only discreetly increased during the 45-day follow-up<sup>[53]</sup>. In the same study, HBV mRNA increased from 3 days p.i. and plateaued at 6 days p.i., while viral proteins showed the same profile of increasing levels. Another study in HepG2-NTCP cells demonstrated presence of intracellular protein-free relaxed circular HBV DNA as early as in 12 h p.i., well before the detection of cccDNA at 2-3 days p.i.<sup>[54]</sup>. In yet another recent work investigating HepG2-NTCP cells, HBV DNA became detectable at 6 h p.i., whereas cccDNA at 24 h p.i.<sup>[55]</sup>. By using assay detecting HBV cccDNA by inv-PCR followed by NAH analysis of amplicons, the appearance of cccDNA was reported at 16 h p.i. in HepG2-NTCP cell line by another group<sup>[56]</sup>. Furthermore, it has also been recently shown that 3.5-Kb HBV mRNA can be detected at 18 h p.i. when examining *de novo* infected primary human hepatocytes and HepG2-NTCP-A3 cell clone<sup>[57]</sup>. In our study of the timeline of formation of the earliest HBV-host DNA integration sites in HepaRG cells investigated applying high sensitivity PCR/NAH-based assays, HBV DNA and its RNA transcripts became detectable from 1 h p.i., while HBV cccDNA from three days p.i. onwards<sup>[51]</sup> [Figure 1].

The model infections with other hepadnaviruses showed comparable timelines of the first detection of viral DNA and its replication intermediates in the early stages of infection to those mentioned above. Thus, infection with duck hepatitis B virus (DHBV) of duck hepatocytes documented generation of virus pre-genomic RNA from 12 h p.i., which peaked at 20 h p.i., and was followed by detection of DHBV core protein 12 h later<sup>[58]</sup>. In livers of Peking ducklings inoculated with DHBV, the appearance of supercoiled



**Figure 1.** Schematic presentation of the detection and evolution of markers of virus infection and its genomic integration, hepatocyte DNA damage, and indicators of hepatocyte oxidative stress and activity of DNA repair machinery in the first 72 h after contact with infectious HBV or WHV. The graphs are based on combined results from HBV and WHV infections in hepatocyte-compatible cells and from woodchucks experimentally infected with wild-type WHV, as detailed in the text. Star on bar represents the time at which peak expression or activity was observed. HBV: hepatitis B virus; WHV: woodchuck hepatitis virus; ROS: reactive oxygen species; RNS: reactive nitrogen species; HO1: heme oxygenase-1; PARP1: poly(ADP-ribose) polymerase 1; XRCC1: X-ray repair cross-complementing protein 1; NAD<sup>+</sup>: nicotinamide adenine dinucleotide; OGG1: 8-oxoguanidine DNA glycosylase 1; nt: not tested beyond the time point indicated.

virus DNA was observed at 6 h p.i., whereas that of viral 3.5-, 2.7-, and 2.4-Kb mRNA transcripts at 12 h p.i., which was shortly after followed by detection of single-stranded DHBV DNA. In addition, prominent increases in DHBV RNA in liver tissue was reported between 12 h and 72h p.i.<sup>[59]</sup>. In the early stage of WHV infection, WHV DNA and mRNA became detectable in liver biopsies obtained at 1 or 3 h p.i. using PCR/NAH assays, which remained virus WHV cccDNA negative at these time points. WHV cccDNA became detectable in the subsequent liver biopsies collected at six weeks p.i.<sup>[51,60]</sup>. In woodchuck WCM260 line derived from primary hepatocytes isolated from a healthy animal<sup>[61,62]</sup>, quantifiable levels of WHV DNA became detectable from 6 h p.i., while its unquantifiable signals were seen from 30 min p.i. by real-time PCR (qPCR) and NAH analysis of the resulting amplicons<sup>[52]</sup>.

### TIME OF THE FIRST APPEARANCE OF HEPADNAVIRUS-HOST GENOMIC MERGES

Definition of time of the appearance of a given virus-host junction in our studies was based on the time in minutes (min), hours (h), or days which lapsed between the first contact of cell or animal with

virus inoculum and the detection of junction. Based on this, the integrations were categorized to three groups following the refined scheme previously applied<sup>[51]</sup>. Thus, the junctions found up to 24 h p.i. were designated as very early integration site (VEIS), those after 24 h and until 72 h p.i. as early integration sites (EIS), and those beyond 72 h p.i. as late or not-early integration sites (NEIS). In addition, the first or earliest virus-host genomic fusions were also called initial integration sites (IIS).

Our original study explored human hepatocyte-like HepaRG cells *de novo* infected with wild-type HBV<sup>[51]</sup>. The IIS became detectable at 1 h after exposure to virus and HBV DNA integrations into five different host genes were identified [Table 1]. At the same time, HBV DNA and its transcripts became detectable<sup>[51]</sup>. Other time points investigated in this study were three and seven days, and two, four, and seven weeks p.i. No integration signals were detected up to 1 h p.i., as well as in control cells collected at time 0 or those after mock infection with normal human plasma (NHP). Overall, 9 HBV-host DNA fusions were classified as VEIS, 5 as EIS, and 11 as NEIS. There was a weak trend towards an increase in the number of HBV integrations over the time examined. In the same study, woodchuck liver biopsies collected prior to *i.v.* infection with WHV and at 1 h or 3 h p.i., and at 6 weeks p.i. were examined for virus-host genomic junctions using WHV-specific inv-PCR/NAH. The biopsies collected at 1 h p.i. revealed the presence of WHV DNA fusions with multiple ( $n = 8$ ) host genes spreading across different chromosomes [Table 1]. Overall, there were 10 sites classified as VEIS in biopsies acquired at 1 h or 3 h p.i. and 7 as NEIS in biopsies acquired at six weeks p.i. Thus, the time of the appearance of the first virus-host merges was the same in cultured HepaRG cells infected with HBV and in woodchucks infected with WHV.

To advance characterization of the human genomic sites forming initial fusions with HBV, we further examined HBV integrations into genome of HepG2-C4 cell clone stably transfected with NTCP (HepG2-NTCP-C4) in the subsequent study<sup>[27]</sup>. These cells have shown a high susceptibility to HBV infection and capability of an efficient production of infectious HBV virions<sup>[27]</sup>. As an inoculum, HBV genotype D secreted by HepG2.2.15.7 cells was used. Infected HepG2-NTCP-C4 cells were investigated for HBV-host junctions at 15 min and 30 min, 3 h and 24 h, and 13 days p.i. The initial virus insertional sites (i.e., IIS) into two different genes were identified 30 min after exposure to HBV [Table 1 and Figure 1]. In general, from the 15 integration sites detected across nine chromosomes, six were identified as VEIS and nine as NEIS [Table 1]. This study for the first time showed that the first fusions of HBV DNA into host's genome could occur as early as 30 min after contact with virus. They also suggested that *in vitro* infection of highly prone cells by recombinant HBV could be superior over natural infection with wild-type virus in enabling hepadnavirus-host genomic merges.

Another research group also investigated timeline of HBV integration<sup>[63]</sup>. The *in vitro* models utilizing recombinant HBV and primary human hepatocytes or hepatocyte-like HepaRG, HepG2, and Huh7 cells transfected with NTCP provided evidence of HBV integration into numerous sites of the host's genome between one and nine days after infection. The first junctions were detected at the first time point investigated in Huh7-NTCP cells, i.e., at 24 h p.i. In addition, the authors used the HBV entry inhibitor Myrcludex B (MyrB), which likely targets HBV hepatocyte NTCP receptor, and showed that HBV DNA integration can be blocked at this early time point, however interpretation was based on agarose gel visualization of the inv-PCR products<sup>[63]</sup>.

Another of our studies examined the timeline of the appearance of the earliest hepadnavirus-host integrations in cultured woodchuck hepatocytes infected *de novo* with wild-type WHV<sup>[52]</sup>. Although the main purpose of this particular study was to recognize a mechanism involved in the formation of the initial virus-host fusions, we also examined the timeframe when these fusions were for the first time assembled. For this study, hepatocytes were examined between 15 min and 72 h after exposure to WHV, while the presence of virus-host integrations was analyzed at 15 min, 30 min, and 1 h p.i. to focus our efforts on the



**Table 1. *In vitro* and *in vivo* data on the initial and very early hepadnavirus DNA integration into host genomic sequences detected up to 24 h after infection with HBV or WHV**

Cell target/ Animal	Time post infection	Infecting virus	Integrated virus sequence (nt position)	Joined host sequence (nt position)	Host gene	Type of virus- host junction	Number of clones or sequencing approach	Ref.			
HepaRG	1 h	HBV	1616-1659	131410505-131410385	NTM	OIJ	8	[51]			
			1720-1603	420-437	ANP32E	HTJ	1				
			1639-1829	98457063-98457081	S3A-26	HTJ	1				
			1647-1720	111339527-111339567	ANK3	HTJ	1				
			1662-1603	102521074 102521122	FGF14	HTJ	1				
	3 h		1645-1700	27-54	UI	HTJ	1				
			1603-1624	86038392-86038596	FLRT2/LINE2	HTJ	1				
	24 h		1603-1623	7874512-7874653	LINE1	HTJ	8				
			1370-1423	60109931-60109947	RNY-1	HTJ	1				
			1603-1624	86038392-86038596	FLRT2/LINE2	HTJ	1				
			1246-1390	97732036-97732056	DPYD	HTJ	1				
			3200-3256	95720503-95720470	MAML2	HTJ	1				
Woodchuck	1 h	WHV	1853-1876	UI	UI	HTJ	1	[51]			
			2785-2934	84783482-84783518	EFTUD1P1	HTJ	1				
			1913-1853	69784039- 69784157	AAK1	HTJ	2				
			1913-1853	83777393-83777412	KIAA1117	HTJ	1				
			1910-1853	39989149-39989169	LPIN3	HTJ	1				
			3030-3009	178193731-178193748	PHACTR3	HTJ	1				
			3250-3308	UI	AAK1	HTJ	1				
	3 h		3251-3308	UI	UI	HTJ	1				
			3251-3308	UI	UI	HTJ	1				
			1729-1821	NA	LINE1	HTJ	Direct sequencing		[63]		
			1729-1821	140957103	RP11-63E9.1	HTJ	Direct sequencing				
			1729-1821	29304718	SNX29P2	HTJ	Direct sequencing				
Huh-7-NTCP	24 h	HBV	1729-1821	11892442	AC116618.1	HTJ	Direct sequencing				
			1647-1774	16583566-16583474	NBPF-1	HTJ	4	[27]			
			1647-1687	87153566-87154156	SINE	HTJ	17				
			1647-1725	82086308-82086501	THE1B-LTR	HTJ	7				
			1866-1945	51769849-51769877	PRKG-1	HTJ	10				
			1647-1820	119935846-119935884	PRR-16	HTJ	10				
			1647-1792	35712645-35713126	RunX1	OIJ	2				
			WCM260	15 min	WHV	1676-1809	72 bp	UI	HTJ	1	[52]
						3010-3028	197 bp	UI	HTJ	1	
						1360-1523	67 bp	UI	HTJ	1	
						1894-1934	98 bp	UI	HTJ	1	
						1894-1934	98 bp	UI	HTJ	1	
30 min	1676-1809	285 bp		UI		HTJ	1				
	2980-2895	23 bp		UI		HTJ	1				
	1647-1667	112 bp		UI		HTJ	1				
	3261-3308	134 bp		UI		HTJ	1				
	3320-3230	123 bp		UI		HTJ	1				
1 h	2904-2975	94 bp		UI		HTJ	1				
	3317-3227	199 bp		OR6C66P		HTJ	1				

HBV: hepatitis B virus; WHV: woodchuck hepatitis virus; min: minutes, h: hour; WCM260: hepatocyte cell line from a woodchuck male; nt: nucleotide; bp: length in base pairs of unidentified host gene sequence; OIJ: overlapping homologous junction; HTJ: head-to-tail junction; NTM: neurotrimin; ANP32E: acidic (leucine rich) nuclear phosphoprotein 32 family, member E; S3A-26: ribosomal protein S3A pseudogene 26; ANK3: ankyrin3; FGF14: fibroblast growth factor 14; UI: unidentified sequence; FLRT2: fibronectin leucine rich transmembrane protein; LINE2: long interspersed nuclear element-2; DPYD: dihydropyrimidine dehydrogenase; RNY-1: Ro-associated Y pseudogene 1; L1: long interspersed nuclear element-1; long interspersed nuclear element-2; MAML2: mastermind-like 2; EFTUD1P1: elongation factor Tu GTP binding domain containing 1 pseudogene; AAK1: AP-1 associated kinase; LPIN3: lipin-3; KIAA1117: Kazusa cDNA 1117; PHACTR3: phosphatase and actin regulator 3; RP11-63E9.1: Long non-coding RNA gene RP11-63E9.1; SNX29P2: sorting nexin 29 pseudogene-2; AC116618.1: hominid apparent LTR retrotransposon; NBPF-1: neuroblastoma breakpoint family member-1; SINE: short-interspersed nuclear element; THE1B-LTR: mammalian apparent LTR retrotransposon; PRKG-1: protein kinase cGMP-dependent type-1; PRR-16: proline rich-16; RunX1: runt-related transcription factor-1; OR6C66P: olfactory receptor family 6 subfamily C member 66 pseudogene.

events occurring at the time of initiation of infection. As controls, WCM-260 cells not exposed to WHV (time 0) and those incubated with normal woodchuck plasma (NWP) were examined. In this infection system, the IIS were detected 15 min p.i., and virus fusions were identified in four different host genomic sequences [Figure 1]. Furthermore, four other junctions were detected at 30 min p.i. and three more at 1 h p.i. By definition, all of them belonged to the VEIS category. WCM260 cells not exposed to WHV (time 0) and those subjected to mock infection did not show integration signals. These results further modified our perception about the time required for hepadnaviral DNA to integrate into the host's genome and strongly suggested that hepadnavirus DNA integration occurs immediately after virus entry into cell, which appears even before initiation of its replication [Figure 1]. This extremely short time period after the first contact with virus was no longer surprising when kinetics of virus-induced oxidative stress and DNA repair machinery, determined in the same infection system, became known<sup>[52]</sup>.

## HBV EARLY IN INFECTION FREQUENTLY INTEGRATES INTO HOST NON-CODING DNA TRANSPOSABLE ELEMENTS

In recent studies, we identified that at the very early stages of infection HBV frequently integrates into human mobile genetic elements containing repetitive non-coding genomic sequences, such as retrotransposon and transposon elements, and into genes with translocation potential<sup>[27,51]</sup>. Thus, in the HBV-HepaRG cell infection model, HBV DNA integrations with or in close proximity to LINE1 (L1) and LINE2 (L2) were identified at 3 and 24 h p.i.<sup>[51]</sup>. These merges were sometimes evident in the majority of clones derived from the particular time point p.i., as for HBV-LINE1 fusion identified at 24 h p.i.<sup>[51]</sup> HBV DNA junctions with LINE1 or LINE2 were also detected in later time points p.i. in this model. In another study, HBV junctions with LINE1 were also detected in Huh7-NTCP cells at 24 h p.i.<sup>[63]</sup>. It is of note that over one-fifth of the human genome is comprised by DNA elements belonging to the family of LINE1, LINE2, or LINE3 (i.e., chicken repeat-1, CR1)<sup>[64]</sup>. Among these three families, LINE1 is the most abundant and represented by an estimated 500,000 copies, which overall constitutes about 18% of the human genome, while LINE2 and LINE3 make up 3% and 0.3% of the genome, respectively<sup>[65,66]</sup>. LINE1 is mostly autonomous and displays endonuclease and reverse transcriptase activity, and it transposes through the mechanism termed as target-primed retrotransposition (TPRT)<sup>[67]</sup>. LINE2 is a fossil representative and is commonly located between intronic regions of the human genome<sup>[68,69]</sup>.

In addition, HBV fusions with human satellite II DNA (HSAT-II), another retrotransposable element, were detected at later time points, i.e., 3 days and 14 days p.i. Representation of this merge was particularly strong at 14 days, as it was identified in 12 clones. The integration of HBV DNA with HSAT-II has not been described before, however HBV merges with the HSAT-III sequence were reported in a hepatoma cell line and in HCC tissue<sup>[51,70,71]</sup>. Overall, taking into account all HBV insertional sites detected in HepaRG infected with native HBV, 5 (23%) of the 22 unique integration sites uncovered were transposonable elements. Considering the number of clones carrying junctions with tandemly repeating non-coding DNA sequences, they represented 46% of all clones with virus–host merges detected (35/76) and 37.5% of those in which integration sites were classified as VEIS (9/24).

In the more efficient infection of HepG2-NTCP-C4 cells with HBV, as judged by the twice-shorter time of the appearance of the first HBV-host genomic fusions, one of two IIS detected at 30 min p.i. was a junction with short-interspersed nuclear element (SINE). The HBV-SINE fusion was represented in 17 (85%) of 21 clones carrying IIS. SINE is a retrotransposon belonging to the non-long terminal repetitive (non-LTR) category that is abundant in human genome and may significantly influence its size<sup>[72,73]</sup>. It is also expected that SINE advances oncogenic transformation<sup>[74]</sup>. Among VEIS identified in HepG2-NTCP-C4 cells, there was also HBV fusion with the retrotransposon known as the mammalian apparent retrotransposon long terminal repetitive (THE-1B-LTR) element, belonging to a mammalian apparent LTR retrotransposon

(MaLR) family<sup>[75]</sup>. This junction was the only one detected at 1 h p.i. but was well represented since seven separate clones confirmed this sequence<sup>[27]</sup>. THE-1B-LTR appears to play an important role in the development of non-Hodgkin's lymphoma<sup>[76]</sup>. Since HBV DNA merged with THE-1B-LTR encompassed virus enhancer II (Enh-II), a possible pathogenic relevance of this fusion might lay in modulations of the retrotransposon activity and hepatocyte functions<sup>[27]</sup>. At one late-time point after infection, i.e., 13 days p.i., either direct or indirect merges of HBV DNA with three other elements of the retrotransposon or transposon class were identified in HepG2-NTCP-C4 cells. Hence, HBV DNA was integrated with hobo activator-18 *Salmo salar* long terminal repeat (hAT-18-SsA). There are about seven dozen hAT elements in humans which transpose through DNA-DNA fusions; in contrast, retrotransposons usually rearrange the genome via DNA-RNA merges<sup>[77]</sup>. Interestingly, hAT-18-Ssa was joined by non-coding sequence of chromosome-2 (CH-2) and that by medium reiterated frequency repeat 5B (MER-5B), another non-coding sequence of the transposon class. This resulted in a complex structure formed by HBV DNA and host sequence trimera. It is of interest to note that MER-5B is known to control expression of alpha fetoprotein (AFP) gene, the protein of which is plentifully displayed in human fetal liver<sup>[78]</sup>. It has also been shown that retroviral insertions may cause protracted expression of AFP as well as H19 in liver<sup>[79]</sup>. In this regard, we documented previously WHV DNA integration into woodchuck H19 gene<sup>[16]</sup>. Taken together, this may suggest a link between hepadnaviral DNA insertions and elevated AFP levels. The molecular foundation of this possible relation would require future investigations. In addition, HBV DNA-LINE2 merge was yet another fusion with retrotransposon identified among NEIS in HepG2-NTCP-C4 cells. In general, among the 15 HBV-host DNA integration sites identified in total, five (33%) were merges with transposable elements. Clones carrying these fusions comprised 41% of all clones with virus-host junctions (32/78) and 49% of those in which merged sequences were classified as VEIS (24/49). If any comparison between HBV integration profiles in HepaRG and HepG2-NTCP-C4 cells could be made, the data indicate that HBV infection in HepG2-NTCP-C4 cells was characterized by a greater proportion of virus fusions with transposable elements (33% vs. 23%) and by a larger proportion of clones carrying these fusions among the clones with integration sites identified as VEIS (49% vs. 37.5%). Considering the nature of transposable elements joined with HBV, only the LINE retrotransposon family was identified in both cell lines.

However, when the above results were compared with those from another study investigating HBV integration after *de novo* infection of HepG2-NTCP or Huh7-NTCP cells<sup>[63]</sup>, four of the same or similar retrotransposon or transposon elements were found. These were SINE detected at five and seven days, THE-1B-related THE-Int at five days, MER-5B similar MER52D/41A/90A/4E1/4A at seven days, and LINE1 between one and seven days p.i. In general, there was a good agreement between the findings from infection systems utilizing cells expressing NTCP and as inocula either patient-derived or recombinant HBV, which jointly further ascertained authenticity of the findings.

## HEPATOCTE GENES TARGETED FOR INTEGRATION BY HBV AND WHV IN THE FIRST 24 H AFTER INFECTION

Based on the clonal sequencing analysis or direct sequencing of virus-host junctions, many different host genes, other than genomic repetitive elements, were found to be insertional sites with which HBV DNA initially (i.e., IIS) or very early (i.e., VEIS) post-infection has fused<sup>[27,51,63]</sup>. Thus, in our study of HepaRG cells infected with authentic HBV, viral integrations into five different genes were detected at 1 h p.i. These genes were neurotrimin (NTM) located at chromosome (Ch)-11q25, acidic (leucine rich) nuclear phosphoprotein 32 family (ANP32E) on Ch-1q21.1, ribosomal protein S3A pseudogene 26 (S3A-26) on Ch-2q22.1, ankyrin3 (ANK3) on Ch-10q21.2, and fibroblast growth factor 14 (FGF14) on Ch-13q33.2 [Table 1]. Two other genes, dihydropyrimidine dehydrogenase (DPYD) on Ch-1q21.3 and Ro-associated Y pseudogene (RNY-1) at the q36.1 locus of Ch-7, were detected at 24 h p.i.<sup>[51]</sup>. Interestingly, the profiles of these very early integrations were much different after infection with two HBV inocula. Hence, inoculum



containing HBV genotype C produced initial virus-host fusions with just two genes, but their sequences were displayed in multiple clones. The two sites identified were NTM encoding a neuronal adhesion molecule at 1 h p.i. and retrotransposon LINE1, mentioned in the preceding section, at 24 h p.i. [Table 1]. Contrastingly, the second inoculum that carried HBV genotype A generated junctions with several host genes, the sequences of which were detectable in singular clones only [Table 1]. This observation potentially represents a valuable finding suggesting that the virus itself could predetermine the pattern of virus-host fusions.

Huh7-NTCP cells infected with recombinant HBV for 24 h also demonstrated virus-host integrations into more than one of the host genes [Table 1]. Hence, HBV DNA fusions to long non-coding RNA gene RP11-63E9.1 (RP11-63E9.1), sorting nexin 29 pseudogene-2 (SNX29P2), and homo sapiens BAC clone RP11-98L17 (AC116618.1) were identified and their existence confirmed by direct Sanger sequencing<sup>[63]</sup>. In another study in which HepG2-NTCP-C4 cells infected with HBV were examined, virus junctions with multiple host genes were detected within 24 h p.i. and almost all of them were identified in multiple clones sequenced<sup>[27]</sup>. The first fusions became detectable in 30 min after exposure to virus and HBV DNA insertions into neuroblastoma breakpoint family member-1 (NBPF-1) gene on Ch-1p36.13 and retrotransposon SINE at the q23.2 locus of Ch-10 were detected [Table 1]. Parenthetically, NBPF-1 is a pseudogene encoding a tumor suppressor for neuroblastoma<sup>[80]</sup>. Other HBV DNA insertions into protein kinase cGMP-dependent type 1 (PRKG1) gene located at Ch-10q11.23 and into protein rich 16 (PRR16) gene on Ch-5q23.1 were found at 3 h p.i. and into run-related transcription factor 1 (RunX1) gene on Ch-21q22.12 at 24 h p.i. PRKG1 is a cyclic GMP-dependent protein kinase that regulates cell signaling and growth mainly in skeletal muscle and neuronal cells<sup>[81]</sup> and RunX1 plays a role in hematopoiesis and possibly in the pathogenesis of acute myeloid leukemia and HCC<sup>[82]</sup>.

Furthermore, three of four biopsies obtained at 1 or 3 h p.i. from woodchucks *i.v.* infected with wild-type WHV also showed viral DNA fusions into multiple host sequences<sup>[51]</sup>. It was possible to assign woodchuck genes and allocate chromosome locus for some of them. The host genes with WHV DNA insertions were mastermind-like 2 (MAML2), elongation factor Tu GTP binding domain containing 1 pseudogene (EFTUD1P1), AP-1 associated kinase (AAK1), Kazusa cDNA 1117 (KIAA1117), lipin-3 (LPIN3), and phosphatase and actin regulator 3 (PHACTR3).

Finally, woodchuck WCM260 hepatocyte line infected with wild-type WHV demonstrated virus-host joints which were identifiable from 15 min p.i. by inv-PCR/NAH followed by cloning of the WHV reactive amplicons and sequencing of the clones<sup>[52]</sup>. In total, 12 clones carrying 11 unique WHV-host DNA fusions were detected between 15 and 60 min p.i. Five of the junctions were with WHV X gene and the other five with WHV preS region sequence. Among the host sequences fused, one was identified as woodchuck olfactory receptor family 6 subfamily C member 66 pseudogene (OR6C66P) [Table 1].

Of all the host's genes mentioned in the section above [Table 1], NTM<sup>[83]</sup>, ANP32E<sup>[84,85]</sup>, S3A<sup>[86]</sup>, and FGF14<sup>[87]</sup> were found in HBV-infected HepaRG cells at 1 h p.i.<sup>[51]</sup>; MAML2<sup>[88]</sup> and PHACTR3<sup>[32]</sup> were detected at 1 h p.i. in liver biopsies of WHV-infected woodchucks<sup>[51]</sup>; and NBPF-1<sup>[89]</sup> and RunX1<sup>[90]</sup> in HBV-infected HepG2-NTCP-C4 cells at 30 min or 24 h p.i.<sup>[27]</sup> were identified as those to be directly or indirectly linked to cellular gene translocation. Thus, while excluding non-coding retrotransposable and transposable elements detected up to 24 h p.i., such as FLRT2/L2, LINE1, RNY-1, RP11-63E9.1, SNX29P2, AC116618.1, SINE, THE1B-LTR, and 11 unidentified woodchuck genomic elements [Table 1], the analysis showed that 8 (50%) of the remaining 16 sequences had predicted translocation potential. Taking into account both non-coding DNA transposable elements, which by their nature are mobile and prone to translocation, and the genes with the predicted translocation potential that were found in the studies discussed, the great majority of the identifiable host sequences joined by HBV or WHV were those that could translocate genomic and

inserted exogenous sequences across hepatocyte genome. Their complete list included NTM, ANP32E, S3A-26, FGF14, FLRT2/LINE2, LINE1, MAML2, PHACTR3, NBPF-1, SINE, THE1B-LTR, and RunX1, and they accounted for 12 of 19 (63%) sequences identified as hepadnaviral insertional sites. This further strengthened the hypothesis proposed in our first study on this subject<sup>[51]</sup> that HBV can engage from the beginning of infection mobile genetic elements, including genes with translocation capabilities, to prompt pro-oncogenic perturbations throughout the host genome which may compromise genome overall stability and either augment or silence expression of individual genes important to the development of HCC.

## CHROMATIN MARKS ON HOST GENOMIC SEQUENCES TARGETED BY HBV INTEGRATION

Whether the hepadnaviral sequences integrated into the host genome will be able to transcribe or not depends on the characteristics of the sites with which virus DNA merged. By identifying the presence or absence of epigenetic modifications on the joined host sequences, transcriptional activity or latency of the integration sites can be predicted. In one of our studies investigating the kinetics of the formation and nature of the earliest HBV-host genome junctions<sup>[27]</sup>, *in silico* analysis was performed to identify histone chromatin signatures H3K4me3 (Tri-methylation of lysine 4 on histone H3) and H3K27ac (histone H3 lysine 37 demethylase), where H3K4 methylation mark suggests transcriptionally repressed sites, whereas presence of H3K27 acetylation mark is attributable to transcriptionally active state<sup>[91,92]</sup>. In addition, we tracked for CCCTC-binding factor (CTCF) clusters linked to insulator activity, enhancer of zeste homolog 2 (EZH2) sequences playing a main role in methylation by activating methyl groups, and for the presence of DNase binding regions<sup>[93-95]</sup>. In the HBV infection model in HepG2-NTCP-C4 cells, in which six distinct virus-host integration sites were identified in the first 24 h p.i., two of the sites encompassing host SINE and NBPF-1 sequences were detected at 30 min p.i. and classified as IIS [Table 1]. The analysis of these two sites revealed commonalities in the state of H3K4me3, EZH2, and DNase marks, however there were differences in the status of H3K27ac. Thus, in contrast to NBPF-1, SINE sequence exhibited presence of acetylation mark H3K27ac, suggesting that the HBV-SINE DNA merge could be in the transcriptionally active state. On the remaining four integration sites detected between 1 h and 24 h p.i., H3K4me3 mark was detected while acetylation mark H3K27Ac was absent. Accordingly, this profile implied that the sequences forming junctions with HBV DNA during this time were unlikely transcriptionally inactive.

CTCF has important regulatory functions including long-range gene activation, insulation, imprinting, and cell differentiation<sup>[96,97]</sup>. Interestingly, CTCF binding motifs were found on almost all host sequences detected up to 24 h p.i., except for PRR-16, in the study mentioned above [Table 1]. Regarding DNase hypersensitive motifs, the data show their absence on all sequences forming sites classified as VEIS, as well as on seven of eight sequences merged with HBV identified at 13 days p.i. in the same study<sup>[27]</sup>. These results suggest that the sequence regions investigated were likely in closed conformation restricting chromatin modifications. With the advancement of recognition of functional significance of chromatin marks, the prominent role of polycomb repressive complex 2 (PRC2) has been uncovered<sup>[98]</sup>. In this regard, we analyzed presence of EZH2 mark, as EZH2 is a key catalytic subunit of PRC2 involved in gene methylation while activating methyl group<sup>[95]</sup>. Although we could not find EZH2 binding sites at the initial stage of HBV DNA integration, i.e., at 30 min p.i., we found them in three of four host sequences merged with HBV DNA detected between 1 h and 24 h p.i. and in seven of eight sequences detected at 13 days p.i. The recent findings progressively illuminate a role of key chromatin marks that may also govern HBV driven initiation of hepatocyte oncogenic transformation culminating in liver cancer.

Regarding the above, some advances have been made in recognition of the methylation status of integrated HBV sequences. One of the studies found presence of methylation on the integrated HBV DNA in SNU-398 cells derived from HBV-associated HCC<sup>[99]</sup>. However, this mark was not found on the integrated HBV sequence in another Hep3B cell line which also expressed HBV envelope proteins<sup>[99]</sup>. Methylation of the integrated HBV DNA might be related to the methylation state of the adjacent host sequences. To

investigate this, a study used a next-generation sequencing-based method for structural methylation analysis of integrated viral genomes<sup>[100]</sup>. It was uncovered that integrated HBV DNA is significantly methylated when the fused host genomic sites are already highly methylated. However, if HBV integrates into the unmethylated sites, such as promoters and enhancers, integrated HBV DNA do remain unmethylated<sup>[100]</sup>.

### HBV AND WHV DNA BREAKING POINTS AT WHICH VIRUS-HOST JUNCTIONS HAS FORMED

Considering the sites of HBV and WHV genomes, which formed junctions in the first 24 h p.i., the sequences of the virus-host DNA fusions were analyzed to determine where the breaking points in viral DNA occurred and if they were clustered in particular regions. However, there is not yet a study based on the whole hepadnavirus genome analysis, and only some parts of hepadnaviral sequence, particularly the X gene, were investigated using sensitive inv-PCR-based methods. Therefore, it is likely that other regions where virus DNA breaks emerge capable of forming fusions with host sequences will be found when more thoughtful analyses of the earliest stages of infection become feasible.

Using the HBV-HepaRG cell infection system<sup>[51]</sup>, the majority of HBV-host junctions were formed between the HBx DNA gene sequence comprising the enhancer-II (Enh-II) and basal core promoter (BCP) regions located between nucleotides (nt) 1246-1829 (nucleotide positions according to HBV DNA GeBank X70185 and AB033556 sequences)<sup>[51]</sup>. Within this fragment, 91.6% (22/24) of all DNA breaking points identified within the HBx 1603-1829 sequence were found. Further, 25% (6/24) were confined to Enh-II between nts 1659 and 1739, while 41.6% (10/24) to BCP between nts 1764 and 1829. This showed that 66% (16/24) DNA breaks were within the HBx gene sequence overlapping HBV regulatory elements BCP and Enh-II and hence these elements appeared most prone to breaks that formed junctions with host genome. The breaking points in the BCP were further divided into two clusters. The first cluster spanned nts 1764-1808 that contained the HBV TATA-like binding sequences (TA2-TA4) between nts 1758 and 1795 and pre-core mRNA initiation sites between nts 1788 and 1795. Six breaking points were identified in this region, including one with six hits at position 1764. The second cluster encompassed nts 1816-1829 and contained the HBV pre-genomic RNA initiation site at nt 1818<sup>[101]</sup>, including one HBV DNA breaking point found at this position in this sequence. In addition to the mentioned findings, this study showed that enumeration of the HBV DNA breaking points according to the virus genomic regions is feasible as well as validated the methodology used, which included clonal sequencing of the inv-PCR/NAH products displaying HBV-specific signals.

By analyzing the data from Huh7-NTCP cells infected by HBV, which were reported by another group<sup>[63]</sup>, we found that of the four VEIS identified at 24 h p.i. the HBV fragment spanning nts 1790-1809 was fused with retrotransposon LINE1, HBV 1793-1822 nt sequence with host RP11-63E9.1, HBV 1760-1790 nt sequence with host SNX29P2, and HBV fragment between nts 1698 and 1726 with host AC116618.1 [Table 1]. The cumulative size of the HBV fragment with DNA breaking points engaged in joining with Huh7-NTCP cell genome was nts 1698-1822. Furthermore, in our study applying HepG2-NTCP-C4 cells as HBV infection targets<sup>[27]</sup>, six different host genes or genetic elements were fused with HBV from 30 min to 24 h p.i. [Table 1]. The HBV DNA breaking points forming these junctions were located between nts 1647 and 1945. All nucleotide positions were enumerated according to the same HBV reference, as cited previously<sup>[51]</sup>. Considering the above findings, the great majority of the DNA breaking points found fused with Huh7-NTCP and HepG2-NTCP-C4 cell genomes were within the same HBV genomic region as the fusions identified in HepaRG cells. Specifically, HBV Enh-II and BCP sequences appeared to be most prone to the DNA breakages.

In the woodchuck infection model, WHV DNA breaks engaged in formation of junctions with host genomic sequences in the first 3 h p.i. were predominantly located in the WHx gene in the sequence

between nts 1853 and 1876 containing the virus BCP region<sup>[51]</sup>. This WHV DNA fragment created fusions with five different VEIS. However, since we also applied invPCR/NAH with primers specific for WHV preS genomic region, it was possible to identify host fusions with nucleotides within this and the downstream P gene sequence. The data show the WHV DNA preS breaking points forming fusions with the host's sites classified as VEIS were located between nts 3300 and 3309 (nucleotide positions enumerated according to GenBank sequence AY334075).

In yet another model, an *in vitro* infection of WCM260 hepatocyte line was applied to recognize the mechanism of hepadnaviral integration in the initial infection. In this study, WHV DNA integrations were uncovered as early as in 15 min p.i. The WHV DNA breaking points were found in the WHx gene sequence of nts 1360-1934 and in the sequences predominantly spanning the preS1 region of nts 2904-3308.

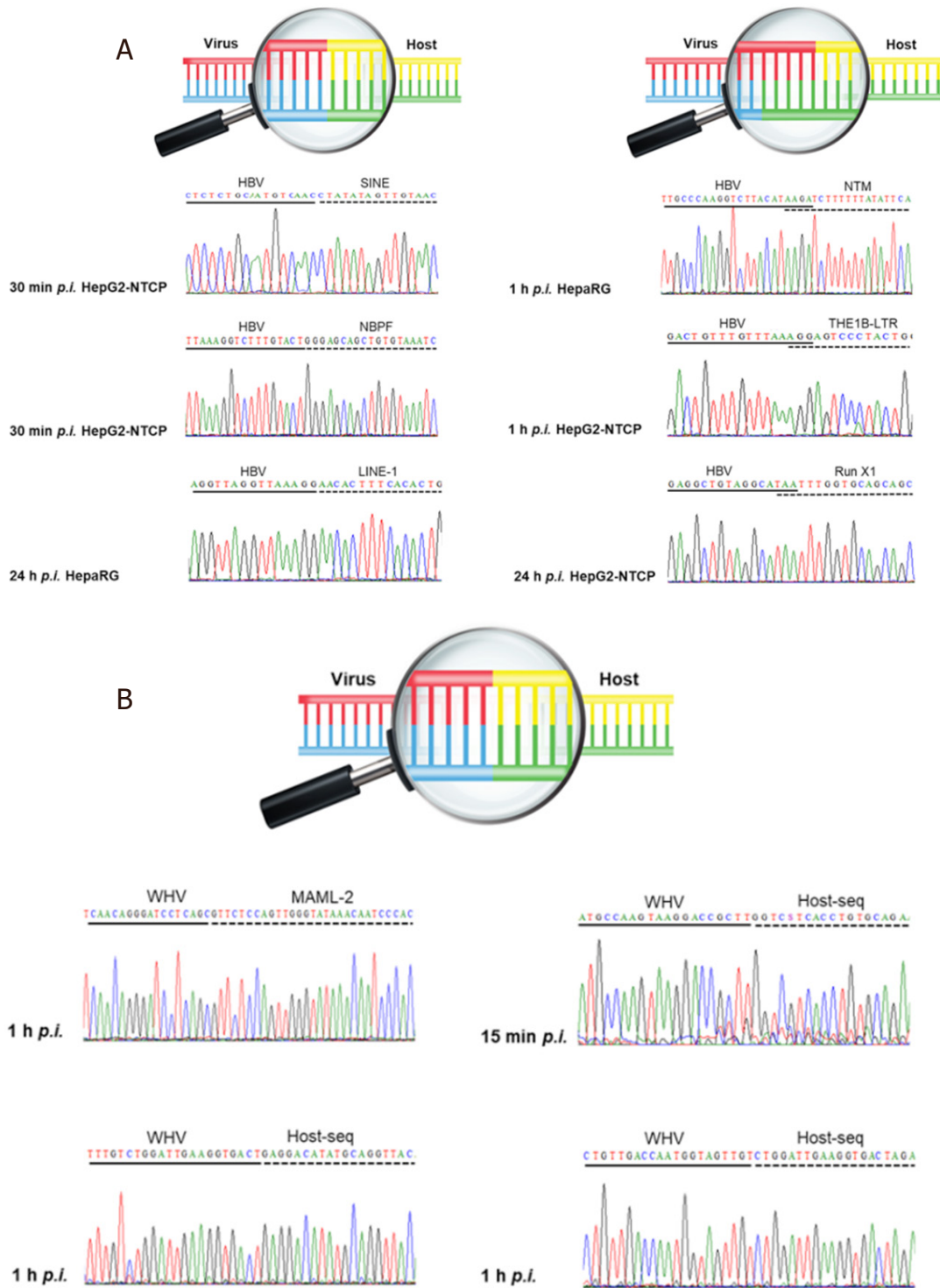
## MOLECULAR FORMATS OF THE VERY EARLY HEPADNAVIRUS-HOST GENOMIC JUNCTIONS

The recent data from analyses at the single nucleotide resolution level demonstrate that the HBV-host or WHV-host DNA fusions created in the first 24 h p.i. were mostly of the head-to-tail (HTJ) type, while overlapping homologous junctions (OHJ), also termed as micro-homology overlapping junctions (MHOJ), were rarely detected [Figure 2]<sup>[27,51,52,63]</sup>. Thus, among 11 different HBV insertional sites identified in HepaRG cells as VEIS, only one was of the OHJ type and the remaining were HTJ<sup>[51]</sup>. In the same study, analysis of liver biopsies obtained from woodchucks at 1 or 3 h p.i. showed that all (10/10) WHV-host junctions detected had the HTJ format [Figure 1 and Table 1]. The HTJ format of the earliest HBV-host fusions was also apparent in Huh7-NTCP cells examined at 24 h p.i.<sup>[63]</sup>. In the subsequent study of the HepG2-NTCP-C4 cell clone infected with HBV, six virus-host junctions were detected between 30 min and 24 h p.i., and, except for the single fusion with RunX-1 site, all others were formed by HT joints [Table 1]<sup>[27]</sup>. Finally, WCM260 hepatocyte line infected with WHV demonstrated 12 virus insertions into hepatocyte genome which were identified between 15 min and 1 h p.i., and all of them were of the HTJ type<sup>[52]</sup>. Therefore, among 43 virus-host genomic fusions identified in total during the first 24 h p.i. in all four studies discussed, only two (4.6%) of the merges were formed by micro-homology overlapping joining. This clearly showed that both HBV and WHV DNA integrate into hepatocyte genome in the initial stages of infection almost exclusively via HT joining. The creation of these joints is a strong indication that the non-homologous end-joining (NHEJ) pathway was involved in their formation.

## MECHANISM OF INITIAL HEPADNAVIRUS INTEGRATION INTO HOST GENOME

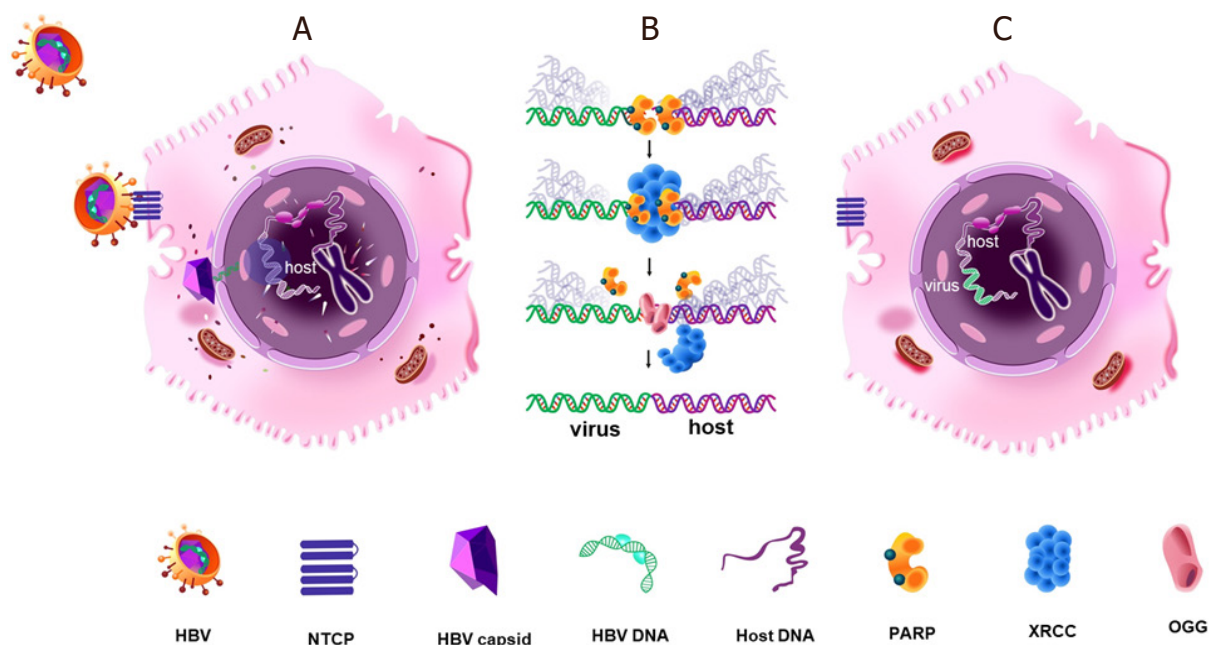
Guided by the finding that the great majority of the very early HBV-host fusions were of the HTJ type, which implied their formation via the NHEJ pathway, and knowing that this pathway is primarily involved in repair of cell double-stranded DNA breaks<sup>[27,51,52]</sup>, which are considered to be precursors for HBV DNA integration<sup>[102]</sup>, we searched for a possible explanation connecting these two events. This task was relatively straightforward since several studies have shown that HBV infection can induce oxidative stress by prompting intracellular production of reactive oxygen species (ROS) and reactive nitrogen species (RNS) leading to DNA oxidation and oxidation-caused DNA breakages<sup>[103]</sup>. To explore if activation of this mechanism in fact occurred and could contribute to the creation of the very early virus-host DNA fusions, we examined woodchuck WCM260 hepatocytes infected with WHV beginning from 15 min p.i.<sup>[52]</sup>. Remarkably, a strong and protracted induction of ROS and transient generation of iNOS, coinciding with microscopically detectable DNA damage, became evident at 15 min after exposure to virus [Figure 1]. While ROS reactivity progressively increased for up to 6 h p.i., reactivity of iNOS was only elevated until 30 min p.i. In addition, cellular DNA damage, as assessed by the nuclear tail moment length using alkaline comet assay, radically increased during the time investigated, i.e., from 15 min to 1 h p.i. [Figure 1]. This suggested that ROS played a primary role in triggering DNA breakages immediately after exposure of WCM260 cells to WHV. Further, as already indicated in the preceding sections, the first WHV-host DNA





**Figure 2.** Molecular formats of the earliest HBV-host and WHV-host genomic junctions: (A) Examples of the HBV-host DNA fusions formed (left) by the NHEJ or (right) by MHMOJ detected between 30 min and 24 h post-infection in hepatocyte-like HepG2-NTCP-C4 clone cells or HepaRG cells; (B) Examples of the WHV-host genomic merges created by NHEJ detected (top left) in woodchuck liver biopsy obtained at 1 h post-infection and (bottom left and right) in woodchucks WCM260 hepatocytes at 15 min and 1 h post-infection. Magnifying glasses at the top of each panel show blown-up regions representing virus-host DNA junctions created by NHEJ or MHMOJ. HBV sequences are depicted as continuous lines, while host sequences are marked by dashed lines. HBV: hepatitis B virus; WHV: woodchuck hepatitis virus; MHMOJ: micro-homology-mediated-overlapping-joining; NHEJ: non-homologous end joining; *p.i.*: post-infection; SINE: short interspersed nuclear element; NBPF: neuroblastoma breakpoint family member; LINE-1: long interspersed nuclear element; NTM: neurotrimin; THE1B-LTR: mammalian apparent LTR retrotransposon; RunX1: runt-related transcription factor-1; MAML-2: mastermind-like 2; host-seq: unidentified woodchuck genomic sequence.





**Figure 3.** Mechanism of the earliest HBV-hepatocyte and WHV-hepatocyte DNA integration by non-homologous end joining: (A) Within 15 min upon HBV entry into hepatocytes via sodium NTCP receptor, ROS and iNOS are produced which shear cellular DNA within the nucleus. With simultaneous entry of the dsHBV DNA into the nucleus, the NHEJ host DNA repair machinery is activated; (B) PARP, a host DNA repair enzyme, is triggered in response to host DNA breakage, and, while performing its host DNA repair function, it also binds to viral dsDNA and forms a complex with the host DNA, which is further joined by XRCC. Since host DNA shearing is mediated by oxidative stress, OGG is also involved in repair; (C) The coordinated DNA repair action results in fusion of viral DNA fragment with host DNA sequence, creating virus-host junction within hepatocyte nucleus. HBV: hepatitis B virus; WHV: woodchuck hepatitis virus; NTCP: taurocholate co-transporting polypeptide; ROS: reactive oxygen species; iNOS: inducible nitric oxide; ds: double-stranded linear; NHEJ: non-homologous end joining; PARP1: Poly(ADP-ribose) polymerase 1; XRCC1: X-ray repair cross-complementing protein 1; OGG: 8-oxyguanine DNA glycosylase.

fusions predominantly with WHV X gene sequence were also detected at 15 min p.i., as well as at 30 min and 1 h p.i., and all of them were of the HTJ type [Table 1]. In this context, it has been previously shown that HBV infection increases levels of ROS and iNOS in human hepatoma HepAD38 cells, which started at the first time point observed at 24 h p.i. and peaked at 72 h p.i. during the 96-h observation period<sup>[104]</sup>. The induction of oxidative stress coincided with augmented expression of genes encoding proteins associated with response to oxidative and metabolic stresses, as well as heat shock proteins<sup>[104,105]</sup>. Considering other viruses with which a similar concept was tested, infection of hepatocyte-like HepG2 cells and J774 cells with Mayaro arbovirus showed increases in activity of the known markers of oxidative stress, including ROS, total superoxide dismutase, and malondialdehyde, within 1 h of infection<sup>[106]</sup>.

To ascertain the repair of DNA breakages caused by oxidative stress was in fact involved in generation of very early virus-host DNA integration, transcription of poly(ADP-ribose) polymerase 1 (PARP1), which plays a central role in recognition of double-strand DNA breaks and in their repair *via* the alternative NHEJ pathway, and transcription of X-ray repair cross-complementing protein 1 (XRCC1), which is the binding partner of PARP1, were examined in WHV-infected WCM260 hepatocytes [Figure 3]<sup>[52]</sup>. In addition, kinetics of nicotinamide adenine dinucleotide (NAD<sup>+</sup>), an indicator of PARP1 activation; heme oxygenase-1 (HO1), a marker of pro-oxidative stress; 8-oxyguanine DNA glycosylase 1 (OGG1), an indicator of response to oxidative damage; and PARP1 cleavage were evaluated in this model. The results have shown the time-synchronized induction of the *PARP1* and *XRCC1* genes accompanied by significantly upregulated activity of NAD<sup>+</sup> and HO1, which all occurred in 15-30 min p.i., while PARP1 cleavage and *OGG1* gene expression became significantly augmented at 5 h and 6 h p.i., respectively [Figure 1]. The results of these

complementary quantitative measurements strongly indicate that WHV is an instant and very potent inducer of oxidative stress in the cells tested and that the PARP1/XRCC1-initiated NHEJ DNA repair machinery is involved in creation of the initial WHV-host genomic junctions.

However, the kinetics of PARP1 transcription after activation and a progressive increase for up to 12 h p.i. subsequently subsided and leveled at approximately the same levels as in uninfected WCM260, as it was evident up to the end of the 72-h observation period<sup>[52]</sup>. In addition, the decline in *PARP1* gene expression coincided with an increase in PARP1 protein cleavage that was significantly augmented between 30 min and 12 h p.i. [Figure 1]. This together suggested that the PARP1-dependent DNA repair may operate chiefly in the initial stages of infection. This also raised a possibility that *de novo* infection might be the main trigger of creation of virus-host fusions *via* the NHEJ mechanism. Interestingly, it has been shown that HBx protein can bind to PARP1 protein and inhibit PARP1 enzymatic activity and repair of DNA breakages<sup>[107]</sup>. The study suggested that there is a physical interaction between PARP1 and HBx proteins and that this may interfere with the recruitment of the DNA repair complex. In this regard, HBx RNA and HBx protein have been identified in infected cells from 4 h p.i. onwards after HBV infection<sup>[108,109]</sup>. It was also uncovered that HBx protein could play a role in initiation of oxidative stress upon HBV infection<sup>[110]</sup>. This was confirmed following transfection of HepG2 cells with HBx protein which induced ROS and activated the oxidative stress pathway within 48 h p.i.<sup>[108,111]</sup>. Similar evidence came from ChangX-34 cells infected with HBV<sup>[112]</sup>. This particular study also showed that ROS production had led to intracellular accumulation of HBx protein, which might be of a pathogenic significance in liver disease<sup>[112]</sup>. Thus, the data accumulated indicate that formation of the very early virus-host DNA fusions *via* the PARP1-initiated DNA repair mechanism is a multifactorial process, which could be potentially influenced by HBx protein.

In summary, recent studies offer the first recognition of the mechanistic aspects of formation of the initial and very early HBV DNA insertions into human hepatocyte genome. In contrast to previous notions, this happens in minutes after the first contact of virus with hepatocyte. The results imply the central roles for the virus-prompted breakages of cellular DNA caused by virus-triggered oxidative stress and the consequential activation of the PARP1/XRCC1-mediated DNA repair machinery. Finding that the very early virus-host DNA junctions have predominantly HT format well collaborates with the direct link between the PARP1 recognition of DNA damage and the NHEJ pathway of DNA repair. The involvement of virus X protein in the formation of the earliest virus-host merges remains uncertain, but this protein has potential to limit PARP1 activity for several hours post-infection.

## CONCLUSIONS

The ability of HBV DNA to integrate into human hepatocyte genome was recognized from the beginning as a characteristic of this virus' biological nature, nonetheless the time when this occurred remained unknown for decades until recently. Historically, it was thought that HBV might integrate not earlier than when chronic hepatitis B is established, hence in months after infection. This unsolved issue prompted our and another group's interest in examining emergence of virus-host genomic fusions in the earliest stages of infection. Human hepatocyte-like cell lines infected with HBV, woodchucks experimentally infected with WHV, and a woodchuck hepatocyte line infected with WHV were utilized as infection models. In our studies, virus-specific inv-PCR amplification was adopted and supplemented with NAH for detection of amplicons displaying virus specific-host DNA fusions, which was followed by their clonal sequencing to identify the joined viral and host sequences.

The data show that HBV integrates into hepatocyte genome as early as 30 min after infection of HepG2 cells overexpressing NTCP and in 1 h when HepaRG cells served as infection targets. Another group reported a similar finding for HBV-infected Huh7-NTCP cells. Importantly, such immediate viral DNA insertions were also evident in WHV infection in woodchucks when their liver biopsies were analyzed in 1 or 3 h p.i.

Further, we uncovered that host non-coding DNA elements, such as retrotransposons, particularly those belonging to the LINE family, transposons, and genes with transposable capabilities were prevailing targets for HBV and WHV DNA integration in the very early stages of infection. They represented overall more than 60% of all insertional sites detected up to 24 h p.i. in four independent studies. These data suggest that HBV can engage from the beginning of infection mobile genetic elements and genes with translocational potential to prompt pro-oncogenic perturbations throughout genomes of infected cells. These perturbations could compromise cell genome stability and augment or silence expression of individual genes important to the development of HCC, and possibly other HBV infection-associated liver and extrahepatic cancers. In addition, a variety of host genes encoding physiologically vital proteins was identified as the very early fusion partners for HBV or WHV DNA.

To recognize a mechanism facilitating formation of the first virus-host DNA merges, our approach was based on the observation that the great majority of the initial and very early fusions were created by non-homologous end joints, also called head-to-tail joints (HTJ), and just very few by overlapping homologous end joining. This implied that their formation involved the NHEJ PARP1/XRCC1-dependent DNA repair pathway and that oxidative stress could be a culprit in this process. In fact, analysis of kinetics of ROS and iNOs levels, cellular DNA damage, and expression of genes associated with oxidative DNA damage response in the WHV-WCM260 infection model demonstrated that all of them significantly increased at the time of formation of the first virus-host junctions at 15-30 min p.i. The results show that the virus is an immediate and very potent inducer of oxidative cell DNA damage and swiftly triggers DNA repair machinery that with a high probability facilitates formation of the earliest virus-host fusions.

The initial and very early HBV-host genomic junctions likely set the stage for the following pro-oncogenic perturbations resulting in HCC. Recognition of their profiles, considering both viral and host sequences involved, frequency of their occurrence, and distribution throughout the infected liver should bring new insights into understanding the pathogenesis of pre-cancerous changes, predict dynamics of the oncogenic process, and prompt ideas regarding novel biomarkers and therapeutic approaches either slowing down or inhibiting progression of these transformations.

## **DECLARATIONS**

### **Authors' contributions**

Made substantial contributions to the concept, design and writing of the article: Chauhan R, Michalak TI

### **Availability of data and materials**

Not applicable.

### **Financial support and sponsorship**

The research from Michalak's laboratory was supported by an operating grant (PIN 22346) from the Cancer Research Society Inc., Canada and the Environmental Cancer Fund - Read for the Cure, Canada and in part by an operating grant (PJT-153001) from the Canadian Institutes of Health Research awarded to Michalak TI.

### **Conflicts of interest**

Both authors declared that there are no conflicts of interest.

### **Ethical approval and consent to participate**

Not applicable.

## Consent for publication

Not applicable.

## Copyright

© The Author(s) 2021.

## REFERENCES

1. Yuen MF, Ahn SH, Chen DS, et al. Chronic hepatitis B virus infection: disease revisit and management recommendations. *J Clin Gastroenterol* 2016;50:286-94.
2. Torresi J, Tran BM, Christiansen D, Earnest-Silveira L, Schwab RHM, Vincan E. HBV-related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. *BMC Cancer* 2019;19:707.
3. Tarocchi M, Polvani S, Marroncini G, Galli A. Molecular mechanism of hepatitis B virus-induced hepatocarcinogenesis. *World J Gastroenterol* 2014;20:11630-40.
4. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol* 2019;16:589-604.
5. Fujiwara N, Friedman SL, Goossens N, Hoshida Y. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. *J Hepatol* 2018;68:526-49.
6. Thomas DL. Global elimination of chronic hepatitis. *N Engl J Med* 2019;380:2041-50.
7. Jia L, Gao Y, He Y, Hooper JD, Yang P. HBV induced hepatocellular carcinoma and related potential immunotherapy. *Pharmacol Res* 2020;159:104922.
8. Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 2002;31:339-46.
9. Kao JH. Hepatitis B vaccination and prevention of hepatocellular carcinoma. *Best Pract Res Clin Gastroenterol* 2015;29:907-17.
10. Michalak TI, Pasquinelli C, Guilhot S, Chisari FV. Hepatitis B virus persistence after recovery from acute viral hepatitis. *J Clin Invest* 1994;93:230-9.
11. Mak LY, Wong DK, Pollicino T, Raimondo G, Hollinger FB, Yuen MF. Occult hepatitis B infection and hepatocellular carcinoma: Epidemiology, virology, hepatocarcinogenesis and clinical significance. *J Hepatol* 2020;73:952-64.
12. Yip TC, Wong GL. Current knowledge of occult hepatitis B infection and clinical implications. *Semin Liver Dis* 2019;39:249-60.
13. Raimondo G, Locarnini S, Pollicino T, Levrero M, Zoulim F, Lok AS; Taormina Workshop on Occult HBV Infection Faculty Members. Update of the statements on biology and clinical impact of occult hepatitis B virus infection. *J Hepatol* 2019;71:397-408.
14. Allweiss L, Strick-Marchand H. In-vitro and in-vivo models for hepatitis B cure research. *Curr Opin HIV AIDS* 2020;15:173-9.
15. Michalak TI. Diverse Virus and Host-Dependent Mechanisms Influence the Systemic and Intrahepatic Immune Responses in the Woodchuck Model of Hepatitis B. *Front Immunol* 2020;11:853.
16. Mulrooney-Cousins PM, Chauhan R, Churchill ND, Michalak TI. Primary seronegative but molecularly evident hepadnaviral infection engages liver and induces hepatocarcinoma in the woodchuck model of hepatitis B. *PLoS Pathog* 2014;10:e1004332.
17. Coffin CS, Pham TN, Mulrooney PM, Churchill ND, Michalak TI. Persistence of isolated antibodies to woodchuck hepatitis virus core antigen is indicative of occult infection. *Hepatology* 2004;40:1053-61.
18. Michalak TI, Mulrooney PM, Coffin CS. Low doses of hepadnavirus induce infection of the lymphatic system that does not engage the liver. *J Virol* 2004;78:1730-8.
19. Michalak TI. Occult persistence and lymphotropism of hepadnaviral infection: insights from the woodchuck viral hepatitis model. *Immunol Rev* 2000;174:98-111.
20. Mulrooney-Cousins PM, Michalak TI. Persistent occult hepatitis B virus infection: experimental findings and clinical implications. *World J Gastroenterol* 2007;13:5682-6.
21. 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;286:533-5.
22. 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;286:535-8.
23. Marion PL, Salazar FH, Alexander JJ, Robinson WS. State of hepatitis B viral DNA in a human hepatoma cell line. *J Virol* 1980;33:795-806.
24. Shafritz DA, Shouval D, Sherman HI, Hadziyannis SJ, Kew MC. Integration of hepatitis B virus DNA into the genome of liver cells in chronic liver disease and hepatocellular carcinoma. Studies in percutaneous liver biopsies and post-mortem tissue specimens. *N Engl J Med* 1981;305:1067-73.
25. Budzinska MA, Shackel NA, Urban S, Tu T. Cellular Genomic Sites of Hepatitis B Virus DNA Integration. *Genes (Basel)* 2018;9:365.
26. Tu T, Budzinska MA, Shackel NA, Urban S. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses* 2017;9:75.
27. Chauhan R, Shimizu Y, Watashi K, Wakita T, Fukasawa M, Michalak TI. Retrotransposon elements among initial sites of hepatitis B virus integration into human genome in the HepG2-NTCP cell infection model. *Cancer Genet* 2019;235-236:39-56.
28. Tsukuda S, Watashi K. Hepatitis B virus biology and life cycle. *Antiviral Res* 2020;182:104925.
29. Schluter V, Meyer M, Hofschneider PH, Koshy R, Caselmann WH. Integrated hepatitis B virus X and 3' truncated preS/S sequences

- derived from human hepatomas encode functionally active transactivators. *Oncogene* 1994;9:3335-44.
30. Saigo K, Yoshida K, Ikeda R, 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;29:703-8.
  31. Sung WK, Zheng H, Li S, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012;44:765-9.
  32. Ding D, Lou X, Hua D, 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:e1003065.
  33. Yang M, Yang G, Li F, et al. HBV integrated genomic characterization revealed hepatocyte genomic alterations in HBV-related hepatocellular carcinomas. *Mol Clin Oncol* 2020;13:79.
  34. Ishii T, Tamura A, Shibata T, et al. Analysis of HBV genomes integrated into the genomes of human hepatoma PLC/PRF/5 Cells by HBV sequence capture-based next-generation sequencing. *Genes (Basel)* 2020;11:661.
  35. Matsuda Y, Ichida T. Impact of hepatitis B virus X protein on the DNA damage response during hepatocarcinogenesis. *Med Mol Morphol* 2009;42:138-42.
  36. Wollersheim M, Debelka U, Hofschneider PH. A transactivating function encoded in the hepatitis B virus X gene is conserved in the integrated state. *Oncogene* 1988;3:545-52.
  37. 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;87:5628-32.
  38. Kekulé 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;343:457-61.
  39. Yamamoto S, Mita E, Nakatake H, Takimoto M, Koshy R, Matsubara K. Transactivating function of integrated hepatitis B virus. *Biochem Biophys Res Commun* 1993;197:1209-15.
  40. Twu JS, Lai MY, Chen DS, Robinson WS. Activation of protooncogene c-jun by the X protein of hepatitis B virus. *Virology* 1993;192:346-50.
  41. Lauer U, Weiss L, Hofschneider PH, Kekulé AS. The hepatitis B virus pre-S/S(t) transactivator is generated by 3' truncations within a defined region of the S gene. *J Virol* 1992;66:5284-9.
  42. Henkler F F, Koshy R. Hepatitis B virus transcriptional activators: mechanisms and possible role in oncogenesis. *J Viral Hepat* 1996;3:109-21.
  43. Caselmann WH. Transactivation of cellular gene expression by hepatitis B viral proteins: a possible molecular mechanism of hepatocarcinogenesis. *J Hepatol* 1995;22:34-7.
  44. Brechot C, Scotto J, Charnay P, et al. Detection of hepatitis B virus DNA in liver and serum: a direct appraisal of the chronic carrier state. *The Lancet* 1981;318:765-8.
  45. Michalak TI, Churchill ND. Interaction of woodchuck hepatitis virus surface antigen with hepatocyte plasma membrane in woodchuck chronic hepatitis. *Hepatology* 1988;8:499-506.
  46. Gripon P, Rumin S, Urban S, et al. Infection of a human hepatoma cell line by hepatitis B virus. *Proc Natl Acad Sci U S A* 2002;99:15655-60.
  47. Yan H, Zhong G, Xu G, et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife* 2012;1:e00049.
  48. Wettengel JM, Burwitz BJ. Innovative HBV animal models based on the entry receptor NTCP. *Viruses* 2020;12: 828.
  49. Li F, Wang Z, Hu F, Su L. Cell culture models and animal models for HBV study. In: Tang H, editor. Hepatitis B virus infection. Singapore: Springer; 2020. pp. 109-35.
  50. Hayes CN, Chayama K. HBV culture and infectious systems. *Hepatology* 2016;10:559-66.
  51. Chauhan R, Churchill ND, Mulrooney-Cousins PM, Michalak TI. Initial sites of hepadnavirus integration into host genome in human hepatocytes and in the woodchuck model of hepatitis B-associated hepatocellular carcinoma. *Oncogenesis* 2017;6:e317.
  52. Chauhan R, Michalak TI. Kinetics of DNA damage repair response accompanying initial hepadnavirus-host genomic integration in woodchuck hepatitis virus infection of hepatocyte. *Cancer Genet* 2020;244:1-10.
  53. Ko C, Chakraborty A, Chou WM, et al. Hepatitis B virus genome recycling and de novo secondary infection events maintain stable cccDNA levels. *J Hepatol* 2018;69:1231-41.
  54. Dezhbord M, Lee S, Kim W, Seong BL, Ryu WS. Characterization of the molecular events of covalently closed circular DNA synthesis in de novo Hepatitis B virus infection of human hepatoma cells. *Antiviral Res* 2019;163:11-8.
  55. Chakraborty A, Ko C, Henning C, et al. Synchronised infection identifies early rate-limiting steps in the hepatitis B virus life cycle. *Cell Microbiol* 2020;22:e13250.
  56. Tu T, Zehnder B, Qu B, et al. A novel method to precisely quantify hepatitis B virus covalently closed circular (ccc)DNA formation and maintenance. *Antiviral Res* 2020;181:104865.
  57. Khakpoor A, Ni Y, Chen A, et al. Spatiotemporal differences in presentation of CD8 T cell epitopes during hepatitis B virus infection. *J Virol* 2019;93:e01457-18.
  58. Liu Q, Huang J, Jia R, et al. The pregenome/C RNA of duck hepatitis B virus is not used for translation of core protein during the early phase of infection in vitro. *Virus Res* 2015;196:13-9.
  59. Tagawa M, Omata M, Okuda K. Appearance of viral RNA transcripts in the early stage of duck hepatitis B virus infection. *Virology* 1986;152:477-82.
  60. Guy CS, Mulrooney-Cousins PM, Churchill ND, Michalak TI. Intrahepatic expression of genes affiliated with innate and adaptive immune responses immediately after invasion and during acute infection with woodchuck hepadnavirus. *J Virol* 2008;82:8579-91.



61. Diao J, Churchill ND, Michalak TI. Complement-mediated cytotoxicity and inhibition of ligand binding to hepatocytes by woodchuck hepatitis virus-induced autoantibodies to asialoglycoprotein receptor. *Hepatology* 1998;27:1623-31.
62. Mulrooney-Cousins PM, Michalak TI. Repeated passage of wild-type woodchuck hepatitis virus in lymphoid cells does not generate cell type-specific variants or alter virus infectivity. *J Virol* 2008;82:7540-50.
63. 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:e02007-17.
64. Shedlock AM. Phylogenomic investigation of CR1 LINE diversity in reptiles. *Syst Biol* 2006;55:902-11.
65. Goodier JL, Kazazian HH Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* 2008;135:23-35.
66. Gentles AJ, Wakefield MJ, Kohany O, et al. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res* 2007;17:992-1004.
67. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 2001;35:501-38.
68. Smit AF. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* 1996;6:743-8.
69. Lovsin N, Gubensek F, Kordi D. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in *Deuterostomia*. *Mol Biol Evol* 2001;18:2213-24.
70. Shaul Y, Garcia PD, Schonberg S, Rutter WJ. Integration of hepatitis B virus DNA in chromosome-specific satellite sequences. *J Virol* 1986;59:731-4.
71. Nagaya T, Nakamura T, Tokino T, et al. The mode of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. *Genes Dev* 1987;1:773-82.
72. Terai Y, Takahashi K, Okada N. SINE cousins: the 3'-end tails of the two oldest and distantly related families of SINEs are descended from the 3' ends of LINEs with the same genealogical origin. *Mol Biol Evol* 1998;15:1460-71.
73. Naville M, Henriot S, Warren I, et al. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr Biol* 2019;29:1161-1168.e6.
74. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2011;13:36-46.
75. Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* 1988;57:159-97.
76. Martín-Moreno AM, Roncador G, Maestre L, et al. CSF1R protein expression in reactive lymphoid tissues and lymphoma: its relevance in classical hodgkin lymphoma. *PLoS One* 2015;10:e0125203.
77. Rubin E, Lithwick G, Levy AA. Structure and evolution of the hAT transposon superfamily. *Genetics* 2001;158:949-57.
78. de Souza FS, Franchini LF, Rubinstein M. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol* 2013;30:1239-51.
79. Perincheri S, Dingle RW, Peterson ML, Spear BT. Hereditary persistence of alpha-fetoprotein and H19 expression in liver of BALB/cJ mice is due to a retrovirus insertion in the *Zhx2* gene. *Proc Natl Acad Sci U S A* 2005;102:396-401.
80. Vandepoele K, Staes K, Andries V, van Roy F. Chibby interacts with NBPF1 and clusterin, two candidate tumor suppressors linked to neuroblastoma. *Exp Cell Res* 2010;316:1225-33.
81. Lutz SZ, Hennige AM, Feil S, et al. Genetic ablation of cGMP-dependent protein kinase type I causes liver inflammation and fasting hyperglycemia. *Diabetes* 2011;60:1566-76.
82. Miyagawa K, Sakakura C, Nakashima S, et al. Down-regulation of RUNX1, RUNX3 and CBFbeta in hepatocellular carcinomas in an early stage of hepatocarcinogenesis. *Anticancer Res* 2006;26:3633-43.
83. Luukkonen TM, Pöyhönen M, Palotie A, et al. A balanced translocation truncates Neurotrimin in a family with intracranial and thoracic aortic aneurysm. *J Med Genet* 2012;49:621-9.
84. Asmann YW, Necela BM, Kalari KR, et al. Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res* 2012;72:1921-8.
85. Li C, Ruan HQ, Liu YS, et al. Quantitative proteomics reveal up-regulated protein expression of the SET complex associated with hepatocellular carcinoma. *J Proteome Res* 2012;11:871-85.
86. Kho CJ, Wang Y, Zarbl H. Effect of decreased *fte-1* gene expression on protein synthesis, cell growth, and transformation. *Cell Growth Differ* 1996;7:1157-66.
87. Shimojima K, Okumura A, Natsume J, et al. Spinocerebellar ataxias type 27 derived from a disruption of the fibroblast growth factor 14 gene with mimicking phenotype of paroxysmal non-kinesigenic dyskinesia. *Brain Dev* 2012;34:230-3.
88. Tonon G, Modi S, Wu L, et al. t(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. *Nat Genet* 2003;33:208-13.
89. Vandepoele K, Andries V, Van Roy N, et al. A constitutional translocation t(1;17)(p36.2;q11.2) in a neuroblastoma patient disrupts the human NBPF1 and ACCN1 genes. *PLoS One* 2008;3:e2207.
90. De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris MJ, Férec C, De Braekeleer M. RUNX1 translocations and fusion genes in malignant hemopathies. *Future Oncol* 2011;7:77-91.
91. Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* 2004;6:73-7.
92. Jenuwein T. The epigenetic magic of histone lysine methylation. *FEBS J* 2006;273:3121-35.
93. Kim TH, Abdullaev ZK, Smith AD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007;128:1231-45.

94. Laugesen A, Højfeldt JW, Helin K. Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. *Mol Cell* 2019;74:8-18.
95. Mehra M, Chauhan R. Long noncoding RNAs as a key player in hepatocellular carcinoma. *Biomark Cancer* 2017;9:1179299X1773730.
96. Dubois-Chevalier J, Staels B, Lefebvre P, Eeckhoutte J. The ubiquitous transcription factor CTCF promotes lineage-specific epigenomic remodeling and establishment of transcriptional networks driving cell differentiation. *Nucleus* 2015;6:15-8.
97. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 1999;98:387-96.
98. Wu L, Murat P, Matak-Vinkovic D, Murrell A, Balasubramanian S. Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. *Biochemistry* 2013;52:9519-27.
99. Jain S, Chang TT, Chen S, et al. Comprehensive DNA methylation analysis of hepatitis B virus genome in infected liver tissues. *Sci Rep* 2015;5:10478.
100. Watanabe Y, Yamamoto H, Oikawa R, et al. DNA methylation at hepatitis B viral integrants is associated with methylation at flanking human genomic sequences. *Genome Res* 2015;25:328-37.
101. Homs M, Buti M, Quer J, et al. Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. *Nucleic Acids Res* 2011;39:8457-71.
102. Mason WS, Gill US, Litwin S, et al. HBV DNA integration and clonal hepatocyte expansion in chronic hepatitis B patients considered immune tolerant. *Gastroenterology* 2016;151:986-998.e4.
103. Alavian SM, Showraki A. Hepatitis B and its relationship with oxidative stress. *Hepat Mon* 2016;16:e37973.
104. Severi T, Ying C, Vermeesch JR, et al. Hepatitis B virus replication causes oxidative stress in HepAD38 liver cells. *Mol Cell Biochem* 2006;290:79-85.
105. Yuan K, Lei Y, Chen HN, et al. HBV-induced ROS accumulation promotes hepatocarcinogenesis through Snail-mediated epigenetic silencing of SOCS3. *Cell Death Differ* 2016;23:616-27.
106. Camini FC, da Silva Caetano CC, Almeida LT, et al. Oxidative stress in Mayaro virus infection. *Virus Res* 2017;236:1-8.
107. Na TY, Ka NL, Rhee H, et al. Interaction of hepatitis B virus X protein with PARP1 results in inhibition of DNA repair in hepatocellular carcinoma. *Oncogene* 2016;35:5435-45.
108. Niu C, Livingston CM, Li L, et al. The Smc5/6 complex restricts HBV when localized to ND10 without inducing an innate immune response and is counteracted by the HBV X protein shortly after infection. *PLoS One* 2017;12:e0169648.
109. Korniyev D, Ramakrishnan D, Voitenleitner C, et al. Spatiotemporal analysis of hepatitis B virus X protein in primary human hepatocytes. *J Virol* 2019;93:e00248-19.
110. Ren JH, Chen X, Zhou L, et al. Protective role of Sirtuin3 (SIRT3) in oxidative stress mediated by hepatitis B virus X protein expression. *PLoS One* 2016;11:e0150961.
111. Hu X, Jiang J, Ni C, et al. HBV integration-mediated cell apoptosis in HepG2.2.15. *J Cancer* 2019;10:4142-50.
112. Wang JH, Yun C, Kim S, et al. Reactive oxygen species modulates the intracellular level of HBx viral oncoprotein. *Biochem Biophys Res Commun* 2003;310:32-9.