Cancer Genetics 235–236 (2019) 39–56

Cancer Genetics

ORIGINAL ARTICLE

Retrotransposon elements among initial sites of hepatitis B virus integration into human genome in the HepG2-NTCP cell infection model

Ranjit Chauhan[a], Yoshimi Shimizu[b], Koichi Watashi[c], Takaji Wakita[c], Masayoshi Fukasawa[b], Tomasz I Michalak[a,*]

[a] *Molecular Virology and Hepatology Research Group, Division of BioMedical Sciences, Faculty of Medicine, Health Science Centre, Memorial University, St. John's, NL, Canada;* [b] *Department of Biochemistry and Cell Biology, National Institute of Infectious Diseases, Tokyo, Japan;* [c] *Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan*

Abstract

Integration of hepatitis B virus (HBV) DNA into host's genome is evident in all stages and models of HBV infection. Investigations of the initial virus-host junctions have been just recently initiated since their nature may promote liver oncogenesis immediately following infection. We examined the time-frame and host sites at which HBV integrates in HepG2 cells overexpressing sodium taurocholate co-transporting polypeptide (NTCP) receptor mediating HBV entry. HepG2-NTCP cells were analyzed from 15 min to 13 days post-infection (p.i.). The results showed that except for 15 min p.i., HBV-host integrations were detected at all time points thereafter. At 30 min p.i., virus junctions with retrotransposon SINE and with neuroblastoma breakpoint family member 1 gene were detected. At one-hour p.i., HBV integration with retrotransposon THE-1B-LTR was identified, while virus insertions into proline-rich protein and protein kinase cGMP-dependent type 1 encoding genes were found at 3 h p.i. Fusion with runt-related transcription factor 1 was detected at 24 h p.i. and merges with 9 different genes at 13 day p.i. The data showed that retrotransposon elements are frequent among first-hit sites of HBV insertion. This may suggest a mechanism by which HBV DNA may spread across host's genome from earliest stages of infection.

Keywords Hepatitis B virus, Hepatocellular carcinoma, Sodium taurocholate co-transporting polypeptide receptor, Virus initial integration sites, Retrotransposon, Chromatin marks.

Introduction

Hepatitis B virus (HBV) is one of the most common and deceitful human pathogens that causes liver injury ranging from an asymptomatic disease through acute and chronic hepatitis type B, to liver cirrhosis and primary hepatocellular carcinoma (HCC) [1]. HBV is recognized as the most potent viral oncogene responsible for thousands of liver cancer-related deaths annually [2]. Its oncogenic potency has been primarily aligned with integration of viral DNA into hepatocyte genome capable of compromising cell genome stability and modifying expression of individual genes [3]. In this regard, HBV-host DNA fusions have been identified in all the stages and in both *in vitro* and natural animal models of HBV infection investigated to date [4,5]. From the beginning of studies on the HBV-associated carcinogenesis, it has been hypothesized that HBV integrates in early stages of infection and this can culminate in liver cancer after prolonged periods of virus latency [6-8]. By applying *in vitro* models of HBV infection, recent findings supported this hypothesis by showing that HBV-host DNA fusion can be indeed a very early event [9,10]. The same was found in the woodchuck model of hepatitis B soon after *de novo* infection with woodchuck hepatitis virus (WHV) [9]. In addition, a direct oncogenic role for integrated hepadnaviral DNA in the development of HCC was substantiated by demonstrating HCC development in the course of primary occult infection in the absence of chronic liver inflammation in the woodchuck model [11].

Received February 14, 2019; received in revised form April 11, 2019; accepted April 18, 2019
* Corresponding author.
E-mail address: timich@mun.ca

2210-7762/$ - see front matter
https://doi.org/10.1016/j.cancergen.2019.04.060

It was reported that HBV DNA integration occurs spontaneously in approximately 1–5% of HBV-infected hepatocytes when evaluated by inverse-polymerase chain reaction (inv-PCR) [12]. Assuming that relatively small number of hepatocytes harbour HBV DNA integration, it is crucial to determine the host's genomic sites at which HBV initially integrates since this could be of a paramount importance to the initiation and spread of pro-oncogenic perturbations in hepatocytes which may underline cancerous growth. Most of the studies searching for sites of HBV-host fusions were done on HBV-infected HCC and adjacent non-cancerous tissues or liver biopsies collected during chronic stages of HBV infection, but they proved inconclusive in determining the sites involved in the virus early integration events [13,14]. However, with the advent of liver cell lines capable of supporting productive virus replication after exposure to either authentic (naturally occurring) or cell culture-derived infectious HBV, such studies became feasible and they already generated valuable data [9,10].

In the previous study utilizing human hepatocyte-like HepaRG cells infected with authentic HBV and woodchucks infected with wild-type WHV, sequences of some early integrations were dissected on the single-nucleotide-level [9]. Both the culture and the natural *in vivo* model investigated revealed hepadnavirus DNA fusions with variety of host's genes as early as one-hour post-infection. Interestingly, mobile genetic elements in a form of tandemly repeating non-coding DNA sequences were found among the sites targeted by HBV DNA for integration in the early stages of infection. In the current work, we explored HepG2 cells stably transfected with sodium taurocholate co-transporting polypeptide (NTCP), a molecule serving as a HBV receptor [15], to advance characterization of the host's genomic sequences forming initial fusions with HBV. HepG2-NTCP cells demonstrate a high susceptibility to HBV infection and an efficient production of biologically competent virions. Their applicability for studies on HBV pathobiology and testing efficacy of cell HBV entry inhibitors and antivirals has been documented [16].

The role of chromatin modifications in transcription of cellular and HBV genes, including covalently closed circular DNA (cccDNA), has been investigated and to some degree deciphered [17–19]. Nevertheless, there is no data regarding the status of chromatin marks at the initial time points post-infection when HBV become integrated into the host's genome. Thus, in addition to the main purpose of the study, we performed *in silico* analysis to identify whether common chromatin modifications may characterize targets of HBV integration in HepG2 cells. Respectively, the presence or the absence of the marks suggesting the host's sequence acetylation, methylation and DNase hypersensitivity, as well as potential enhancer properties were assessed [20–22].

Application of the HepG2-NTCP cell-based HBV infection system in our study resulted in new finding that the first fusions of HBV with host's genome may take place as early as in 30 min after exposure to virus. The results further confirmed the fact that integration of virus DNA into retrotransposon sequences is a common event in the initial stages of infection. The results also suggested apparent lack of the common chromatin marks possibly modifying transcription of the host sequences targeted by HBV integration. Overall, the data obtained indicate that unlocking the intertwined interactions between HBV and retrotransposon elements and delineating consequences of these interactions could be one of the key factors in deciphering the mechanism initiating HCC development in HBV infection.

Materials and methods

HepG2-NTCP cells and HBV infection

HBV infection experiments were carried out in HepG2-NTCP cell line created by stably transfecting and overexpressing HBV receptor NTCP in these cells [16]. HepG2 cell line is universally used for liver and liver-related research and is catalogued in American Type Culture Collection (ATCC) repository with catalogue number ATCC HB-8065. The HepG2 cells were obtained from and authenticated by Cellular Engineering Technologies Inc. (Coralville, IA). HepG2-NTCP-C4 cell line was established in the National Institute of Infectious Diseases Japan in 2013. The parental cells were authenticated at that time according to the instruction from Cellular Engineering Technologies Inc. Clone C4 showed the highest efficiency of supporting HBV infection and the cell line derived from this clone was designated as HepG2-NTCP-C4.

The HepG2-NTCP-C4 cell line was found to be susceptible to infection with both authentic patient-derived HBV at 100 virus genome equivalents vge (vge)/cell and cell culture-derived HBV at 6×10^3 vge/cell as has shown in the previous studies [16,23].

HBV of genotype D secreted by HepG2.2.15.7 cell line served as inoculum [24]. HBV virions were concentrated by Amicon Ultra-15 Centrifugal Filters (Merck Millipore Ltd., Cork, Ireland) and stored in 10% fetal calf serum (FCS) at −80 °C until use. Infection was carried out in infection medium (IM) comprised by Dulbecco's modified Eagle's medium (DMEM)/F12 with GlutaMAX, 10% FCS, 10 mM HEPES, 200 units/mL penicillin, 200 μg/mL streptomycin, 5 μg/mL insulin, 4% polyethylene glycol-8000 (PEG-8000), and 3% dimethyl sulfoxide (DMSO). HepG2-NTCP-C4 cells were plated at the density of 5×10^5 cells/well in collagen-coated 6-well culture plates. After achieving 50–60% confluence, the cells were infected with HBV at estimated 1000 vge/cell in 1 mL of IM per well, and incubated in a cell culture incubator for 15 min, 30 min, 1 h, 3 h, 24 h and 13 days. HBV infection was carried out in duplicate for each time point investigated. HepG2-NTCP-C4 cells cultured for 15 min, 1 h and 24 h in the absence of virus served as mock infection controls. After completion of infection and control experiments, cells were washed five times with serum-free DMEM by slow-speed centrifugation. Total DNA was extracted from the final cell pellets using Blood/Cultured Cell Genomic DNA Extraction Mini Kit (FAVORGEN Biotech Corp., Pingtung City, Taiwan).

Inv-PCR for detection of HBV-host DNA junctions

The inv-PCR was applied to identify HBV-host junctions using HBV X gene-specific primers capable of amplifying all HBV genotypes and conditions reported in detail previously [9,11]. Two rounds of amplifications, direct and nested, were normally performed. The exception was a situation when amplicon bands were detectable on agarose gel after the first (direct) run of inv-PCR. As controls, DNA from cells subjected

to mock infection, and water added to the inv-PCR mixtures instead of test DNA or to the amplicons after direct inv-PCR were used [9,11].

Validation of inv-PCR products by NAH

After visualization of inv-PCR amplicons on the agarose gels and documentation, the products were transferred onto Amersham Hybond-N+ membrane (GE Health Care Bio-Sciences, Piscataway, NJ) and analyzed by nucleic acid hybridization (NAH). As a probe, radiolabelled, full-length recombinant HBV DNA was used [9,11]. It is important to note that only the NAH-positive bands carrying virus-specific sequences were excised from agarose gels and processed for analysis, similar as in our previous studies [9,11]. The subsequent steps included the amplicon DNA purification, digestion, circularization with T4 DNA ligase, and linearization, as detailed previously [9].

Verification of HBV-host DNA junctions by clonal sequencing

The resulting DNA were cloned using the TOPO-TA system (Invitrogen Life Technologies, Burlington, Canada). Screening of the clones for the insert was done, in this first instance, by their sub-cloning into the second LB agar plate. Then, each clone was further expanded in the LB Luria broth overnight. Screening of the resulting clones was done by colony PCR with inv-PCR nested primers. Clones showing amplification bands on agarose gel were purified using Qiagen gel purification kit (Invitrogen). The amplicon DNA was sequenced bi-directionally using inv-PCR nested primers [11].

Identification and analysis of joined HBV and human DNA sequences

HBV sequences were aligned using the BioEdit software (Ibis Biosciences, Carlsbad, CA) and GenBank reference sequence X72702 (genotype D) as a reference. Human DNA sequences were mapped using NCBI BLAST (National Center for Biotechnology Information, Bethesda, MD) and coordinates assigned with BLAST genome browser (University of California at Sacramento, UCSC) using human GRCh38 as reference sequence. Retrotransposable sequences were identified using Dfam browser for repetitive elements and SINEBase online software for short interspersed sequences, as indicated previously [9]. To determine the presence of chromatin marks indicative of the HepG2 sequences with potential enhancer activity, and those methylated, acetylated and hypersensitive to DNase, each of the host nucleotide sequences found to be joined by HBV DNA (see Table 1) was fed into the UCSC server with ENCODE program [25] to track for the CCCTC-binding factor (CTCF*)* clusters, Tri-methylation of lysine 4 on histone H3 (H3K4me3) hot spots, enhancer of Zeste homolog 2 (EZH2) sequences, histone H3 lysine 27 demethylase (H3K27ac) clusters, and DNase binding regions [26].

Deposition numbers for HepG2 cells in the ENCODE database used for analysis were ENCSR575RRX, ENCSR000DUF, ENCSR00AMP, and ENCSR000AMO. In addition, to precisely determine the nucleotides where *CTCF* may bind on the genes targeted by HBV integration, analysis utilizing the CTCFBSDB 2.0 online software was performed [27].

Results

Overall profile of HBV integration in HepG2-NTCP cells

Hepatocyte-like HepG2-NTCP-C4 cells exposed to infectious HBV derived *in vitro* from HepG2.2.15.7 cells have displayed signals indicative of virus-host genome fusions from 30 min post-infection (p.i.) onwards, as became apparent by NAH analysis of the inv-PCR products. The presence of HBV-host DNA integrations was confirmed by clonal sequencing and only the junctions successfully sequenced were accepted and reported. There were no HBV DNA signals detected in the nested inv-PCR mixtures containing DNA from HepG2-NTCP-C4 cells collected at 15 min p.i. or the cells subjected to mock infections carried out for 30 min, 1 h or 24 h.

Overall, 15 distinctive sites of HBV DNA insertion into 9 different human chromosomes (Ch) were identified. The nine chromosomes with HBV hits were Ch-1, Ch-2, Ch-4, Ch-5, Ch-8, Ch-10, Ch-11, Ch14, and Ch-21 (Table 1). Considering all sites identified in this study, single HBV integration sites were detected in Ch-1, Ch-4, Ch-5, Ch-14, and Ch-21. Two integrations per chromosome were found on Ch-2, Ch-8 and Ch-11, and three on Ch-10. There also was a singular junction formed between HBV and host's sequence which could not be defined and, therefore, it was marked as unidentified (UI). Based on the time of the detection in relation to the first contact of cells with HBV inoculum, the junctions were categorized into two main categories, following the scheme previously proposed [9]. Hence, the junctions detected at 30 min p.i. until 24 h.p.i. were designated as very early integration sites (VEIS), while those at 13 dp.i. as late or not-early integration sites (NEIS). In the current study, the first virus-host DNA fusions detected at 30 min p.i. were also designated as the initial integration sites (IIS). Overall, there were 6 sites classified as VEIS (i.e., SINE (30 min), NBPF-1 (30 min), THE-1B-LTR (1 h), PRR-16 (3 h), PRKG-1 (3 h) and RunX-1 (24 h)) and 9 as NEIS (i.e., *h*AT-18-Ssa-LTR, DNTNP, FAM90A, PEB-4, PCDH15, LINE2, UI, lnc101929653 and C14Orf29), as well as two of the six VEIS also were IIS (i.e., SINE and NBPF-1) (Table 1).

The initial sites of HBV DNA integration

Twenty-one out of 40 bi-directionally sequenced clones, which were selected by colony PCR after cloning of the inv-PCR amplicons demonstrating NAH-detectable HBV DNA signals at 30 min p.i. that carried virus-host junctions. Among these clones, 17 showed HBV DNA integration at short-interspersed nuclear element (SINE) (Table 1). This element belongs to non-tandem genomic repeats termed as retrotransposons. The HBV sequence forming junction with SINE was 40 base pairs (bp) in length and corresponded to nucleotides (nts) of HBV X gene (*HBx*) at positions 1647–1687. The length of the joined host sequence was 294-bp. The sequence was

Table 1 HBV DNA integrations detected in HepG2-NTCP cells at different time points post infection.

Time post infection	Virus-host junctions detected (no)	Integrated HBV sequence (nt position)	Host sequence (bp length)	Chromosome number	Sequence locus	Host sequence (nt position)	Gene	Type of virus-host junction
30 min	2	1647-1774	92	Ch-1	p36.13	16583566-16583474	NBPF-1	HTJ
		1647-1774	92	Ch-1	p36.13	16583566-16583474	NBPF-1	HTJ
		1647-1774	92	Ch-1	p36.13	16583566-16583474	NBPF-1	HTJ
		1647-1774	92	Ch-1	p36.13	16583566-16583474	NBPF-1	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
		1647-1687	294	Ch-10	q23.2	87153862-87154156	SINE	HTJ
1 h	1	1647-1725	193	Ch-2	p12	82086308-82086501	THE1B-LTR	HTJ
		1647-1725	193	Ch-2	p12	82086308-82086501	THE1B-LTR	HTJ
		1647-1725	193	Ch-2	p12	82086308-82086501	THE1B-LTR	HTJ
		1647-1725	193	Ch-2	p12	82086308-82086501	THE1B-LTR	HTJ
		1647-1725	193	Ch-2	p12	82086308-82086501	THE1B-LTR	HTJ

(*continued on next page*)

Table 1 (*continued*)

Time post infection	Virus-host junctions detected (no)	Integrated HBV sequence (nt position)	Host sequence (bp length)	Chromosome number	Sequence locus	Host sequence (nt position)	Gene	Type of virus-host junction
		1647-1725	193	Ch-2	p12	82086308-82086501	THE1B-LTR	HTJ
		1647-1725	193	Ch-2	p12	82086308-82086501	THE1B-LTR	HTJ
3 h	2	1647-1820	38	Ch-5	q23.1	119935846-119935,884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866–1945	54	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
		1647-1820	38	Ch-5	q23.1	119935846-119935884	PRR-16	HTJ
		1866-1945	23	Ch-10	q11.23	51769849-51769877	PRKG-1	HTJ
24 h	1	1647-1792	481	Ch-21	q22.12	35712645-35713126	RunX-1	OHJ
		1647-1792	481	Ch-21	q22.12	35712645-35713126	RunX-1	OHJ
13 days	9	1647-1754	220	Ch-2	q33.1	199220298-199220518	*hAT*-18 Ssa	OHJ
		1647-1754	220	Ch-2	q33.1	199220298-199220518	*hAT*-18 Ssa	OHJ
		1647-1754	220	Ch-2	q33.1	199220298-199220518	*hAT*-18 Ssa	OHJ

(*continued on next page*)

Table 1 (*continued*)

Time post infection	Virus-host junctions detected (no)	Integrated HBV sequence (nt position)	Host sequence (bp length)	Chromosome number	Sequence locus	Host sequence (nt position)	Gene	Type of virus-host junction
		1647-1754	220	Ch-2	q33.1	199220298–199220518	*hAT*-18 Ssa	OHJ
		1605-1623	289	Ch-4	p16.3	1668784-1669073	DNTNP	HTJ
		1605-1623	289	Ch-4	p16.3	1668784-1669073	DNTNP	HTJ
		1605-1623	289	Ch-4	p16.3	1668784-1669073	DNTNP	HTJ
		1605-1623	289	Ch-4	p16.3	1668784-1,669073	DNTNP	HTJ
		1606-1664	19	Ch-8	p21.3	22796773-22796792	PEB-4	HTJ
		1606-1664	19	Ch-8	p21.3	22796773-22796792	PEB-4	HTJ
		1602-1622	214	Ch-8	p23.1	8016999-8017213	FAM90A	HTJ
		1647-1764	277	Ch-10	q21.1	476850-476573	PCDH15	OHJ
		1647-1764	277	Ch-10	q21.1	476850–476573	PCDH15	OHJ
		1647-1818	542	Ch-11	q13.4	71323803-71323261	L2	OHJ
		1647-1818	542	Ch-11	q13.4	71323803-71323261	L2	OHJ
		1647-1818	542	Ch-11	q13.4	71323803-71323261	L2	OHJ
		1647-1818	542	Ch-11	q13.4	71323803-71323261	L2	OHJ
		1603-1659	93	Ch-11	p15.5	233-140	lnc101929653	OHJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1602-1620	206	Ch-14	q22.1	94423774-94423980	C14Orf29	HTJ
		1615-1668	12	–	–	–	UI	HTJ
		1615-1668	12	–	–	–	UI	HTJ
		1615-1668	12	–	–	–	UI	HTJ

Abbreviations: no, number; nt, nucleotide; bp, base-pair; Ch, chromosome; HTJ, head-to-tail junction; OHJ, overlapping homologous junction; 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; RunX-1, runt-related transcription factor 1; hAT-18-Ssa-LTR, hobo activator *Salmo salar*-18 long terminal repeat; DNTNP, dorsal neural tube nuclear protein; PEB4, phosphatidylethanolamine binding protein 4; FAM90, family with sequence similarity 90; PCDH15, protocadherin 15; LINE2, long-interspersed nuclear element 2; lnc101929653, long non-coding RNA 101929653; C14Orf29, alpha/beta hydrolase domain-containing protein-12B; UI, unidentified.

located on Ch-10, locus q23.2 at nts 87153862-87154156 (Fig. 1A). The HBV-SINE junction was in a head-to-tail orientation referred to as head-to-tail junction (HTJ). The remaining 4 clones displayed HBV integration into the neuroblastoma breakpoint family member 1 (NBPF-1) (Table 1). NBPF-1 is a pseudogene that functions as a tumor suppressor for neuroblastoma [28]. This fusion was of the HTJ type**.** It was formed by a 127-bp-long HBV sequence matching *HBx* nts 1647-1774 and a 92-bp-long host DNA spanning nts 16583566-16583474 on Ch-1p36.13 (Fig. 1B). Collectively, among the clones carrying IIS detected at 30 min p.i. prevailed (85%) those in which HBV DNA merged with retrotransposon.

HBV very early integration sites identified between one and 24 h post-infection

Sequencing of 48 clones derived following cloning of HBV DNA-positive inv-PCR products from cells collected at one h.p.i. resulted in identification of 7 clones with virus-host junctions (Table 1). In all 7 clones, a 78-bp fragment of HBV DNA corresponding to nts 1647–1725 was integrated with a 193-bp sequence of host DNA located on Ch-2p12 at nts 82086308–82086501 (Fig. 2A). This sequence was identified as the mammalian apparent retrotransposon long terminal repetitive (THE-1B-LTR) element belonging to a mammalian apparent LTR retrotransposon (MaLR) family [29]. This fusion was of the HTJ type.

Seventy one colonies were screened by colony PCR after transfection with the inv-PCR products carrying HBV DNA signals obtained at 3 h.p.i. Among them, 36 were found suitable for sequencing. The virus-host junctions were successfully identified in 10 of them and all displayed the same integration pattern (Table 1). The arrangement detected was complex and contained two HBV DNA fragments and two host DNA sequences intervolving with each other in the order: virus-host-virus-host. The first HBV-host junction was created by a 173-bp sequence of *HBx* corresponding to nts 1647-1820 and by a host DNA fragment of 38-bp located between nts 119935846-119935884 on Ch-5q23.1 (Fig. 2B). This was HTJ fusion. The host DNA fragment was homologous with the gene encoding for proline-rich protein-16 (PRR-16). The expression of this gene is known to be associated with cancer development and regulation of cell size [30]. In the second HBV-host fusion, a 79-bp HBV DNA sequence spanning nts 1866-1945 of *HBx* gene formed HTJ with a host gene encoding for protein kinase cGMP-dependent type 1 (PRKG-1). The host sequence was located on Ch-10q11.23 and aligned with nts 51769849-51769,877 (Fig. 2B). PRKG-1 is a cyclic GMP-dependent protein kinase that regulates cell growth and cell signaling, primarily in neuronal and skeletal muscle cells [31]. Further, the left-hand junction between PRR-16 and the second HBV sequence detected was formed at nt 119935884 of PRR-16 and nt 1867 of *HBx*. This was again the HTJ type fusion.

Of the 138 colonies screened by colony PCR, 61 were sequenced to detect virus-host DNA merges in cells collected at 24 h.p.i. The virus-host fusions were evident in two clones (Table 1). The HBV integrant found had a 145-bp *HBx* sequence matching nts 1647-1792 joined with a 481-bp host DNA spanning nts 35712645-35713126 on Ch-21q22.12 (Fig. 3). This was overlapping homologous junction (OHJ) formed by the 3-bp TAA sequence. The host's sequence corresponded to that encoding human runt-related transcription factor-1 (RunX-1). RunX-1 reportedly functions in hematopoiesis and has been implicated in the pathogenesis of acute myeloid leukemia and HCC [32].

The late HBV-host junctions identified at 13 days post-infection

HepG2-NTCP cells exposed to HBV which were obtained at 13 dp.i. provided a total of 120 colonies which were screened by colony PCR. Forty of them were sequenced and out of them 29 carried clearly identifiable HBV-host junctions. The virus-host fusions were identified with 9 different genes, among which two sequences belonged to retrotransposon or transposon class. To simplify presentation, the description of these junctions was arranged according to a chromosome number on which they were found. Thus, considering Ch-2, *HBx* 107-bp-long fragment spanning nts 1647-1754 was integrated at q33.1 locus (Table 1). The length of the host's sequence was 220 bp and corresponded to nts 199220298-199220518 (Fig. 4A). The fusion was 3-bp OHJ and was identical in four clones. The analysis showed that HBV integrated with hobo activator-18 *Salmo salar* long terminal repeat (hAT-18-Ssa). However, hAT-18-Ssa was joined downstream by non-coding Ch-2 sequence and that by medium reiterated frequency repeat 5B (MER-5B) forming complex host sequence trimera (Fig. 4A).

Another gene forming junction with HBV at 13 dp.i was that located on Ch-4 encoding dorsal neural tube nuclear protein (DNTNP) (Table 1). There were four clones demonstrating this fusion. The length of the host's sequence was 289 bp and corresponded to nts 1668784-1669073. The virus sequence was 18-bp-long and spanned nts 1605-1623. The fusion was in the HTJ orientation (Fig. 4B).

Two different genes on Ch-8 formed junctions with HBV DNA. The host sequences encoded phosphatidyl ethanolamine binding protein 4 (PEB-4) on p21.3 arm and primate-specific gene family with sequence similarity 90A (FAM90A) on p23.1 arm of the chromosome. There were two clones with *HBx* integration into PEB-4 gene and a singular clone with virus sequence fused with FAM90A (Table 1). In the first case, the length of the HBV sequence was 58 bp (nts 1606–1664) and the PEB-4 19-bp sequence mapped at positions 22796773-22796792 (Fig. 5A). In the second case, HBV sequence was of 20 bp (nts 1602–1622), while the FAM90 gene fragment was 214-bp-long and spanned nts 8016999-8017213 (Fig. 5B). Both fusions were HTJ.

Another HBV-host integration was detected at Ch-10q21.1. There were two clones carrying this junction (Table 1). The host's sequence was 277-bp-long and aligned with protocadherin-related 15 (PCDH15) gene at nts 476850-476573. The merge was formed by 3-bp OHJ and the overlapping nucleotides were AGG. PCDH15 is a member of the cadherin superfamily involved in calcium-dependent cell-cell adhesion [33].

Two genes with integrated HBV DNA were also identified on Ch-11 (Table 1). In the first instance, four clones with identical integration into the retrotransposon long-interspersed nuclear element 2 (LINE2) were identified (Table 1). The LINE2 542-bp long sequence was located at positions 71323803-

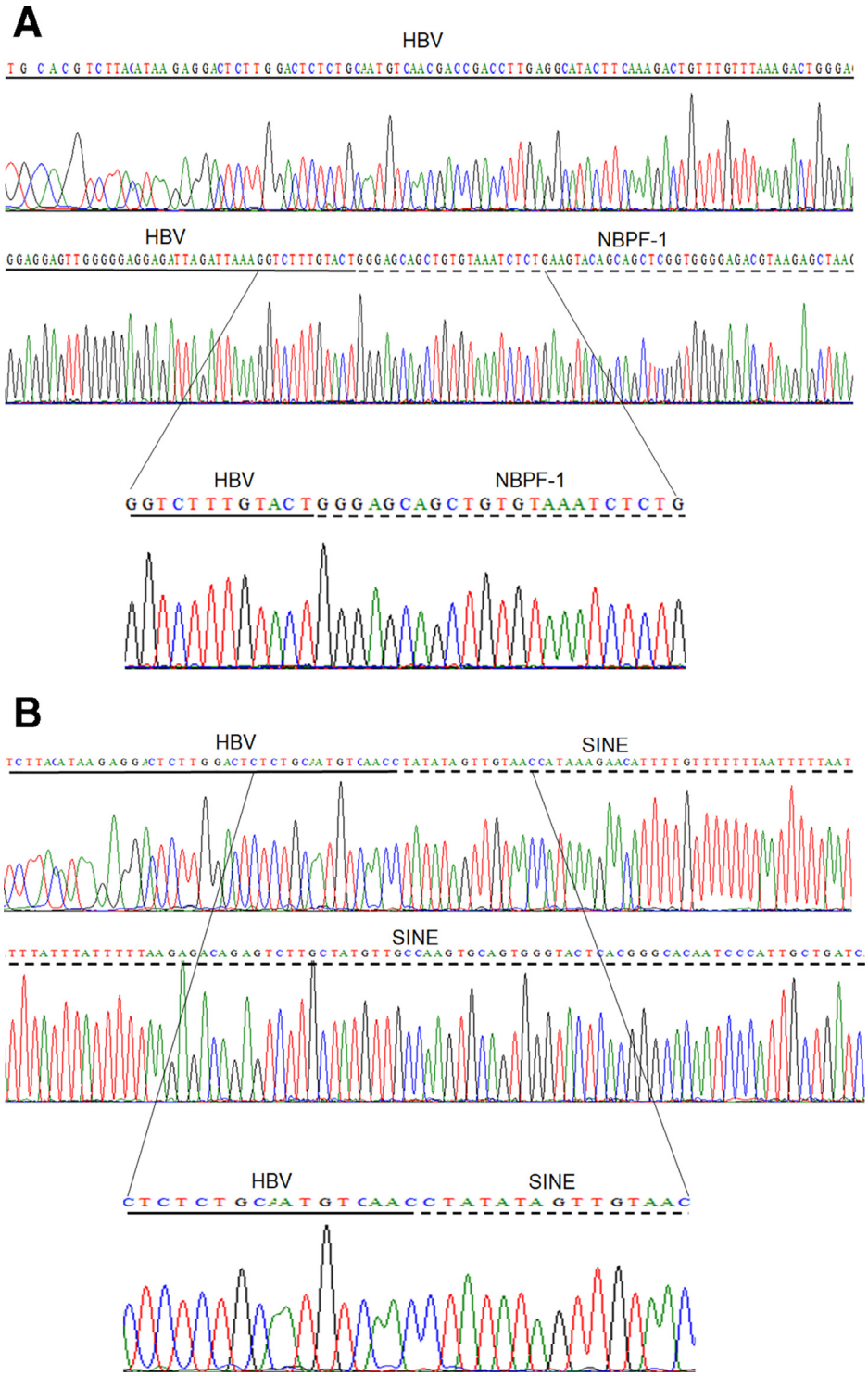

Fig. 1 Initial integration sites (IIS) of HBV identified at 30 min after exposure of HepG2-NTCP cells to HBV inoculum. **(A)** HBV junction with neuroblastoma breakpoint family member 1 (NBPF-1). **(B)** HBV DNA fusion with retrotransposon short-interspersed nuclear element (SINE). Sequencing electropherograms show *HBx* gene sequences underlined with continuous line and the host's genomic sequences underlined by dashed line. The inserts presented in lower panels detail the breaking points between HBV and host gene sequences. Both junctions shown were the head-to-tail fusions. See Table 1 for more details.

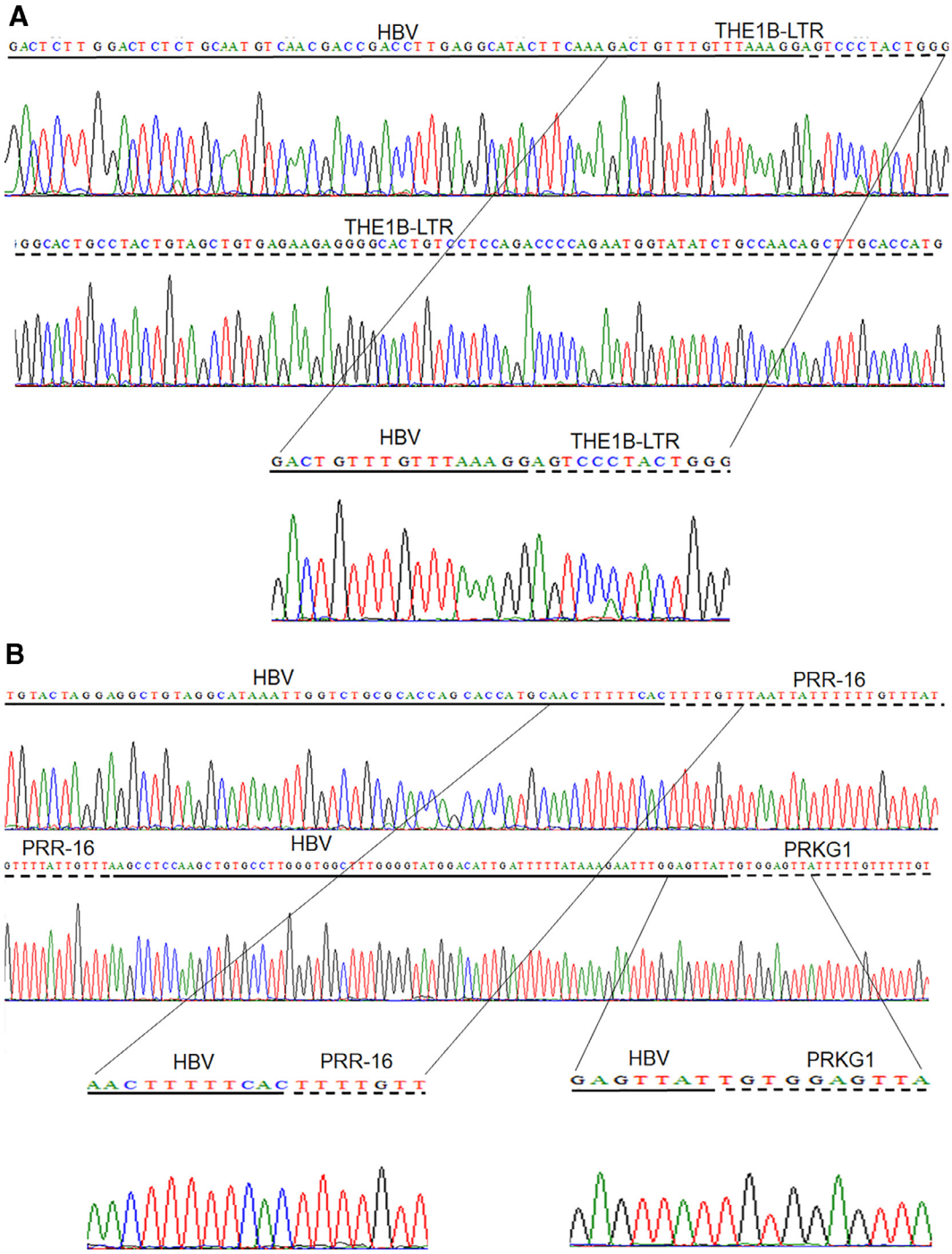

Fig. 2 Very early sites of HBV integration (VEIS) in HepG2-NTCP cells found between one and 24 h after exposure to virus. **(A)** HBV DNA head-to-tail fusion with the mammalian apparent retrotransposon long terminal repetitive (THE-1B-LTR) element located on Ch-2 and identified at one hour after exposure to HBV. **(B)** HBV integration with proline-rich 16 protein (PRR-16) and protein kinase, cGMP-dependent type I (PRKG-1) encoding genes detected at 3 h post-infection. Electropherograms detail sequences of the HBV-THE-1B-LTR and HBV-RunX-1 junctions, as well as show the complex fusion of two HBx sequences with two different genes normally located on chromosomes 5 and 10, indicating a severe host genome rearrangement. All virus-host merges forming this complex virus-host fusion were head-to-tail joins, including that between PRR-16 and the second *HBx* gene fragment detected. For more details, see legend to Fig. 1 and Table 1.

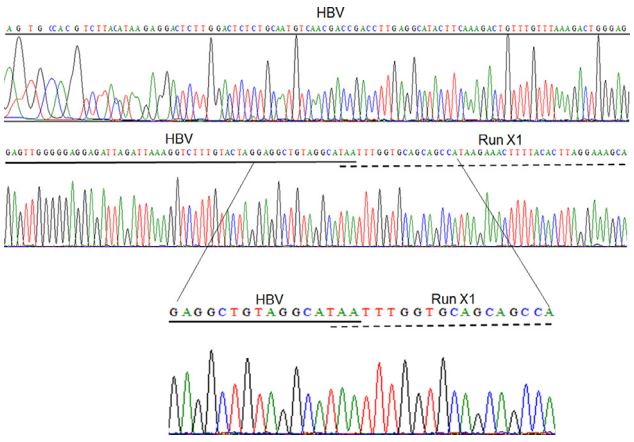

Fig. 3 HBV DNA forming an overlapping homologous junction with gene encoding runt-related transcription factor 1 (RunX-1) detected at 24 h after infection. Electropherogram showing *HBx* sequences underlined with continuous line and that of the human gene underlined with dashed lines. Additional details are in Table 1.

71323261. The *HBx* sequence was 171-bp-long and corresponded to nts 1647-1818. This fusion was a 3-bp OHJ. In the second case, a singular clone showed a 93-bp host's DNA sequence joined with a 56-bp long *HBx* fragment spanning nts 1603-1659. The host's sequence aligned to nts at positions 233-140 in the long non-coding RNA 101929653 (lnc101929653). The fusion was created by an overlapping 4-bp OHJ with AAGA sequence.

Another site at which HBV integrated at 13 dp.i. was alpha/beta hydrolase domain containing protein 12B (C14Orf29) located on Ch-14 at locus q22.1 (Table 1). The host sequence was 206-bp-long and corresponded to nts 94423774-94423980. This sequence was fused with 18-bp *HBx* fragment spanning nts 1602–1622. This was the HTJ type of junction.

Lastly, a short 12-bp host's integrant was found in 3 clones which, due to its short sequence and annealing with several human chromosomes, was classified as UI (unidentified) (Table 1). The left-hand *HBx* fragment joining the sequence was located between nts 1615–1668. The junction was in the HTJ orientation.

Analysis of chromatin marks on HepG2 nucleotide sequences targeted by HBV integration

Whether the HepG2 genes into which HBV integrated were in poised quiescent or active state should depend on accessibility of chromatin to transcription factors which, in turn, depends on modifications of histone proteins. We performed *in silico* analysis of the selected marks to find out the status of histone codes on HBV integration sites. Five different chromatin modification markers, such as CTCF, H3K4me3, EZH2, H3K27Ac and DNAse, which were linked to insulator activity, methylation, acetylation and sensitivity to DNase, respectively, were examined**.** Table 3 is detailing detection or absence of each of the marks. Analysis revealed presence of the CTCF binding clusters on 10 (71.4%) and absence in 4 (28.6%) of the HBV targeted sequences. H3K27me3 and H3K27Ac are linked to opposing states in gene transcription, i.e., methylation and acylation, respectively. Thus, based on the presence or absence of these marks, three scenarios are possible: (1) inactive state, where H3K4me3 is present and H3K27Ac is absent; (2) active state, where H3K27Ac present but H3K4me3 is absent, and (3) neutral state, where both H3K27Ac and H3K4me3 are present or absent [34]. In this context, switch from H3K27me to H3K27Ac positivity could be suggestive of a change from silenced to activated state of a given gene [35]. Based on the above scheme and the results obtained (Table 3), SINE targeted by HBV integration at 30 min p.i. was found to be only the one in active transcriptional state. Conversely, THE-1B, PRR-16, PRKG-1, RunX-1, *h*AT-18-Ssa, PEB-4 and PCDH-15 appeared transcriptionally inactive. Both the methylation and acetylation marks were absent on three targets, i.e., NBPF-1, FAM90A and lnc101929653. Furthermore, both the methylation and acetylation marks coexisted on three host sequences, i.e., DNTNP, LINE2 and C14Orf29 (Table 3). We also analyzed the presence of DNase sensitivity marks. Except for the PEB-4, DNase marks were absent on all genes targeted by HBV integration in this study. EZH2, which has prominent role in methylation [36], was present on 10 (71.4%) of the HBV-targeted genes, while absent on 4 (28.6%). The genes on which EZH2 binding sites detected were THE-1B, PRKG-1, RunX-1, hAT-18- Ssa, DNTNP, PEB-4, PCDH-15, LINE2, lnc101929653, and C14Orf29, whereas those on which these sites were absent were NBPF-1, SINE, PRR-16, and FAM90. We extended analysis by checking CTCF marks on the HBV sequence (nt 1602-1945) which integrated with HepG2 genes from 30 min to 13 day p.i. For this analysis, HBV with GenBank accession number X72702 (genotype D) served as reference. Contrary to expectation, a single CTCF binding motif of 9-bp GGCACAGCT spanning nts 1602-1945 was only detected.

Discussion

Integration of HBV into host's genome for long has been considered as an event closely coinciding with the life cycle of the virus. The recent availability of efficient *in vitro* HBV infection systems and advancements in techniques detecting with virus-host genomic junctions showed that integration is a consistent feature of infection with HBV and related viruses. Importantly, our and others recent studies showed that HBV integration occurs very early in the course of infection and it could be identified even earlier than detection of virus covalently closed circular DNA (cccDNA) that is the canonical marker of hepadnaviral replication [9,10,37]. In addition, it was uncovered that even infection with very small quantities of hepadnavirus, i.e., 10 or 100 virions, leads to virus DNA integration into host's genome [11].

The primary focus of this study was to identify the initial sites of HBV integration and the time of their formation in HepG2-NTCP cells which demonstrated superior sensitivity to HBV infection and efficient HBV replication comparing to the previous *in vitro* models. Our results showed that HBV integrates as early as in 30 min p.i. in these cells. In general, this finding is consistent with our previous study in which hepadnavirus-host DNA junctions became detectable in one hour p.i. in both HepaRG cells exposed to patient-derived

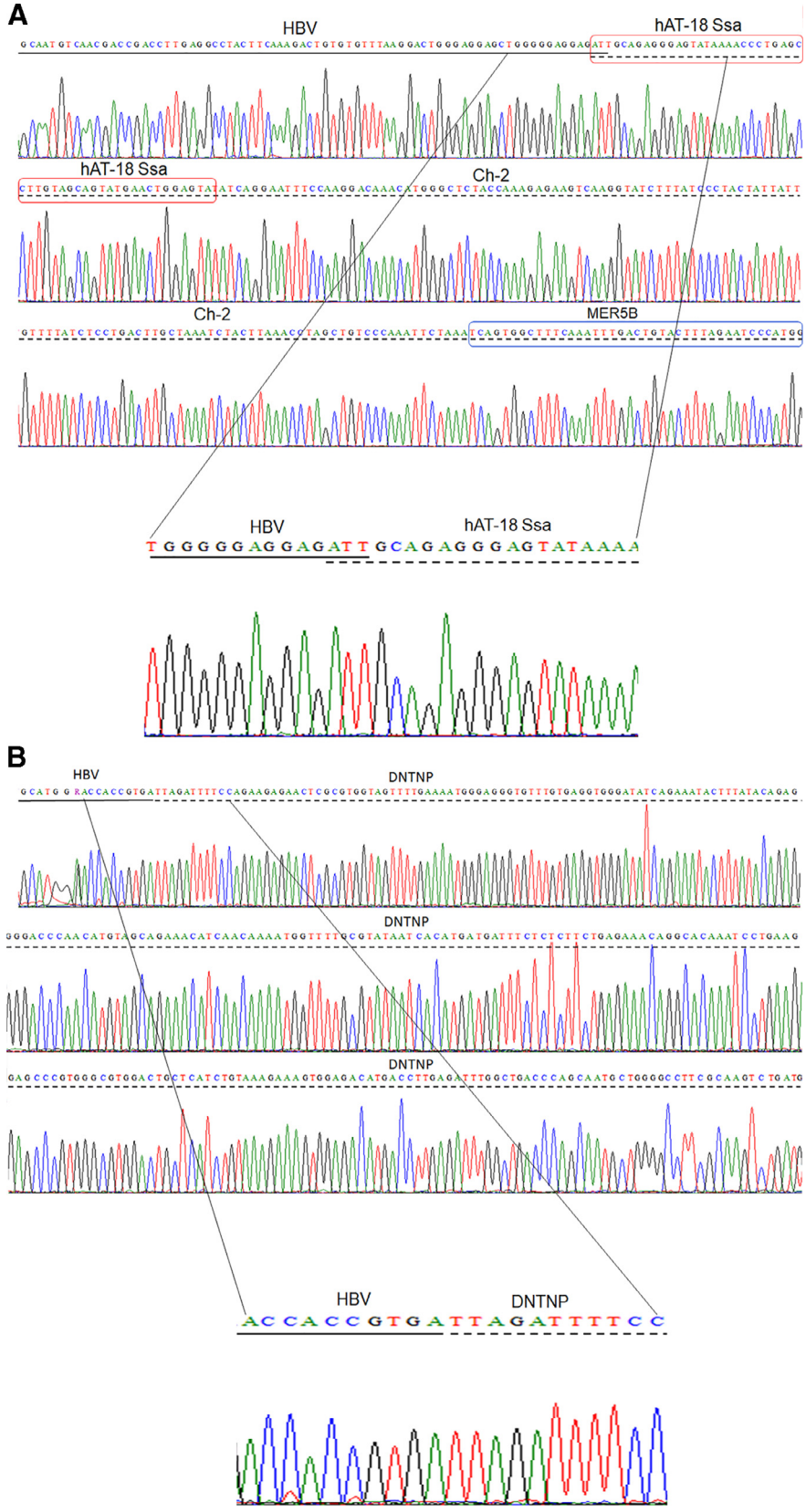

Fig. 4 Examples of the late or not early HBV-host genome integrations (NEIS) detected in HepG2-NTCP cells at 13 days post-infection. **(A)** HBV DNA junction with the hobo activator *Salmo salar*-18 long terminal repeat (hAT-18 Ssa) fused with retrotransposon element called medium reiteration frequency interspersed repeat 5B (MER-5B) (sequence in box). **(B)** HBV junction with gene encoding dorsal neural-tube nuclear protein (DNTNP). Electropherograms showing *HBx* sequences underlined with continuous line and those of the human genes underlined with dashed lines. Both junctions were of the HTJ type. See Table 1 for more details.

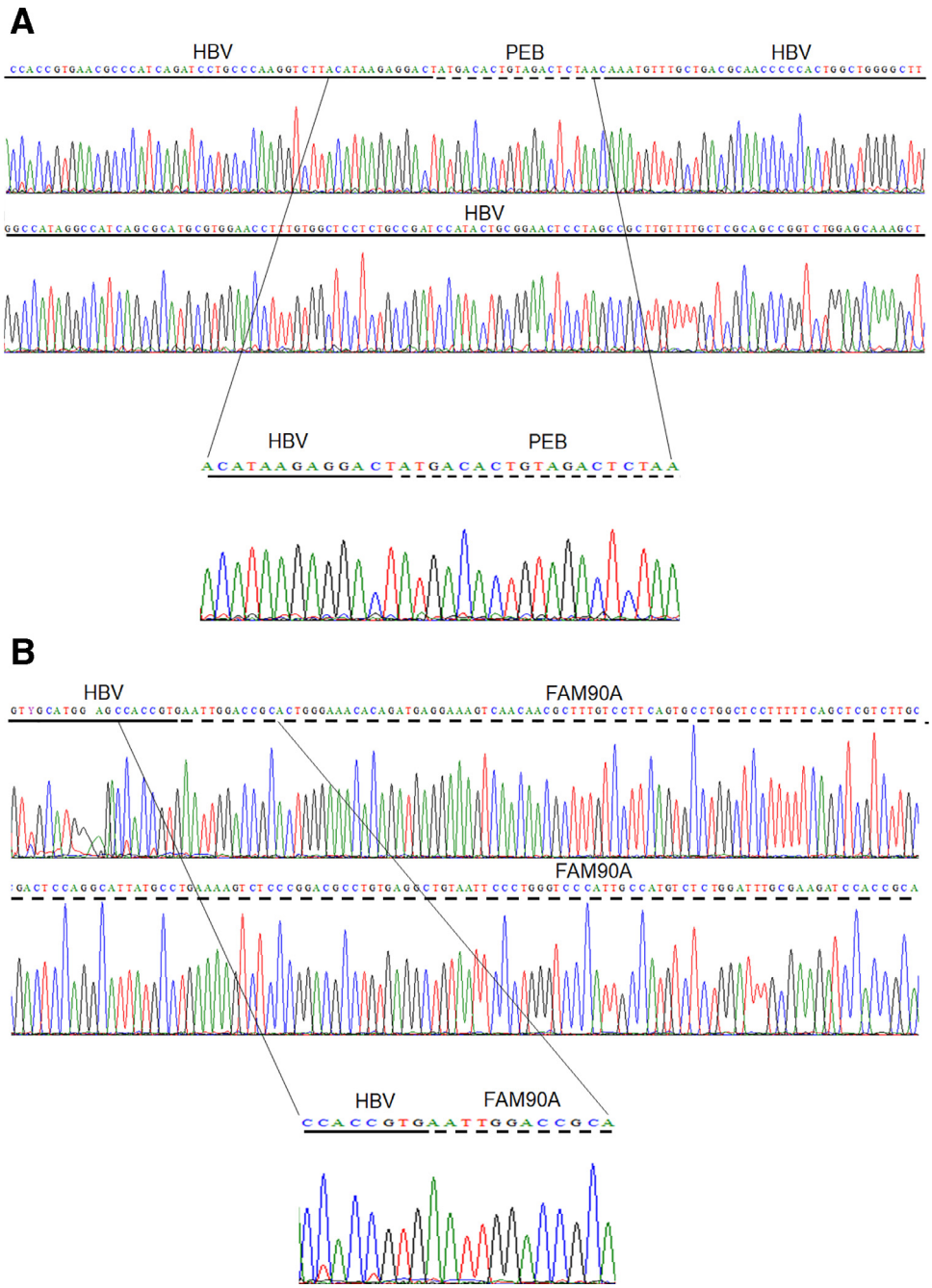

Fig. 5 Other examples of the late or not early HBV-host genome integrations (NEIS) detected in HepG2-NTCP cells at 13 days post-infection. **(A)** HBV integration with gene encoding phosphatidyl ethanolamine binding protein 4 (PEB-4). **(B)** HBV integration with primate-specific gene family with sequence similarity 90A (FAM90A) encoding gene. Electropherograms showing *HBx* sequences underlined with continuous line and those of the human genes underlined with dashed lines. Both junctions were of the HTJ type. See Table 1 for additional details.

HBV and woodchucks *de novo* infected with wild-type WHV [9]. However, twice shorter time period from the first contact with virus to detection of virus-host fusions suggests that the HepG2-NTCP-HBV infection system is better predisposed to facilitate virus integration than natural infection in the woodchuck model. The second major finding of our study was that HBV initially preferably integrates into the host non-coding "jumping" elements, especially into retrotransposons. This is not entirely surprising as nearly 50% of the human genome is constituted by the retrotransposon/transposon sequences [38].

Considering the initial integration sites detected at 30 min p.i., SINE sequence and NBPF-1 gene were identified as the sites of HBV DNA insertion. SINE belongs to the retrotransposon non-LTR category. Importantly, SINE sequences constitute over one tenth of human genome [39]. The recent

reports suggest that small transposons like SINE may have significant influence on the overall size of genome [40] and may drive oncogenic transformation [41]. Regarding NBPF-1, it is a transcription factor and a suppressor linked to neuroblastoma [42]. NBPF-1 protein derives from a translocation from Ch-1p36.2 to Ch-17q11.2 that coincides with a 5215-bp deletion in Ch-17. Interestingly, the breakpoints in each chromosome are located within the long interspersed nuclear elements [43]. Thus, the breakpoint in Ch-1 is located within the LINE1 retrotransposon and that in Ch-17 within LINE2. When we reanalyzed our data for a sequence similarity with LINE1 or LINE2, we didn't found any homology. It is of note that HBV integration into neuroblastoma genome has been reported [44].

At one hour p.i., HBV formed junction with THE-1B-LTR belonging to a MaLR family of the LTR retrotransposons. Origin of THE-1B-LTR is ancestral and was found at high copy numbers in both new and old world primates [45]. The pro-carcinogenic role of THE-1B-LTR has been postulated in non-Hodgkin's lymphoma (NHL) where its hypomethylation activates colony stimulating factor 1 receptor (CSF1R), which acts as a tyrosine kinase receptor [45]. The ectopic expression of CSF1R protein has been detected in 39-48% of patients with NHL [45]. Considering a possible pathogenic relevance of the HBV-THE-1B-LTR fusion, the merged HBV sequence spanned nts 1647-1725, which covers enhancer II (Enh-II) region of virus genome. Enh-II regulates activity of virus pregenomic RNA and virus surface (S) gene promoter, as well as it has long-range regulatory effects. In this regard, Enh-II is recognized by transcription factors, such as hepatocyte nuclear factor-1 and −4 enhancer II (Enh-II) [46]. Therefore, HBV Enh-II when fused with THE-1B-LTR may modulate activity of the retrotransposon and exert hepatocyte-specific regulatory effects. In this context, a relatively high incidence of HBV infection in patients with NHL [47] and data suggesting that HBV can enter B cells and infect them [48] could be interpreted in support of a pathogenic role for the HBV EnhII-THE-1B-LTR fusion at the extrahepatic location.

At three hours post-infection, an interesting pattern of the complex HBV-host genome integration was identified. The merge had the HBV-host-HBV-host sequence order that appeared as a single continuous nucleotide stretch which was identical in 10 separate clones. The stretch contained two host sequences encoding proline-rich-region-16 (PRR-16) and cGMP-dependent protein kinase 1 (PRKG-1) proteins. Regarding PRR-16, the protein family containing ankyrin repeat, SH3 domain, and proline-rich-region (i.e., ASPP family) was recently identified as apoptosis regulation proteins and have role in increasing cell size [49]. These proteins are frequently down-regulated due to DNA methylation in HBV infection-associated HCC [50]. Considering PRKG-1, proteins from this family play central role in regulating cardiovascular and neuronal functions [51]. PRKG-1 phosphorylation leads to perturbation of cellular calcium, specially lowering of intracellular free calcium levels, controlling vesical trafficking and myosin light chain phosphorylation. It was reported that deletion of PRKG-1 resulted in complex metabolic phenotype including liver inflammation and hyperglycemia [31].

HBV DNA integration with RunX-1 encoding gene was identified at 24 h.p.i. The HBV sequence forming this junction corresponded to nts 1647-1754 and, similar to that merging with THE-1B-LTR at one hour p.i., covered virus Enh-II which may play a part in regulating and enhancing expression of this gene. Although we have not checked if expression of RunX-1 was upregulated, it is possible that it was due to fusion with Enh-II. RunX-1 is a transcription factor and a dominant oncogene [52]. It is required for the generation and maintenance of hematopoietic stem cells phenotype and their differentiation into distinct lineages [53]. Importantly, RunX-1 has been identified as a target of leukemogenic chromosomal translocations in acute myelogenous leukemia [54]. In this context, chromosomal translocations and inverted duplications frequently accompany integration of HBV DNA in HCC [55]. Further, avian leukosis virus (ALV) induces B cell lymphoma by integrating at several host chromosomal sites, including RunX-1 [56]. Although, currently there is no data suggesting role of RunX-1 in carcinogenesis coinciding with HBV infection, there are reports showing that RunX-3 gene can be hypermethylated in HBV-infected HCC patients [57]. Contradictory to RunX-1, RunX-3 is a tumor suppressor gene and have been found to be downregulated in HCC and RunX 1-3 genes are the key player in cancer metastasis [58]. Interestingly, it has been shown that RunX3 plays pivotal role in LINE1 transcription and retrotransposition [59]. Therefore, it would be interesting to know if expression of HBV EnhII-RunX-1 fusion could influence a metastatic potential of HBV-related HCC.

Although the focus of the current research was to decipher formation of the virus-host junctions soon after the first contact of virus with susceptible cell, we also analyzed one later-time point to assess dynamic of the merges and if the junctions detected at early time points persisted. Thus, at 13 day p.i., nine different junctions were identified and there was no similarity between them and those found in the first 24 h.p.i., except that host sequences in two of them represented retrotransposon or transposon. We limited discussion of the HBV-host fusions uncovered at 13 dp.i. to three most interesting examples.

One of the most striking HBV-host junctions identified at 13 day p.i. was complex fusion of HBV with a trimeric host sequence constituted by hAT-18-Ssa, noncoding Ch-2 sequence, and MER-5B (see Fig. 4A). Both hAT-18-Ssa and MER-5B belong to the transposon class. Although hAT-18-Ssa-MER-5B fusion has been reported in one *in silico* study [60], this joint was separated by the Ch-2 sequence insert in our data. There are approximately 70 different types of hAT elements in human and they transpose via DNA-DNA fusions, instead of via DNA-RNA merges as generally seen in retrotransposons [61]. Following hAT-18 Ssa was Ch-2 sequence which aligned to q33.1 arm of the chromosome at positions 199220298-199220518. The third element of the triad was MER-5B that is known to play a role in controlling expression of alpha fetoprotein (AFP) gene, which protein is abundantly expressed in human fetal liver [62]. Interestingly, it has been shown that persistence of AFP and H19 expression in liver can be due to the retroviral insertion [63]. In this context, we have previously documented WHV DNA integration into H19 gene [11]. These indirect evidence may together suggest a connection between elevated AFP levels and hepadnaviral insertional events, and future studies may explain molecular foundation of this relation.

Another gene into which HBV integrated at 13 dp.i. was PEB-4 gene which encodes an anti-apoptotic protein [64]. Its anti-apoptotic function was proven in prostate cancer, breast cancer, as well as liver cancer cell lines. Although PEB-4 is ubiquitously expressed in most of the human tissues, its

Table 2 Comparison of the findings on HBV-host genome integrations identified in the current study with those from two previous *in vitro* studies.

Gene family	HepG2-NTCP-C4 (Current study)	HepG2-NTCT or Huh7-NTCP (Ref. [10])	HepaRG (Ref. [9])	Comparison
Retrotransposons/transposons				
SINE	SINE (30 min)	SINE (5 d,7 d)	n.a.	Common
THE	THE-1B (1 h)	THE-Int (5 d)	n.a.	Similar
MER	MER-5B (13 d)	MER52D/41A/90A/4E1/4A (7 d)	n.a.	Similar
LINE	LINE-2 (13 d)	LINE-1 (1 d - 7 d)	LINE-1 (1 d), LINE-2 (3 d), FLRT2-LINE-2 (1 d)	Common
Other genes				
NBPF	NBPF-1 (30 min)	-n.a.	NBPF25P (14 d)	Similar
PCDH	PCDH-15 (13 d)	PCDH-15 (7 d)	n.a.	Common
FAM	FAM90A (13 d)	FAM85B/183B (7 d)	n.a.	Similar
C-Orf	C14Orf29 (13 d)	C1Orf27/C9Orf72/C20Orf196 (7 d)	n.a.	Similar

Abbreviations: min, minutes; h, hour; d, day; n.a., not applicable (not found); SINE, short interspersed nuclear element; THE, mammalian apparent LTR retrotransposon; MER, medium reiteration frequency interspersed repeat; LINE, long interspersed nuclear element; NBPF, neuroblastoma family member; PCDH, protocadherin; FAM, family with sequence similarity; C-Orf, chromosome-open reading frame.

high expression was observed in tumorous part of the tissues. It was demonstrated that PEB-4 promotes cellular resistance to TNFα-induced apoptosis by inhibiting activation of Raf-1/MEK/ERK pathway [65]. Using the hepatocyte-like cell line, HepG2, it was also shown PEB-4 inhibited growth of HepG2 cells in dose and time-dependent manner by increasing caspase 3 expression [66]. Others found that PEB-4 is expressed in 90% of patients with B-cell lymphoma, but only 16.7% in normal lymph nodes [67]. Interestingly, PEB-4 over-expression inhibited Rituximab-mediated complement dependent toxicity, while silencing of PEB-4 augmented efficacy of Rituximab both *in vitro* and *in vivo* [67].

PCDH15 was yet another gene into which HBV found to be integrated at 13 dp.i. (see Table 1). Several reports documented role of protocadherin in the development of HCC. The protocadherin identified were PCDH-8, PCDH-9, PCDH-10, PCDH-11, PCDH-15 and PCDH-20. One report detailed association of decreased expression of PCDH-20 with poor prognosis of HCC [68]. PCDH-8 was frequently found inactivated due to promoter hypermethylation in liver cancer [69]. PCDH-9 inhibits epithelial-mesenchymal transition and cell migration through activation of GSK-3β in HCC [70]. Loss of PCDH-17 appears to promote metastasis and invasion through hyperactivation of EGFR/MEK/ERK signaling pathway in HCC [71]. PCDH-10 gene inhibits cell proliferation and induces cell apoptosis by inhibiting the PI3K/Akt signaling pathway in HCC [72]. Using the uPA-human hepatocytes engrafted chimeric mice infected with HBV, the recent report has identified HBV integration into PCDH-11 gene [73]. Since this study identified the timing of HBV integration as late it may collaborate with our finding of HBV integration with the PCDH-15 gene at the later time point post-infection.

Taking under consideration the above finding, we compared the data from this study with those from two previous studies in which HBV integration into host genome was investigated soon after *de novo* infection *in vitro* [9,10]. Eight different genes were found to be comparable or belonged to the same gene family among those identified in both our work and at least one of the two studies (Table 2). Among them, four belonged to the retrotransposon/transposon class and included genes within SINE, THE, MER and LINE families, and other four corresponded to genes belonging to NBPF, PCDH, FAM and C-Orf families. There were two genes found to be HBV IIS in the current study, i.e., NBPF-1 and SINE. It is of note that the gene belonging to the same family as NFBF-1, i.e. NBPF-25P, was identified as the HBV integration site at 14 dp.i. in our previous study utilizing HepaRG cells as infection targets [9]. As well, SINE was found as HBV integration site in HBV-infected Huh7-NTCP cells from 5 dp.i. onwards by another group [10]. Overall, there was relatively high degree of uniformity in the findings from three independent studies in which *in vitro* infection systems utilizing either patient-derived or recombinant HBV produced in culture were used as inocula. This ascertained authenticity of the observations made.

Recognition of the nature of epigenetic modifications at the earliest time points post-infection when HBV integrate into human genome could foreshadow whether those viral hits might be transcriptionally active or latent. Accessibility of chromatin to transcription factors depends of the histone marks H3K4me3 and H3K27ac as their activity peak in early phases of the transcriptional elongation [74–76]. Enrichment of H3K4 methylation marks is characteristic of transcriptionally repressed areas, whereas presence of H3K27 is attribute of transcriptionally active state [77]. In addition to the H3K4me3 and H3K27ac, we expanded analysis to find additional marks which may modify genome organization and its functionality, and they were CTCF, EZH2 and DNase [78,79]. As presented in Table 3, the analysis showed commonalities in the status of H3K4me3, EZH2 and DNAse marks at 30 min p.i., however there was a difference in the presence of H3K27ac. In this context, since SINE sequence demonstrated absence of methylation mark but presence of H3K27ac, this may imply that the sequence was transcriptionally active. From one to 24 h p.i., H3K4me3 was detected and acetylation mark H3K27Ac was absent. Accordingly, this profile may imply that the sequences forming junctions with HBV DNA during this time period were transcriptionally inactive. At 13 day p.i., the acetylation marks were either not detectable or accompanied by the methylation marks suggesting that these host's sequences were in the latent state considering transcription.

CTCF is a master weaver of the genome and has regulatory functions [80]. CTCF forms complex topological 3D

Table 3 *In silico* identification of chromatin motifs in sequences of the HepG2 cell genes found to be the sites of HBV integration.

Time post infection	Gene	CTCF	H3K4me3	EZH2	H3K27Ac	DNase
30 min	NBPF-1	Present (1)	Absent	Absent	Absent	Absent
	SINE	Present (3)	Absent	Absent	Present	Absent
1 h	THE1B-LTR	Present (1)	Present	Present	Absent	Absent
3 h	PRR-16	Absent	Present	Absent	Absent	Absent
	PRKG-1	Present (2)	Present	Present	Absent	Absent
24 h	RunX-1	Present (3)	Present	Present	Absent	Absent
13 days	hAT-18 Ssa	Present (1)	Present	Present	Absent	Absent
	DNTNP	Present (1)	Present	Present	Present	Absent
	PEB-4	Absent	Present	Present	Absent	Present
	FAM90	Absent	Absent	Absent	Absent	Absent
	PCDH-15	Present (1)	Present	Present	Absent	Absent
	LINE2	Present (2)	Present	Present	Present	Absent
	lncRNA	Absent	Absent	Present	Absent	Absent
	C14Orf29	Present (2)	Present	Present	Present	Absent

Abbreviations: min, minutes; h, hour; CCCTC-binding factor, CTCF; Enhancer of Zeste homolog 2, EZH2; Tri-methylation of lysine 4 on histone H3, H3K4me3; Histone H3 lysine 27 demethylase, H3K27ac; Deoxyribonuclease, DNase.
Numbers in parenthesis in the CTCF column represent numbers of CTCF binding clusters detected in a given sequence.

structures which are central to the long-range gene activation and have role in cell differentiation [81]. Its regulatory functions includes transcriptional activation or repression, insulation, imprinting and X chromosome inactivation [81]. As CTCF is a sequence-specific DNA binding protein, its binding site can be altered by DNA methylation [81]. Therefore, our analysis of CTCF binding clusters and interpretation of its possible activity included determination of H3Kme3. As shown in Table 3 and Supplementary Table 1, the CTCF binding motifs were found on almost all host sequences detected up to 24 h p.i., except PRR-16. Depending on the location of CTCF binding sites in relation to methylation marks, their existence may indicate a regulatory function. This might be particularly relevant in the case of NBPF-1 and SINE which sequences carried CTCF clusters but not H3K4me3 marks. At 13 days p.i., CTCF binding sites were present on hAT-18-SSa, DNTNP, PCDH-15, LINE2 and C14Orf29, and absent on PEB-4, FAM90 and lnc101929653. However, all five sequences with CTCF clusters also carried methylation marks (Table 3).

Regarding of DNase hypersensitive sites, our analysis revealed their absence on all the sequences examined except PEB-4 (Table 3). This may suggest that the sequence fragments identified were in fact closed for chromatin modifications. Going further, we also determined presence of EZH2 marks, as it is a main catalytic subunit of polycomb repressive complex 2 (PRC2) which plays important role in methylation by activating methyl groups [82]. Notably, we could not find EZH2 binding sites at the initial time of HBV integration, i.e., at 30 min p.i. The absence of both EZH2 and H3K4me3 marks on the initial HBV integration targets may suggest a strict control of their methylation and upholding them transcriptionally active. Since the analysis was done *in silico* on HepG2 genome and not on the genome of HepG2-NTCP-C4 cells, which is not yet determined, further studies might be required to confirm our results. Nevertheless, the recent findings progressively illuminate a role of key chromatin marks which may govern HBV driven initiation of hepatocyte oncogenic transformation finally leading to the development of liver cancer.

The results from this study indicate that retrotransposon elements are frequent among first-hit sites of HBV integration. This may suggest a mechanism by which HBV DNA may spread across host's genome from the earliest stages of hepadnaviral infection, which could involve transposition of the fused HBV-retrotransposon sequences across chromosomes [83]. The results also showed that HBV integrates as early as in 30 min p.i. into genome of HepG2-NTCP-C4 cells. This was at least twice shorter time period than that found in the previous *in vitro* studies utilizing HepaRG cell line and in woodchucks *de novo* infected with WHV [9]. This indicates that the HBV-HepG2-NTCP infection model could be better predisposed to facilitate HBV-host genome fusions likely due to more efficient uptake of virus via NTCP receptor. We also delineated profiles of the most important chromatin marks on the HepG2 genomic sequences targeted by HBV integrations. These marks may significantly influence organization and transcriptional activity of the host's sequences affected by HBV and their recognition should contribute with time to our better understanding of the mechanisms underlying HBV driven oncogenesis. The wealth of the data on the HBV-host genome merges occurring from the earliest stages of virus infection opens new opportunities for investigation on the functional significance of virus-host interactions that predispose to the particular long-term pathological outcomes in HBV-infected patients.

Acknowledgments

This study was supported by an operating grant (PIN 22346) from the Cancer Research Society Inc., Canada awarded to T.I.M. T.I.M. was a recipient of the Senior (Tier 1) Canada Research Chair in Viral Hepatitis/Immunology sponsored by the Canada Research Chair Program and funds from the Canadian Institutes of Health Research, the Canada Foundation for Innovation, and Memorial University, St. John's, NL, Canada. This work was also supported in part by Grants-in-Aid from the Research Program on Hepatitis from Japan Agency for Med-

ical Research and Development (AMED: 18fk0310101j1102 to M.F.).

Disclosure statement

The authors have no conflict of interest.

Conflict of interest statement

The authors have no conflict of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cancergen.2019.04.060.

References

[1] Baumert TF, Thimme R, von Weizsacker F. Pathogenesis of hepatitis B virus infection. World J Gastroenterol 2007;13:82–90.
[2] Kim GA, Lim YS, Han S, Choi J, Shim JH, Kim KM, Lee HC, Lee YS. High risk of hepatocellular carcinoma and death in patients with immune-tolerant-phase chronic hepatitis B. Gut 2018;67:945–52.
[3] Wang M, Xi D, Ning Q. Virus-induced hepatocellular carcinoma with special emphasis on HBV. Hepatology Int 2017;11:171–80.
[4] Chauhan R, Lingala S, Gadiparthi C, Lahiri N, Mohanty SR, Wu J, Michalak TI, Satapathy SK. Reactivation of hepatitis B after liver transplantation: current knowledge, molecular mechanisms and implications in management. World J Hepatol 2018;10:352–70.
[5] Mason WS, Gill US, Litwin S, Zhou Y, Peri S, Pop O, Hong ML, Naik S, Quaglia A, Bertoletti A, Kennedy PT. HBV DNA integration and clonal hepatocyte expansion in chronic hepatitis B patients considered immune tolerant. Gastroenterology 2016;151:986–98.
[6] 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.
[7] 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.
[8] 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. New Engl J Med 1981;305:1067–73.
[9] 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.
[10] 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 NTCP-dependent uptake of enveloped virus particles. J Virol 2018. doi:10.1128/JVI.02007-17.
[11] 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 Path 2014;10:e1004332.
[12] 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;84:8308–15.
[13] Murakami Y, Saigo K, Takashima H, Minami M, Okanoue T, Brechot C, Paterlini-Brechot P. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. Gut 2005;54:1162–8.
[14] Toh ST, Jin Y, Liu L, Wang J, Babrzadeh F, Gharizadeh B, Ronaghi M, Toh HC, Chow PK, Chung AY, Ooi LL, Lee CG. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. Carcinogenesis 2013;34:787–98.
[15] Yan H, Liu Y, Sui J, Li W. NTCP opens the door for hepatitis B virus infection. Antivir Res 2015;121:24–30.
[16] Iwamoto M, Watashi K, Tsukuda S, Aly HH, Fukasawa M, Fujimoto A, Suzuki R, Aizaki H, Ito T, Koiwai O, Kusuhara H, Wakita T. Evaluation and identification of hepatitis B virus entry inhibitors using HepG2 cells overexpressing a membrane transporter NTCP. Bioch Biophys Res Comm 2014;443:808–13.
[17] Pollicino T, Belloni L, Raffa G, Pediconi N, Squadrito G, Raimondo G, Levrero M. Hepatitis B virus replication is regulated by the acetylation status of hepatitis B virus cccDNA-bound H3 and H4 histones. Gastroenterology 2006;130:823–37.
[18] Tropberger P, Mercier A, Robinson M, Zhong W, Ganem DE, Holdorf M. Mapping of histone modifications in episomal HBV cccDNA uncovers an unusual chromatin organization amenable to epigenetic manipulation. Proc Natl Acad Sci USA 2015;112:E5715–24.
[19] Guo Y, Li Y, Mu S, Zhang J, Yan Z. Evidence that methylation of hepatitis B virus covalently closed circular DNA in liver tissues of patients with chronic hepatitis B modulates HBV replication. J Med Virol 2009;81:1177–83.
[20] Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nature Rev Gen 2016;17:487–500.
[21] Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. Ann Rev Biochem 1988;57:159–97.
[22] Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F. The polycomb group protein EZH2 directly controls DNA methylation. Nature 2006;439:871–4.
[23] Umetsu T, Inoue J, Kogure T, Kakazu E, Ninomiya M, Iwata T, Takai S, Nakamura T, Sano A, Shimosegawa T. Inhibitory effect of silibinin on hepatitis B virus entry. Bioch Biophys Rep 2018;14:20–5.
[24] Sells MA, Chen ML, Acs G. Production of hepatitis B virus particles in Hep G2 cells transfected with cloned hepatitis B virus DNA. Proc Natl Acad Sci USA 1987;84:1005–9.
[25] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genom Rev 2002;12:996–1006.
[26] Davis C.A., Hitz B.C., Sloan C.A., Chan E.T., Davidson J.M., Gabdank I., Hilton J.A., Jain K., Baymuradov U.K., Narayanan A.K., Onate K.C., Graham K., Miyasato S.R., Dreszer T.R., Strattan J.S., Jolanki O., Tanaka F.Y., Cherry J.M. The encyclopedia of DNA elements (ENCODE): data portal update. 2018;46:D794–D801.
[27] Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. Nucl Acids Res 2013;41:D188.
[28] 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.
[29] Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, Kochert K, Bouhlel MA, Richter J, Soler E, Stadhouders R, Johrens K, Wurster KD, Callen DF, Harte MF, Giefing M, Barlow R, Stein H, Anagnostopoulos I, Janz M, Cockerill PN, Siebert R, Dorken B, Bonifer C, Mathas S. Derepression of

an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat Med 2010;16:571–9 (one page following 579).
[30] Yamamoto K, Gandin V, Sasaki M, McCracken S, Li W, Silvester JL, Elia AJ, Wang F, Wakutani Y, Alexandrova R. Largen: a molecular regulator of mammalian cell size control. Mol Cell 2014;53:904–15.
[31] Lutz SZ, Hennige AM, Feil S, Peter A, Gerling A, Machann J, Krober SM, Rath M, Schurmann A, Weigert C, Haring HU, Feil R. Genetic ablation of cGMP-dependent protein kinase type I causes liver inflammation and fasting hyperglycemia. Diabetes 2011;60:1566–76.
[32] Miyagawa K, Sakakura C, Nakashima S, Yoshikawa T, Kin S, Nakase Y, Ito K, Yamagishi H, Ida H, Yazumi S, Chiba T, Ito Y, Hagiwara A. Down-regulation of RUNX1, RUNX3 and CBFbeta in hepatocellular carcinomas in an early stage of hepatocarcinogenesis. Anticancer Res 2006;26:3633–43.
[33] Senften M, Schwander M, Kazmierczak P, Lillo C, Shin JB, Hasson T, Geleoc GS, Gillespie PG, Williams D, Holt JR, Muller U. Physical and functional interaction between protocadherin 15 and myosin VIIa in mechanosensory hair cells. J Neurosci 2006;26:2060–71.
[34] Bannister AJ, Kouzarides T. Regulation by chromation by histone modifications. Cell Res 2011;21:381–95.
[35] Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 2011;470:279–83.
[36] Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev 2002;16:2893–905.
[37] Ko C, Chakraborty A, Chou WM, Hasreiter J, Wettengel JM, Stadler D, Bester R, Asen T, Zhang K, Wisskirchen K, McKeating JA, Ryu WS, Protzer U. Hepatitis B virus (HBV) genome recycling and de novo secondary infection events maintain stable cccDNA levels. J Hepatol 2018;69:1231–41.
[38] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 2011;13:36–46.
[39] 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.
[40] Naville M, Henriet S, Warren I, Sumic S, Reeve M, Volff JN, Chourrout D. Massive Changes of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. Curr Biol 2019;29:1161–8.
[41] Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, Zhang D, Li D, Xing X, Kim S, O'Donnell D, Gordon JI, Wang T. Transposable elements drive widespread expression of oncogenes in human cancers. Nat Genetics 2019;51:1611–68.
[42] 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.
[43] Vandepoele K, Andries V, Van Roy N, Staes K, Vandesompele J, Laureys G, De Smet E, Berx G, Speleman F, van Roy F. 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.
[44] Tagieva NE, Gizatullin RZ, Zakharyev VM, Kisselev LL. A genome-integrated hepatitis B virus DNA in human neuroblastoma. Gene 1995;152:277–8.
[45] Martin-Moreno AM, Roncador G, Maestre L, Mata E, Jimenez S, Martinez-Torrecuadrada JL, Reyes-Garcia AI, Rubio C, Tomas JF, Estevez M, Pulford K, Piris MA, Garcia JF. CSF1R protein expression in reactive lymphoid tissues and lymphoma: its relevance in classical Hodgkin lymphoma. PloS One 2015;10:e0125203.
[46] Yuh CH, Ting LP. Differentiated liver cell specificity of the second enhancer of hepatitis B virus. J Virol 1993;67:142–9.
[47] Wang F, Xu RH, Luo HY, Zhang DS, Jiang WQ, Huang HQ, Sun XF, Xia ZJ, Guan ZZ. Clinical and prognostic analysis of hepatitis B virus infection in diffuse large B-cell lymphoma. BMC Cancer 2008;8:115–26.
[48] Wang Y, Wang H, Pan S, Hu T, Shen J, Zheng H, Xie S, Xie Y, Lu R, Guo L. Capable infection of hepatitis B virus in diffuse large B-cell lymphoma. J Cancer 2018;9:1575–81.
[49] Chen R, Wang H, Liang B, Liu G, Tang M, Jia R, Fan X, Jing W, Zhou X, Wang H, Yang Y, Wei H, Li B, Zhao J. Downregulation of ASPP2 improves hepatocellular carcinoma cells survival via promoting BECN1-dependent autophagy initiation. Cell Death Dis 2016;7:e2512.
[50] Zhao J, Wu G, Bu F, Lu B, Liang A, Cao L, Tong X, Lu X, Wu M, Guo Y. Epigenetic silence of ankyrin-repeat-containing, SH3-domain-containing, and proline-rich-region- containing protein 1 (ASPP1) and ASPP2 genes promotes tumor growth in hepatitis B virus-positive hepatocellular carcinoma. Hepatology 2010;51:142–53.
[51] Pfeifer A, Kilic A, Hoffmann LS. Regulation of metabolism by cGMP. Pharmacol Ther 2013;140:81–91.
[52] Cameron ER, Blyth K, Hanlon L, Kilbey A, Mackay N, Stewart M, Terry A, Vaillant F, Wotton S, Neil JC. The Runx genes as dominant oncogenes. Blood Cell Molecul Dis 2003;30:194–200.
[53] Dzierzak E, Speck NA. Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nat Immunol 2008;9:129–36.
[54] Okuda T, Cai Z, Yang S, Lenny N, Lyu CJ, van Deursen JM, Harada H, Downing JR. Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. Blood 1998;91:3134–43.
[55] Nakamura T, Tokino T, Nagaya T, Matsubara K. Microdeletion associated with the integration process of hepatitis B virus DNA. Nucl Acid Res 1988;16:4865–73.
[56] Justice JF, Morgan RW, Beemon KL. Common viral integration sites identified in avian leukosis virus-Induced B-cell lymphomas. mBio 2015;6 e01863-15.
[57] Poungpairoj P, Whongsiri P, Suwannasin S, Khlaiphuengsin A, Tangkijvanich P, Boonla C. Increased oxidative stress and RUNX3 hypermethylation in patients with hepatitis B virus-associated hepatocellular carcinoma (HCC) and induction of RUNX3 hypermethylation by reactive oxygen species in HCC cells. Asian Pac J Cancer Prev. 2015;16:5343–8.
[58] Miyagawa K, Sakakura C, Nakashima S, Yoshikawa T, Kin S, Nakase Y, Ito K, Yamagishi H, Ida H, Yazumi S, Chiba T, Ito Y, Hagiwara A. Down-regulation of RUNX1, RUNX3 and CBFbeta in hepatocellular carcinomas in an early stage of hepatocarcinogenesis. Anticancer Res 2006;26:3633–43.
[59] Yang Nuo, Zhang Lin, Zhang Yue, Kazazian Haig H Jr. An important role for RUNX3 in human L1 transcription and retrotransposition. Nucl Acids Res 2003;31:4929–40.
[60] Giordano J, Ge Y, Gelfand Y, Abrusan G, Benson G, Warburton PE. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. PLoS Comp Biol 2007;3:e137.
[61] Rubin E, Lithwick G, Levy AA. Structure and evolution of the hAT transposon superfamily. Genetics 2001;158:949–57.
[62] 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;30:1239–1251.
[63] 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 USA 2005;102:396–401.

[64] Wang X, Li N, Liu B, Sun H, Chen T, Li H, Qiu J, Zhang L, Wan T, Cao X. A novel human phosphatidylethanolamine-binding protein resists tumor necrosis factor alpha-induced apoptosis by inhibiting mitogen-activated protein kinase pathway activation and phosphatidylethanolamine externalization. J Biol Chem 2004;279:45855–64.
[65] Li H, Wang X, Li N, Qiu J, Zhang Y, Cao X. hPEBP4 resists TRAIL-induced apoptosis of human prostate cancer cells by activating Akt and deactivating ERK1/2 pathways. J Biol Chem 2007;282:4943–50.
[66] Li W, Dong Y, Zhang B, Kang Y, Yang X, Wang H. PEBP4 silencing inhibits hypoxia-induced epithelial-to-mesenchymal transition in prostate cancer cells. Biomed Pharmacother 2016;81:1–6.
[67] Wang K, Jiang Y, Zheng W, Liu Z, Li H, Lou J, Gu M, Wang X. Silencing of human phosphatidylethanolamine-binding protein 4 enhances rituximab-induced death and chemosensitization in B-cell lymphoma. PLoS One 2013;8:e56829.
[68] Wang K, Jiang Y, Zheng W, Liu Z, Li H, Lou J, Gu M, Wang X. Decreased expression of protocadherin 20 is associated with poor prognosis in hepatocellular carcinoma. Oncotarget 2017;8:3018–28.
[69] Zhang C, Peng Y, Yang F, Qin R, Liu W, Zhang C. PCDH8 is frequently inactivated by promoter hypermethylation in liver cancer: diagnostic and clinical significance. J Cancer 2016;7:446–52.
[70] Zhu P, Lv J, Yang Z, Guo L, Zhang L, Li M, Han W, Chen X, Zhuang H, Lu F. Protocadherin 9 inhibits epithelial-mesenchymal transition and cell migration through activating GSK-3beta in hepatocellular carcinoma. Biochem Biophys Res Commun 2014;452:567–74.
[71] Dang Z, Shangguan J, Zhang C, Hu P, Ren Y, Lv Z, Xiang H, Wang X. Loss of protocadherin-17 (PCDH-17) promotes metastasis and invasion through hyperactivation of EGFR/MEK/ERK signaling pathway in hepatocellular carcinoma. Tumour Biol 2016;37:2527–35.
[72] Ye M, Li J, Gong J. PCDH10 gene inhibits cell proliferation and induces cell apoptosis by inhibiting the PI3K/Akt signaling pathway in hepatocellular carcinoma cells. Oncol Rep 2017;37:3167–74.
[73] Furuta M, Tanaka H, Shiraishi Y, Unida T, Imamura M, Fujimoto A, Fujita M, Sasaki-Oku A, Maejima K, Nakano K, Kawakami Y, Arihiro K, Aikata H, Ueno M, Hayami S, Arizumi SI, Yamamoto M, Gotoh K, Ohdan H, Yamaue H, Miyano S, Chayama K, Nakagawa H. Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. Oncotarget 2018;9:25075–88.
[74] Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell 2007;129:823–37.
[75] Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 2007;39:311–18.
[76] 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.
[77] Jenuwein T. The epigenetic magic of histone lysine methylation. FEBS J 2006;273:3121–35.
[78] Kim T.H., Abdullaev Z.K., Smith A.D., Ching K.A., Loukinov D.I., Green R.D., Zhang M.Q., Lobanenkov V.V., Ren B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 2007;128:1231–45.
[79] Laugesen A, Højfeldt JW, Helin K. Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. Mol Cell 2019;74:8–18.
[80] Dubois-Chevalier J, Staels B, Lefebvre P, Eeckhoute J. The ubiquitous transcription factor CTCF promotes lineage-specific epigenomic remodeling and establishment of transcriptional networks driving cell differentiation. Nucleus 2015;6:15–18.
[81] Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 1999;98:387–96.
[82] Mehra M, Chauhan R. Long noncoding RNAs as a key player in hepatocellular carcinoma. Biomark Cancer 2017;9:1–15.
[83] Maxwell PH. Consequences of ongoing retrotranposition in mammalian genomes. Adv Genom Genet 2014;4:129–42.