Review Article

Hepatitis B virus integration into chromosomes in hepatocellular carcinoma cells with a focus on centromeric alpha satellite DNA: Advances in molecular techniques reveal new insights into hepatocarcinogenesis

†Norio Ogata, MD, PhD[1, 2†]*

†Correspondence: *[1]Division of Laboratory Medicine, *[2] Division of Internal Medicine, Niigata Cancer Center Hospital, 2-15-3, Kawagishi-cho, Chuo-ward, Niigata city, Niigata 951-8566, Japan
Tel: +81-25-266-5111; Fax: +81-25-233-3849;
E-mail: nogata@niigata-cc.jp
Received August 15, 2025; accepted October 30, 2025
*[1] Division of Laboratory Medicine, *[2] Division of Internal Medicine, Niigata Cancer Center Hospital, 2-15-3, Kawagishi-cho, Chuo-ward, Niigata city, Niigata 951-8566, Japan

ABSTRACT

The molecular mechanisms by which hepatitis B virus (HBV) contributes to hepatocellular carcinoma (HCC) have been investigated since the late 1970s. This review follows advances in molecular biology and sequencing, particularly since the mid-2000s, which have enabled detailed analysis of HBV DNA integrations into the human genomes.

Molecular cloning and Sanger sequencing have identified that HBV integrates into several gene regions and non-genic regions, mainly repetitive sequences. Chromosomal rearrangements have also been observed. Our group first reported HBV integration into centromeric alpha satellite DNA (αSat) of HCC cells associated with genomic rearrangements. Second- generation (short-read) sequencing has found that many gene regions are involved in HBV integration with *TERT* promoter the most. Approximately 30 HCC-driver genes including *TERT*, *CTNNB1* and *TP53* have been identified. Third-generation (long-read) sequencing has revealed that driver genes are located not only close to HBV integration sites but also distant from HBV integration sites through structural variations (SVs). Also, 10-50% of HBV integrations have been confirmed to occur in centromeric or other repetitive regions, key loci for initiating SVs. Computational methods have emphasized that αSat may be the preferential locus for HBV integration. Recent completion of the full αSat sequence, the "real" goal of the Human Genome Project, allows better exploration of this underappreciated genome. Together, findings in this article refine current views of HBV-related hepatocarcinogenesis.

Ongoing advances in sequencing and centromeric biology are expected to deepen our understanding of virus-induced genomic instability and guide future strategies for early detection, prevention, and treatment of HBV-related HCC.

〔Lab Med Int 2026; 5(1): 3-19〕

Key Words

Hepatocellular carcinoma (HCC), Hepatitis B virus (HBV), α satellite DNA (α Sat), Centromere, Structural variation (SV)

I. Introduction

Epidemiological studies have clearly demonstrated a strong correlation between infection with hepatitis B virus (HBV) and the occurrence of hepatocellular carcinoma (HCC)[1)-3)].

Since the advent of molecular biological techniques applied to human genomes in the late 1970s, the causal relationship between HBV and HCC has been extensively studied. The discovery of HBV DNA integration into the chromosomes of human HCC tissues through Southern blot analysis suggested that integrated viral sequences might contribute to hepatocarcinogenesis.

Three key features have been proposed to define oncogenic viruses[4)]: (1) the presence and persistence of viral DNA or RNA in tumors, (2) the presence of viral oncogenes, and (3) the ability to modify host genes, through insertional mutagenesis (nearby genes) or chromosomal rearrangements (distant genes), which prompt activation of tumor-related genes or inactivation of tumor-suppressor genes.

This review summarizes breakthroughs in molecular biological techniques over the past few decades, that have advanced our understanding of HBV-related hepatocarcinogenesis. Particular emphasis is given to centromeric alpha satellite DNA (αSat) that we identified for the first time as a target of HBV DNA integration in HCC cells.

II. Southern blot analysis of HBV DNA in HCC

Beginning in the early 1980s, Southern blot analysis was widely used to examine the molecular state of HBV DNA in both HCC and non-malignant liver tissues[5)-8)].

Our representative results are shown in **Figure 1**. The HBV genome possesses a unique structure consisting of partially double-stranded DNA with a full genome length of 3.2 Kbp.

Typically, total DNA was extracted from HCC tissues and digested with restriction endonucleases, either those that do not cut within the HBV genome (e.g., *Hin*dIII or *Eco*RI) or those that do (e.g., *Bam*HI or *Bgl*II). The digested DNA was then subjected to agarose gel electrophoresis followed by Southern blot hybridization. Digestion with those restriction enzymes that do not cut HBV DNA allows the detection of integration forms of HBV DNA, which yield signals distinct from those of undigested DNA. Band signals indicate stable, fixed forms of HBV DNA present across clonal cell populations, while smeared signals suggest heterogenous structures of HBV DNA present among heterogenous cell populations. In some cases, replication forms of HBV DNA are also noticed in HCC tissues, which produce signals in both undigested- and enzyme-digested samples when the enzymes do not cut the viral genome, indicating active viral replication.

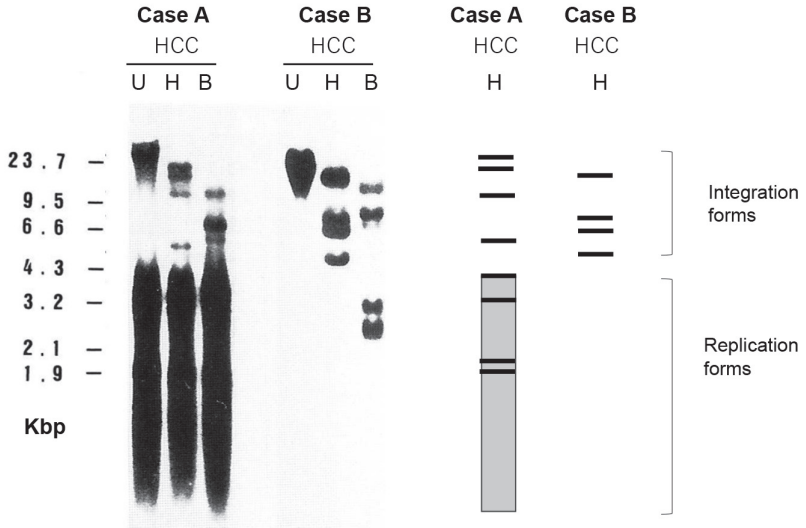

Figure 1 Representative findings of Southern blot hybridization of total DNA extracted from human HCC tissues for the presence and structures of HBV DNA. The extracted DNA (20 μg, each) was either undigested (U), digested with *Hin*dIII (H) or *Bam*HI (B), electrophoresed into 0.8% agarose gels and hybridized with a cloned HBV DNA probe. The *Hin*dIII is known to have no cutting site within the HBV genome, whereas BamHI possesses one cutting site within the HBV genome. Typical integration forms and replication forms of HBV DNA are illustrated at the right part of the figure.

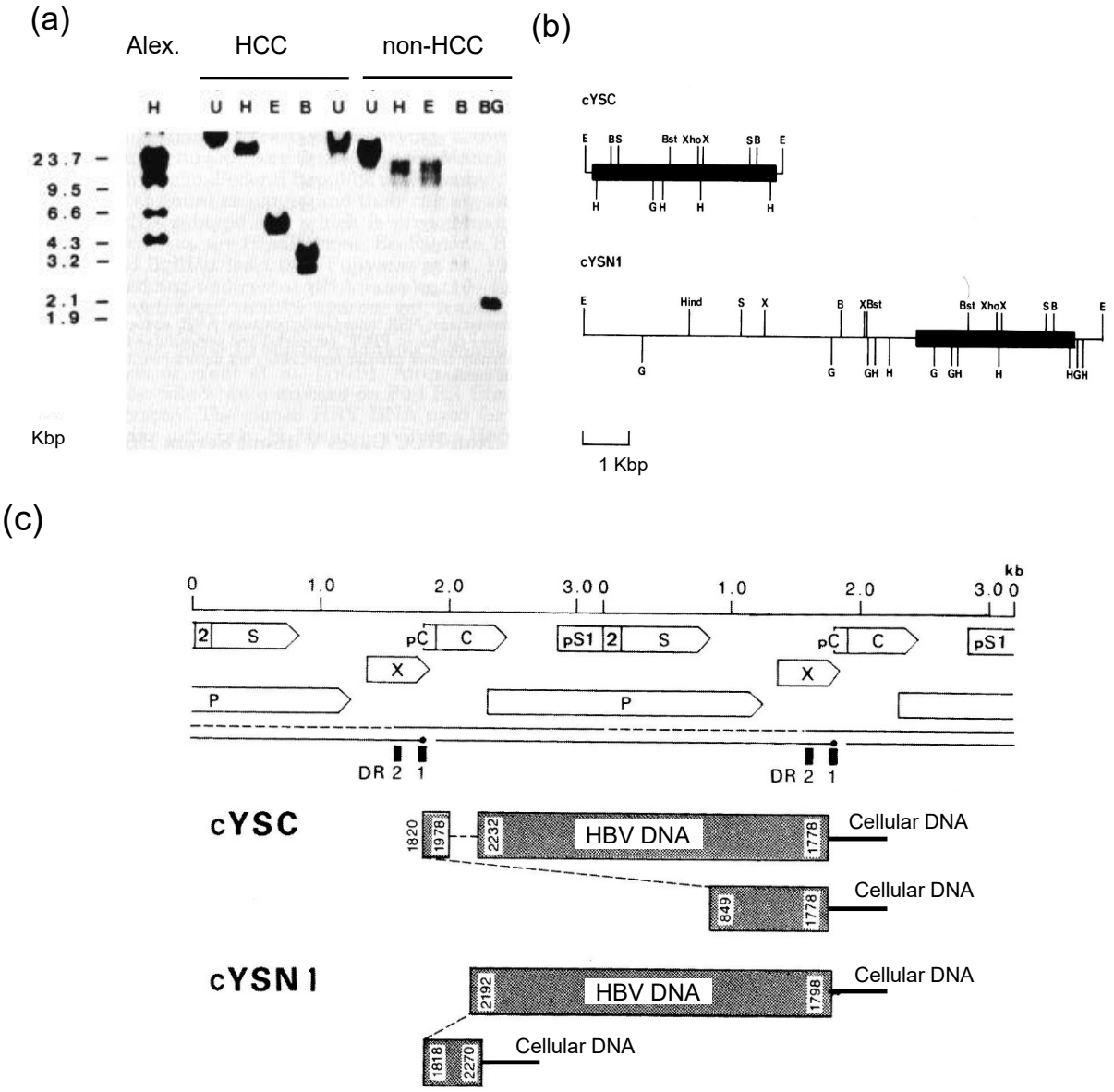

Figure 2

(a) Southern blot analysis of cellular DNA for the presence of HBV DNA. Alexander cell line (PLC/PRF/5 cells) was used as a positive control, of which lane is indicated as Alex. Total celluar DNA was extracted from Alexader cell lines. Also, total cellular DNA was extracted from HCC tissue and non-HCC tissue derived from the same patient. Extracted DNA (20 μg, each) was either undigested (U), digested with *Hin*dIII (H), *Eco*RI (E), *Bam*HI (B), or *Bgl*II (BG), electrophoresed into 0.8% agarose gels, blotted onto a nitrocellulose filter, and hybridized with a [α-^{32}P] cloned HBV DNA probe. *Hin*dIII and *Eco*RI have no cutting site within the HBV genome, *Bam*HI has one cutting site within the HBV genome, and *Bgl*II has more than two cutting sites within the HBV genome.

(b) Restriction endonuclease maps of cloned genomic DNA fragments derived from HCC tissue (cYSC) and from non-HCC tissue (cYSN1). Integrated HBV DNA is denoted by thick line; flanking cellular DNA by thin line. Cleavage sites with restriction endonucleases are indicated with the following abbreviations: E, *Eco*RI; B, *Bam*HI; Hind, *Hin*dIII; G, *Bgl*II; H, *Hinc*II; S, *Sph*I; Bst, *Bst*EII; Xho, *Xho*I, X, *Xba* I.

(c) Structure of the integrated HBV DNA sequences in the clones. In the upper part of the figure, tandem dimers of HBV genomes and coding regions with the direction of transcription of the structural genes are indicated with the following abbreviations; S = S gene coding for HBsAg; pS1 = preS1 region; 2 = preS2 region; C = C gene coding for HBcAg and HBeAg; pC = preC region; X = X gene; and P = P gene coding for a DNA polymerase / reverse transcriptase. The positions of the 11-bp direct repeats (DR2 and DR1) are indicated by small vertical bars that start at nucleotide positions 1590 and 1824, respectively. The region between DR2 and DR1 is the cohesive end region. In the lower part of the figure, organization of integrated HBV DNA clone from HCC tissue (cYSC) and non-HCC tissue (cYSN1) are illustrated by shaded bars with the nucleotide positions at each end. Deletion in the genome in cYSC is represented by a dashed line. The complex-type virus DNA that consists of two HBV genomes whose regions are overlapping, at least in part (cYSC and cYSN1), is represented by multiple solid bars aligned in different levels, each representing different units of the virus genome, and are connected by tilted dashed lines. In the clone cYSC, two incomplete HBV genomes are connected in an inverted manner at nucleotide positions 1820 and 849. Both virus-cellular junctions locate at the same nucleotide position 1778, between DR2 and DR1. In the clone cYSN1, two incomplete HBV genomes are also connected in an inverted manner at nucleotide positions 2192 and 1818. One virus-cellular junction locates at nucleotide position 1798, between DR2 and DR1, and another junction locates at nucleotide position 2270.

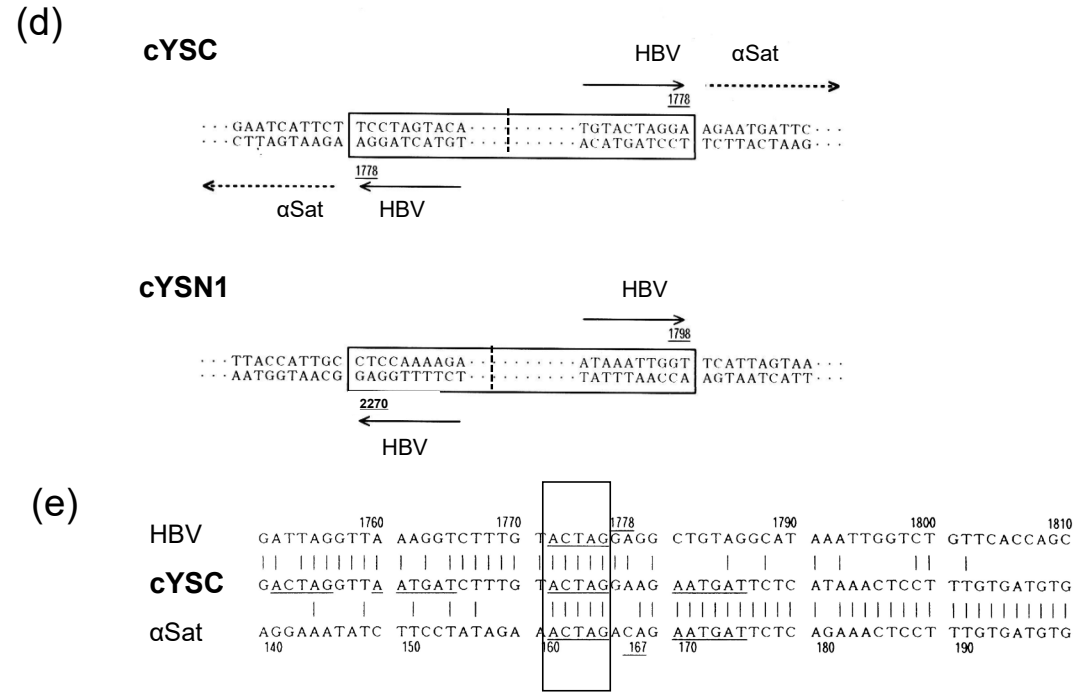

Figure 2 (continue)
(d) Nucleotide sequences around the junction region of HBV-host DNA in the integrated HBV DNA clone from the HCC tissue clone (cYSC) and non-HCC tissue clone (cYSN1). The HBV DNA sequences are boxed. The connection of the two incomplete HBV genomes is indicated by dashed line. In cYSC, HBV DNA and flanking host alpha satellite (αSat) sequences together formed an inverted structure. In cYSN1, the HBV DNA is inverted, but flanking cellular sequences are not inverted. Numbers underlined indicate nucleotide positions at the end of integrated HBV sequences. Arrowheads of solid lines indicate the direction of HBV transcription and those of dashed lines represent the 5' to 3' direction of monomeric units of αSat.
(e) Nucleotides sequences showing 5-bp microhomology (MH), ACTAG, between HBV DNA and αSat around the virus-cellular junction locus. The MH is presented by a box.

III. Sanger sequencing-based detection of HBV integration events

1. Case study: HBV integration into centromeric αSat in a young female patient with HCC

A 27-year-old Japanese woman, known to be a chronic HBV carrier with positive serum antibodies to HBe antigen (anti-HBe) and normal serum ALT values, developed HCC [7]. Total DNA was extracted from surgically resected HCC tissue and adjacent non-tumor tissue and subjected to Southern blot analysis. Histo-pathologically, the non-tumor tissue showed minimum or no necro-inflammatory reactions.

As shown in **Figure 2a**, following digestion with *Hin*dIII or *Eco*RI, enzymes that do not cut within the HBV genome, DNA from the HCC tissue showed a single band, indicating a clonal population harboring a uniform HBV DNA structure, whereas DNA from the non-tumor tissue exhibited a smeared pattern, suggesting heterogeneity in HBV DNA structure among non-cancerous hepatocytes. A DNA clone from HCC tissue (designated cYSC) and a DNA clone from non-tumor tissue (designated cYSN1) were molecularly cloned, subjected to generating restriction endonuclease mappings (**Figure 2b**) and sequenced using the Sanger method. Both clones showed similar integration structures, characterized by inverted HBV sequences, with host-viral DNA breakpoints located near the direct repeat 1 (DR1) and direct repeat 2 (DR2) regions, consistent with previous reports (**Figure 2c**). The flanking cellular DNA of cYSC was identical to αSat and surprisingly was also inverted along with the HBV genome (**Figure 2d**). Further analysis identified microhomology (MH) between HBV and αSat (**Figure 2e**), suggesting MH-mediated HBV integration.

A schematic diagram of the constituents of human centromere, including αSat, and changes of tandem repeats of αSat monomer caused by HBV integration in the presented case is shown in **Figure 3**.

Our findings are the first demonstrating centromeric αSat as a target for HBV DNA integration associated with inverted fusion of both viral and host repeat sequences [7].

2. Summary of direct cloning and Sanger sequencing studies

From the 1980s through the 1990s, a number of stud-

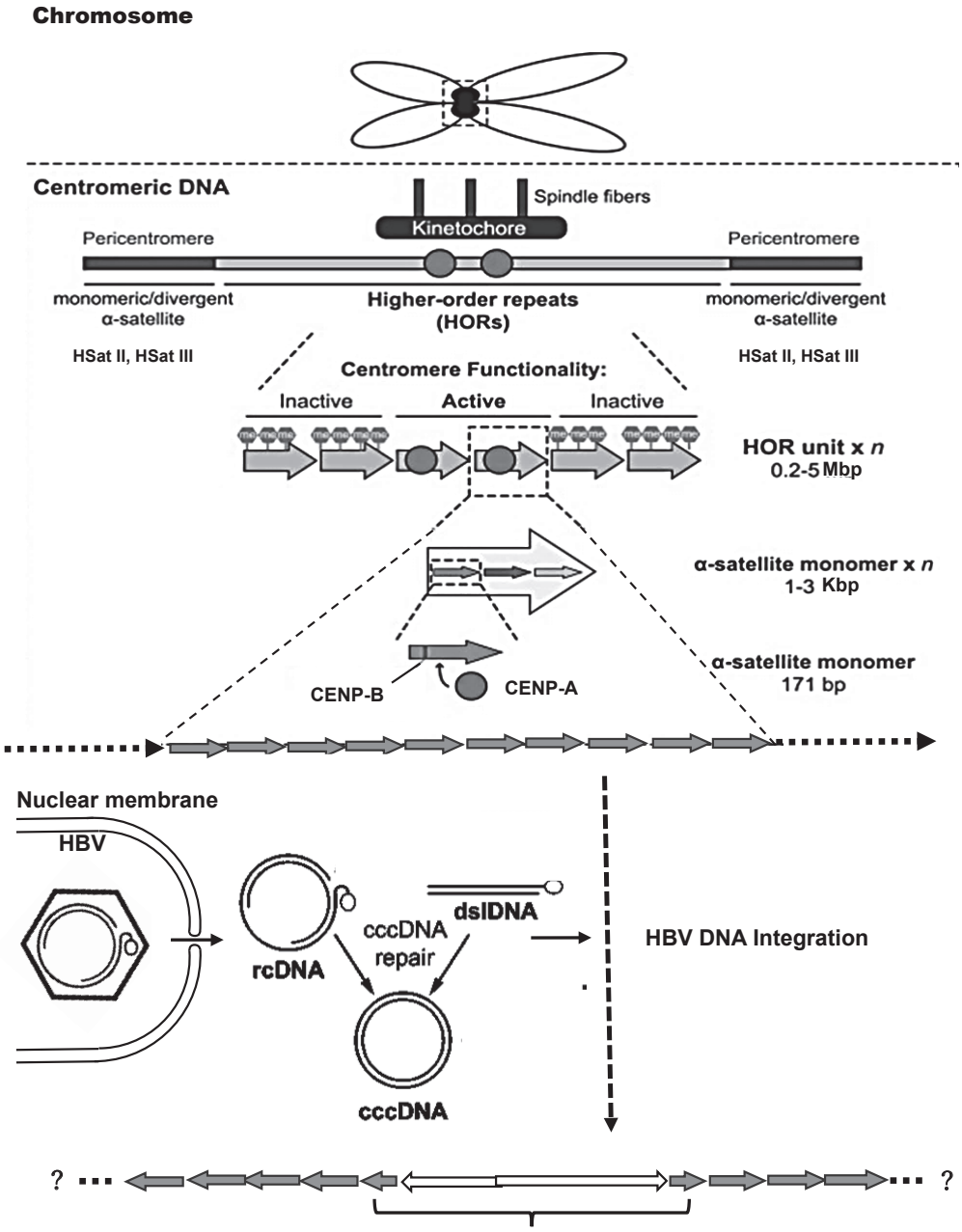

Figure 3 Human centromere structure and kinetochore organization. Human centromeres contain arrays of higher order repeats (HORs) of ~171-bp αSat monomers and arranged as a head-to-tail manner that can span several Mbp of DNA. Kinetochore proteins associate with only a subset of HOR units, which is called the active array. The functionally active centromere is defined epigenetically with centromeric protein (CENP)-A-containing nucleosomes. CENP-A plays central functions of the centromere in chromosome stability and segregation in mitosis. Some of αSat harbor a centromeric protein (CENP)-B binding motif (17-bp motif). CENP-B stabilizes the interaction between the chromatin and kinetochore. Pericentromeric regions contain monomeric or divergent αSat together with human satellite II (HSat II) / satellite III (HSat III) sequences and transposable elements.

Virions of HBV enter hepatocytes, then in the nucleus their genomes, either relaxed circular (rc) form or double-stranded linear (dsl) form, convert into covalently closed circular (ccc) form. All three HBV DNA molecules, rcDNA, dslDNA, and even cccDNA, can be molecular forms for integration into host chromosomes. In the HCC-derived clone in the patient presented in this article, after HBV DNA integration, structures of HBV DNA (possibly two genomes were integrated) underwent rearrangements together with franking αSat, resulting in an inverted fusion structure. Because the sequenced αSat region covered only a short part of the monomer, the extent of inverted repeat structure of αSat was not known.

Table 1 Summary of the results of altered chromosomal DNA regions in HCC cells, into which HBV integrated, that were analyzed using molecular cloning and Sanger sequencing.

gene region	clone No	Ref.	non-genic region	clone No	Ref.	chromosome	clone No	Ref.
erb-A/TR (*RARβ*)	1	10	minisatellite DNA	1	16	inversion	4	7, 25, 26
			satellite III DNA	4	17, 18	amplification	4	26, 27
erb-B (*EGFR*)	1	11	α satellite DNA	3	7, 19, 20	deletion / allelic loss	1	28
			long interspersed	2	13, 21	translocation		
cyclin A (*CCNA2*)	1	12	nuclear element			chr 17;18	1	29
			(LINE)			chr X;17	1	30
RBMY	1	13	short interspersed	8	18, 22, 23, 24	chr 5;9	1	30
p53 (*TP53*)	6	14, 15	nuclear element			chr 7p;17cen	1	19
			(SINE)			chr 3q;8p	1	31
						chr 17p;8p	1	32

No, number.
Current gene names are indicated in parenthesis.

ies used direct molecular cloning followed by Sanger sequencing to characterize integrated HBV DNA and its flanking host sequences. Much of this fundamental work was carried out by Prof. Kenichi Matsubara and colleagues at the Institute for Molecular and Cellular Biology, Osaka University [8) 9)].

A summary of these findings, including our owns, is provided in **Table 1**.

HBV integration sites were identified within or near the five gene regions, including four tumor-related genes involved in cell proliferation or cell cycle regulation and one tumor-suppressor gene, *TP53* [10-15)]. The infrequent detection of gene regions might be attributed to the fact that many genes had not yet been identified in those days.

HBV integration sites were found in non-genic regions more commonly, especially repetitive elements such as satellite DNAs, including αSat detected for the first time in our study, long interspersed nuclear elements (LINEs) of the *Kpn* family and short interspersed nuclear elements (SINEs) of the *Alu* family [7) 16-24)].

Chromosomal rearrangements were also observed in association with HBV integration [7) 19) 25-32)]. Among chromosome translocations, chromosome 17 carrying *TP53* tended to be frequently observed. Interestingly, in one case, the centromere of chromosome 17 served as the origin of a translocation to chromosome 7p [19)].

Thus, HBV integration can contribute both local and distant genomic alterations.

IV. Insights from second-generation (short-read) sequencing technologies (Illumina reads)

1. Impact of The Human Genome Project

In 1989, the White House of the United States announced the Human Genome Project (HGP), which began in 1990 and was completed in 2003. It successfully determined approximately 99% of the 2.85 Gbp nucleotide sequence in the human genome. The HGP used a clone-by-clone shotgun Sanger sequencing method. However, this method was unable to sequence centromeres, telomeres, and ribosome DNAs.

2. Advances of second-generation sequencing technology

Since the mid-2000s, next-generation sequencing, now called second-generation sequencing, has become available. This technology is approximately 300 times faster in analyzing sequences than the Sanger method. However, its limitations include short read lengths and difficulty in accurately aligning repetitive sequences.

In 2006, the National Institutes of Health in the USA launched The Cancer Genome Atlas (TCGA). This is the project to catalogue the genome alterations responsible for cancer. In 2008, for collaborating with world's leading researches of cancer and genomics the International Cancer Genome Consortium (ICGC) was established. Subsequent phases of ICGC included the Pan Cancer Analysis of Whole Genomes (PCAWG) in 2014 and Accelerating Research in Genomic Oncology (ARGO) in 2019.

Based on cited articles, in this review, HBV DNA integrations into gene regions are defined as those within exons, introns, or sequences within 10 Kbp away from the gene. HBV DNA integration is considered colocalized with host DNA structural variants (SVs), including copy number alterations (CNAs), if located within 500 Kbp. Driver genes are defined as genes with mutation frequencies that are statistically significant after accounting for background mutation rates.

Table 2 Summary of the results of altered chromosomal DNA regions in HCC cells, into which HBV integrated, that were analyzed using short-read (Illumina reads) sequencing .

Illumina reads						
	gene region	sample No	non-genic region	sample No	chromosome (gene)	Ref. 33
Patient No n=68	*TERT*	17	near the	ND	recurrent arm level gains in chr 1q, 5p, 6p, 8q	
	KMT2B	6	centromere		recurrent arm level losses in chr 1p, 4q, 6q, 8p, 17p	
	DQ582201	4	on chr 1p, 8p, 10q			
HBV-BP No n=628	*DDX11L11*	3			focal amplification: 25%:	
	DQ590432	2			30 genes (*TERT*, C*CND1*, *EGF19*, etc.)	
	ALOX5	2			homozygous deletion: 17%	
	MYO19	2			28 genes (*CDKN2A/2B*, *MAP2K3*, *PTEN*, etc.)	
	SENP5	2				
	RGS22	2			* relation to HBV integration : ND	
	ZFPM2	2				

Illumina reads						
	gene region	sample No	non-genic region	sample No	chromosome (gene)	Ref. 34
Patient No n=74	*TERT*	14	lnc RNA	ND	tandem duplication	
	KMT2B	5	*NEAT1*		deletion	
	SOX 5	2	*MALAT1/NEAT2*		translocation	
HBV-BP No n=233			CTCF-binding sites		38 genes (*TERT*, *CCND1*, *CDKN2A*, *ASH1L,* etc.)	
			* relation to HBV integratin: ND		* relation to HBV integration: ND	

High throughput viral integration detection (HIVID) method						
	gene region	sample No	non-genic region	sample No	chromosome (gene)	Ref. 38
Patient No n=426	*TERT*	101	preferentially	ND	HBV integration frequently at chr 5, 16, 17, 19	
	KMT2B	31	"repetitive"		especially at chr 17p (*TP53*, *MAP2K4/MMK4*)	
	FN1	17	or			
HBV-BP No n=3486	*CCNA2*	8	rare fragile sites			
	CCNE1	7				
	FAM 157A	7	satellite DNAs			
	PTPRD	2 ≤	tandem repetitive-			
	UNC5D	2 ≤	sequences			
	NRG3	2 ≤				
	CTNND2	2 ≤				
	AHDD	2 ≤				

No, number; BP, breakpoint; ND, not described.
Only recurrently altered gene regions are indicated.
Genes altered by chromosome rearrangements are indicated in parenthesis.

3. ICGC and HBV integration

As shown in **Table 2**, Totoki et al. [33] and Fujimoto et al. [34] contributed HCC sequencing data to the ICGC project. Totoki et al. [33] analyzed a multi-ethnic cohort (Japanese, Asian, and European), while Fujimoto et al. [34] focused exclusively on Japanese patients. Both studies included HCC cases of various etiologies, such as HBV, other viruses, and non-viral.

Regarding gene regions, Totoki et al. [33] identified 308 gene regions and Fujimoto et al. [34] found 94 gene regions, that were targeted by HBV integration, indicating the heterogeneity of HCC genomes. Both studies reported frequent HBV integrations in *TERT* and *KMT2B*. Totoki et al. [33] further analyzed *TERT* promoter mutations

and detected them in 54% (254 /469) of HCC cases, including 37% (44 /120) HBV-positive HCCs. The HBV integration and *TERT* promoter mutations were mutually exclusive. Collectively, *TERT* somatic alterations and enhanced expression were observed in ~70% of HCC cases.

Regarding non-genic regions, Totoki et al.[33] noted some HBV integrations near the centromeres on chromosomes 1p, 8p, and 10q, although frequency and detailed characterization were not reported. Fujimoto et al.[34] did not show HBV integration into repetitive sequences, but identified recurrent alterations in long non-coding RNAs (lncRNAs), such as *NEAT1* and *MALAT1* (*NEAT2*). However, their association with HBV integration was not described.

Chromosomal rearrangements (i.e., SVs including CNAs) occurred frequently in HCC. A broad range of tumor-related genes and tumor-suppressor genes were affected by these molecular events. Totoki et al.[33] reported several chromosomal arm-level gains and losses, affecting 30 genes with focal amplifications and 28 genes with homozygous deletions. Although these SVs were not mentioned in the context of HBV integration, the number of such SVs was found to be higher in HBV-related HCCs than in HCCs of broad range of other etiologies. Fujimoto et al.[34] reported several chromosomal rearrangements, such as tandem duplications, deletions, and translocations, affecting 38 genes with alterations. Although each SV was not described in relation to HBV integration, the number of HBV integration sites per sample was identified to be positively correlated with the number of SVs.

Based on significant alterations, 30 genes by Totoki et al.[33] and 25 genes by Fujimoto et al.[34], were selected as candidate driver-genes. Importantly, mutations in drug target kinase genes such as *EGFR1*, *EGFR2*, *FGFR3*, *KIT*, *JAK1,* and *MET* were rarely found in HCC[33, 34]. This observation suggests limited applicability of current cancer genomic medicine in patients with HCC.

To further explore the genomic landscape and its variability across populations, Shibata et al.[35] expanded upon these findings by incorporating two additional studies[36, 37]. They compiled genomic data of TCGA project from more than 1,000 publicly available HCC cases, which included 1,340 patients from different ethnic groups, including Japanese, Asian, and European cohorts.

In total, 33 driver genes were detected. The most frequently mutated gene region was *TERT* (~ 60%), followed by *CTNNB1* and *TP53* (each ~ 30%). Due to the high GC content of the *TERT* promoter regions, short-read Illumina sequencing had difficulty in detecting mutations, and thus the true frequencies of *TERT* region mutations could not be established. *TERT*, *CTNNB1,* and *TP53* were identified as master genes in HCC across all ethnic groups. The second most frequently (> 7%) mutated driver gene group included *ALB*, *APOB*, *ARID1A*, *ARID2,* and *AXIN1*. Remaining low frequency (< 5%) mutated driver gene group was involved in various biological functions.

Collectively, across the four referenced studies, SVs, including CNAs, occurred frequently in HCC cells and affected a broad range of tumor-related genes and tumor suppressor genes. Also, SVs significantly increased at HBV breakpoint locations, indicating that HBV integration contributed to gene alterations through chromosomal instability in HCC cells.

4. High-throughput viral integration detection

Zhao et al.[38] reported the sequencing results of patients with HBV-related HCC using the high throughput viral integration detection (HIVID) method from a Chinese cohort of 600 patients with HBV-related HCC. Briefly, HIVID uses HBV-specific probes to enrich viral DNA, followed by computational analysis based on TCGA dataset. (see Refs. 38, 39 for methodological details).

As shown in **Table 2**, the study by Zhao et al.[38] examined a relatively large number of patients with HBV-related HCC and studied HCC tissues and paired non-tumor tissues.

In gene regions, they identified 826 genes with HBV insertions in tumor samples and 303 in non-tumor samples, with only 64 genes shared between the two, highlighting distinct integration profiles in tumor versus non-tumor tissues. Notably, there were as many as 88 genes recurrently affected in HCC, many of which were previously unreported. Examples included *PTPRD*, *UNC5D*, *NRG3*, *CTNND2*, and *AHRR*. Expression changes were noticed both at the transcript and protein levels, suggesting that these novel affected genes were functionally relevant to HBV insertion.

Interestingly and importantly, HCCs arising from non-cirrhotic liver displayed a significantly enriched HBV integration in the vicinity of *KMT2B*, *CCNE1*, and *AHDD*.

In non-genic regions, HBV integration occurred preferentially in satellite DNAs and tandem repetitive sequences, possibly including LINEs and SINEs. This is the first study to report that HBV preferentially integrates into repetitive or fragile genomic regions, including satellite DNAs.

Regarding chromosomal rearrangements, HBV integration was significantly enriched in chromosomes 5,16,17,

and 19, especially chromosome 17p, which harbors *TP53* and a newly identified *MAP2K4* (*MKK4*).

Clinically, circulating HBe antigen (HBeAg) levels were positively correlated with the number of HBV integrations in HCC tissues, suggesting that higher HBV replication activity increases the likelihood of integration.

V. Computational tools for detecting HBV integration in repetitive regions: Insertion or integration callers

As described previously, short-read sequencing technologies face inherent limitations, particularly in mapping reads to repetitive genomic regions. When multiple similar copies of a region exist throughout the genome, aligners often fail to determine the correct location, as multiple potential alignment sites can exist for a single read. This challenge becomes especially critical when a virus integrates into repetitive sequences, leading to misaligned or ambiguously mapped reads.

Several computational programs have been developed to analyze large-scale sequencing datasets for viral integration events, but they struggled to accurately predict integrations in repeat-rich regions of the host genome. Among available computational methods, SurVirus has been identified as a particularly sensitive, precise, and efficient virus integration caller to achieve precise results. Using a second-generation paired-end sequencing dataset, along with a host genome reference and a viral database, SurVirus predicts integration events that occurred, providing the precise integration loci on the host genome and the corresponding viral segments involved (see Refs. 40, 42, and 43 for methodological details).

Rajaby et al.[40] applied SurVirus to analyze HBV using the HBV HIVID database[38] and the HBV whole-genome sequencing (WGS) dataset[41]. Among approximately 500 detected HBV insertion events, 7% were novel, and most occurred in repeat regions. Among these HBV insertions, satellite DNAs accounted for 40% of the insertions, followed by LINEs (15%) and SINEs (0.5%).

Because Rajaby and colleagues[40),42),43)] extensively published on similar computational algorisms for detecting insertion elements in host genomes[42),43)], their findings are considered robust and reliable. Importantly, this study is the first to highlight the significant role of repetitive sequences in hepatocarcinogenesis, indicating the need for computational tools capable of resolving complex genome regions.

VI. Findings from third-generation (long-read) sequencing technologies (PacBio or Nanopore reads)

1. Overview of third-generation sequencing

While second generation sequencing technology enabled analyses of HBV integration in HCC, their reliance on short reads introduces several limitations. These include the inability to directly sequence full-length viral genomes, sensitivity to probe-capture bias, sequencing depth, and analysis methods. Importantly, short-read technologies struggle with tandem repeats and repetitive sequences, often leading to errors in sequencing, alignment, assembly, or amplification. This makes it challenging to determine the exact sequence composition and number of repeat units within each repetitive region.

In the mid-2010s, third generation sequencing methods, including PacBio and Nanopore reads, have become available, providing long reads (~20 Kbp). These long-reads enabled more accurate mapping of integration events, especially in complex genomic regions.

2. Key studies using short- and long- read sequencing

As shown in **Table 3**, recent studies, ranging both short-read (Illumina reads) and long-read (PacBio or Nanopore reads) sequencing, have provided crucial insights into HBV integration patterns.

Ethnicities of patients with HBV-related HCC in each report were as follows; Zhuo et al.[44] studied a Chinese cohort, Alverti et al.[45] analyzed a Japanese cohort, Peneau et al.[46] examined multi-ethnic (European, African, and Asian) cohort, and Qian et al.[47] studied a Chinese cohort.

Regarding gene regions, altered gene regions were essentially similar to previous reports, and further, confirmed that HBV enhancers induced high mRNA expression. Important findings from these studies can be summarized as follows; First, HBV integration in the promoter region of *TERT* was not detected in non-tumor tissues[44]. Second, tumors harboring HBV integrations in *CCNE1*, *CCNA2*, or *KMT2B* were more frequently identified in patients without cirrhosis[46]. Third, *KMT2B*-integrated HCCs did not show alterations in the *TERT* promoter or other known driver genes[46], suggesting that *KMT2B*-altered HCC may involve distinct mechanisms of carcinogenesis. Fourth, among young patients with HCC (< 35 years old), *TERT* mutations (1.9%) and *CCNE1* mutations (3.7%) were rare, while over half of these patients had alterations in *TP53*[47]. Thus, in HCCs from a subgroup of patients without cirrhosis, alterations of one or two genes may induce strong proliferation, replication stress, and a rearrangement signature that

Table 3 Summary of the results of altered chromosomal DNA regions in HCC cells, into which HBV integrated, that were analyzed using short-read (Illumina reads) and long-read (PacBio or Nanopore reads) sequencing.

Illumina reads						**Nanopore or PacBio reads**		Ref.
	gene region	sample No	non-genic region	sample No	chromosome		chromosome (gene)	44
Patient No ND	*CCNE1*	153	"repeat sequences"	265	ND	Patient No ND	"complex" all autosomal chromosomes except for chr 15, 16, 21 priority on chr 1, 4, 5, 19	
	TERT	17						
	SYT14	27						
HBV-BP No n=475	*SERTADO*	7				HBV-BP No n=141		
	LINCO 1493	2						
	MIR 12125	2						
	MIR 4457	2						

Illumina reads						**Nanopore reads**		Ref.
	gene region	sample No	non-genic region	sample No	chromosome (gene)		chromosome (gene)	45
Patient No n=51	*TERT*	14	ND	ND	chr 5p amplification (*TERT*)	Patient No n=7/9	translocation chr 1p;9p with chr 1p deletion (*ARID1A*) chr 1p;11q chr 3p;Xq with chr Xp deletion (*RPS6KA3*) chr 4q;7q with (one patient: chr 4q deletion (*IRF2*); chr 8q;4p; chr 8q;10q) chr 8q;13q wirh chr 13q deletion (*RB1*) chr 8p;17p with chr 17p deletion (*TP53*) chr 11q;19p	
	KMT2B	4						
	CCNE1	2						
HBV-BP No n=148								

Illumina reads						**PacBio reads**		Ref.
	gene region	sample No	non-genic region	sample No	chromosome (gene)		chromosome (gene)	46
Patient No n=177	*TERT*	48	centromere/ peri-centromere	10-20 %	chr 17p deletion (*TP53*) :15 cases 8/15 centromere**	Patient No n=3	inverted duplication : 1 sample translocation : 2 samples isochromosome ** : 1 sample	
	CCNE1	4						
	KMT2B	3						
HBV-BP No n=2199	*AHRR*	2			chr 5p amplification (*TERT*) :14 cases	HBV-BP No n=8	amplification: chr 5p (*TERT*), 8q (*MYC*) deletion: chr 17p (*TP53*)	
	NRG1	2						
	TRIM16L	2			chr 8q amplification (*MYC*) :13 cases 4/13 centromere**			
	ST18	2						

Illumina reads						**Nanopore reads**		Ref.
	gene region	sample No	non-genic region	sample No	chromosome (gene)		chromosome (gene)	47
Patient No n=84	*TERT*	28	teromere or centormeric satellite DNA	15%	chr 17p deletion near centromere** (*TP53*)	Patient No n=50	large duplication : 10 samples inverted fusion: 54 samles translocation : 59 samples	
	KMB2	7						
	CCNE1	2						
HBV-BP No n=482	*TSHZ2*	2			chr 8q amplification and / or chr 8p deletion	HBV-BP No n=298	amplification: chr 5p (*TERT*), 8q (*MYC*), 19q, 1q, 7p, 6p, 20q deletion: chr 17p (*TP53*), 8p, 4q, 16q, 6q, 9p (*CDKN2A*), 13q (*RB*), 1p (*ARID1A*)	

No, number; BP, breakpoint; ND, not described; ** HBV integration into centromere
Only recurrently altered gene regions are indicated.
Genes altered by chromosome rearrangements are indicated in parenthesis.

directly promotes cancer development. Further, in HCCs in a subgroup of younger patients, different mechanisms other than common gene alterations may develop cancer. Collectively, these findings suggest that distinct mechanisms may underlie HBV-related HCC development across ethnic groups, non-cirrhotic liver states and age cohorts.

Regarding non-genic regions, Zhuo et al.[44] reported that more than half, 55.8% (265 /475) of HBV breakpoints in tumors were located within repetitive sequences. The frequency of such breakpoints was higher in tumor tissues than in non-tumor tissues. Although the specific characteristics of these repetitive sequences were not clearly identified, these findings suggest that preferential HBV integration into chromosomal repetitive sequences may confer a selective advantage during tumorigenesis. Additionally, Peneau et al.[46] found that approximately 10-20% of HBV breakpoints were located in centromeric or pericentromeric regions. Also, Qian et al.[47] reported that approximately 15% of HBV breakpoints were in telomere or centromeric satellite DNAs. These results indicate that HBV integration into centromeric regions is more frequent than previously reported.

Several studies have reported a relationship between viral insertion events and SVs. For example, Peneau et al.[46] reported that 36% of HBV integration events precisely matched SV boundaries. Three major types of SVs that were bordered by HBV integration in more than 40 samples included deletion of chromosome 17p and amplification of chromosomes 5p and 8q. Notably some of these SVs involved centromeric regions. Qian et al.[47] reported that deletion of chromosome 17p occurred near an HBV-integrated centromere.

In these reports, despite the use of short-read sequencing, a relatively high percentage of HBV integrations into repetitive sequences was observed compared to previous studies. While the reason for this discrepancy remains unclear, possibly improved sequencing performance may partly explain the difference.

Using long-read sequencing, detailed insights into the contributions of HBV integration to SVs have been obtained. Peneau et al.[46] reported that HBV integration-related SVs were prevalent in 41% of HCCs and 48% of them were associated with CNAs. Qian et al.[47] reported that HBV integration-related CNAs occured in 58% of HCCs, especially enriched in younger patients. Both studies identified common CNAs at the chromosome arm level as well as focal copy number changes affecting key cancer-related genes.

In case reports, Alverti et al.[45] reported detailed results. They studied nine Japanese patients with HCC, in whom only one end of the integrated HBV DNA was detected using short-read sequencing. Among these, seven patients showed non-canonical integration involving chromosomal translocations.

Peneau et al.[46] examined three patients with non-canonical integration. They found several types of SVs, including an intriguing finding of generating isochromosome of 8q, which might be caused by HBV integration into centromere.

Qian et al.[47] studied many patients with non-canonical integration. They identified several types of SVs, including amplifications and deletions in a large number of chromosomes, which were accompanied by HBV integration.

Peneau et al.[46] and Qian et al.[47] investigated HBV-integration events in relation to clinical aspects.

Peneau et al.[46] identified a larger amount of replicating HBV DNA in non-tumor tissues correlated with a higher number of HBV integration in HCC cells. They further found the number of HBV integration generating CNAs-was an independent prognostic factor in HCC progression.

Qian et al.[47] compared CNAs between HBsAg-positive and HBsAg-negative groups. HBsAg-positive tumors showed significantly more frequent ($p<0.02$) amplification of chromosome 8q and deletions in chromosomes 4q, 16q, and 17p, which were among the most common HBV-associated CNAs. Although not all chromosomes were studied, these results suggest a contribution of HBV integration-associated CNAs to HCC development. They also reconstructed the likely evolutionary scenario timelines of HBV-integration-induced amplification in chromosome 8q, using changes in the number of the clock-like mutations using COSMIC SBS signature 1 and 5 before and after the amplification. Their results suggested that most of these amplifications likely occurred decades prior to the clinical diagnoses, some as early as before age 10. They proposed that HBV-integration-induced CNAs, especially by ultra-early chromosome 8q amplifications, may promote the initial clonal expansion, shorten the time to malignant transformation, or accelerate tumor progression through the combined effects of viral proteins and overexpression of oncogenic hotspot genes. They also observed a significant correlation between higher HBV DNA levels and *TP53* mutation. Based on these findings, they emphasized importance of early antiviral therapy to prevent HBV-driven genomic alterations.

Collectively, long-read sequencing has revealed that not only gene-regions close to HBV integration sites but also those located at considerable distances from HBV inte-

gration sites can be altered through HBV-related SVs. Importantly, centromeric regions appear to contribute to the generation of multiple SVs, as previously reported by our group [7] and others [19] during the early era of molecular biology using direct cloning and Sanger sequencing.

While many HBV integrations may represent passenger events, some act as cancer drivers that promote HCC initiation. Long-read sequencing is emerging as an essential tool in HBV integration pattern research and may have future research and possibly clinical applications. It is particularly valuable for elucidating HBV integration patterns including SVs and clonal expansion, and finally, the underlying mechanisms of HBV-induced carcinogenesis.

VII. Characterization of genomics and proteomics of integrated HBV

1. Structural features of integrated HBV

It is widely believed that most integrated HBV sequences are derived from the normal double-stranded linear (dsl) DNA and primarily through MH-mediated mechanisms [48]. Li et al. [49] by analyzing long-read sequencing data from 35 HBV-related patients with HCC, identified several structural patterns of HBV integration. In their study, both ends of the integrated HBV fragments were captured using long-read sequencing. Notably, short HBV insertion segments (0–1 Kbp) accounted for ~50% of integration events, while longer viral segments (>3 Kbp) accounted for ~25% of such events. The longest detected HBV insertion measured 5,408 bp.

Normally generated HBV dslDNA is speculated to be in the range of 1816–3182/1–1832. The dslDNA is thought to be difficult to cross the 1600–1900 bp region. However, Li et al. [49] found the HBV insertion fragment in the range from 843 to 2263 of HBV genome. Notably, some integration events included complete pre-surface (S) 1/pre S 2/S gene or pre-core (C)/C gene.

Therefore, they speculated several mechanisms of virus integration as follows. (1) The reverse transcription process leads to the formation of excessively redundant HBV dslDNA sequences, which are integrated into the human genome through the classical non-homologous end joining or MH-mediated end joining (MMEJ) mechanism. (2) The abnormally redundant HBV single-stranded DNA is integrated into the human genome through the single-strand annealing mechanism. (3) The closed circular HBV genome is integrated into the human genome through the MMEJ mechanism.

Due to limited research on the characteristics and mechanisms of HBV integration, the integration mechanism of HBV sequence remains unclear.

2. Mutations and genotypes of integrated HBV

Peneau et al. [46] unexpectedly observed a negative selection of HBV variants that impair HBeAg production, including mutations in the basic core promoter locus and pre-core region [50) 51] or mutations associated with antiviral resistance in reverse transcriptase domains [52] in HBV sequences integrated into HCC genomes. This observation suggests that while these mutations may enhance viral fitness or facilitate immune escape, they do not provide a specific advantage for tumor development.

Qian et al. [46] investigated the genotypes of integrated HBV in a diverse population where nearly all major HBV genotypes (A, B, C, D, E, F, G, and H) were present. Among the integrated HBV sequences, genotype C was the most prevalent, followed by genotype B.

3. Transcripts and proteins of integrated HBV

As part of the Clinical Proteomic Tumor Analysis Consortium, a comprehensive proteogenomic analysis of HBV-related HCCs in a Chinese cohort was performed by Gao et al. [53].

Gao et al. [53] reported that the large envelope protein, external core antigen/capsid protein, and polymerase were detected in both proteomic and RNA sequencing datasets, while the X protein was not detected, but only mRNA was detected. These results raise questions about the role of the X protein in HCC, suggesting that it may not be essential for the maintenance of the malignant phenotype. There were less abundant HBV mRNAs and proteins in HCC tissues than in non-HCC liver tissues. The presence of HBV proteins in HCC tissues did not correlate with the survival of patients with HCC.

They also detected significantly lower protein and mRNA levels of the HBV receptor NTCP (sodium taurocholate co-transporting polypeptide) in HCC tissues than in non-HCC liver tissues. Unlike HBV proteins, reduced NTCP expression in HCC tissue was significantly associated with reduced survival, indicating its potential relevance in disease progression.

VIII. αSat and centromeric biology

1. Complete sequencing of αSat

Human centromeres are located within large arrays of tandemly repeated DNA sequences known as α Sat, which often span millions of base pairs (Mbp) on each chromosome. Although HGP was declared completion in 2003, several genomic regions remained unresolved due to sequencing limitations. These included centromeric, telomeric, and ribosomal DNA regions. The "real" completion of HGP was achieved in 2022 under the telomere-to-telomere assembly project [54), 55]. In this assembly, satellite

repeats constitute 6.2% of the human genome assembly, with αSat representing the largest component, 2.8% of the genome [54]. Each αSat array is more likely to interact with centromere protein A (CENP-A), an essential inner kinetochore protein required for normal chromosome segregation during mitosis [56]. With the availability of complete αSat sequences, the structure and function of αSat DNA have become a rapidly growing area of research, particularly in the context of human diseases, including cancers.

2. Centromere function and its role in aneuploidy

Aneuploidy, defined as an abnormal number of chromosomes, is a hallmark of cancer, that was first described well over a century ago [57]. The causes of aneuploidy have been understood as chromosome instability, resulting in the gain or loss of chromosomes or chromosome fragments. The centromere breakage is a major initiating factor leading to aneuploidy and the resulting changes in the selective landscape that drive most cancers [57].

In this context, HBV DNA integration into αSat sequences, particularly when accompanied by the reforming may disrupt the structural integrity of αSat. This disruption could directly contribute to the generation of aneuploidy by impairing normal centromere function and chromosome segregation.

IX. Clinical Implications

1. Translation of genetic landscapes into clinical practice

Comprehensive molecular and genomic analyses of large HCC cohorts have now uncovered the landscape of driver genes, characteristic mutational signatures, and molecular classification of tumor subtypes. To translate this growing body of knowledge for better diagnosis, treatment, and prevention of this intractable cancer subtype, further advances are needed in genetic screening and clinical sequencing to facilitate early diagnostic tools, such as liquid biopsy, to optimize and individualize treatment strategies, and to accelerate translational drug developments, especially targeting *TERT* (anti-*TERT* drugs), which may represent a promising avenue.

2. Liquid biopsy

Liquid biopsy detects cell free (cf) and/or circulating tumor (ct) DNA, which reflects various types of biological information from the tumor and serves as sensitive tools for non-invasive detection of early-stage HCC [58].

Methylated cf/ctDNA is a cancer-specific DNA fragment that is one of the most advanced tools for early HCC detection.

Suehiro et al. [59] have developed a new assay called combine restriction digital PCR (CORD) assay, which enables counting each copy of a methylated gene in a small amount of DNA. Using this assay, Kotoh et al. [60] found that the median copy number of methylated *SEPT9* (m-*SEPT9*) was significantly higher in patients with HCC than patients with chronic liver diseases and healthy volunteers. This assay showed a sensitivity of 63% and a specificity of 90% for detecting HCC across all stages. A meta-analysis performed by Chandrapalan et al. [61] included six case-control studies of m-*SEPT9* conducted in Japan as mentioned above as well as in the USA and in Europe and showed a pooled sensitivity of 80% and a pooled specificity of 90% for detecting HCC of all stages. Thus, m-*SEPT9* has shown promise in detecting HCC with high estimated diagnostic performance.

Chalasani et al. [62), 63)] have researched methylated marker sets of three genes, *HOXA1*, *B3GALT6*, and *TSPYL5*, combined with sex and serum level of α fetoprotein (AFP) (designated mt-HBT). This algorithm showed a sensitivity of 82% and a specificity of 87% for detecting early-stage HCC. Kunimune at al. [64] have examined a methylated marker of a gene, *HOXA1*, combined with age, sex, and serum levels of AFP together with des-γ-carboxyl prothrombin (DCP) (designated ASDAmH-1). This index demonstrated a sensitivity of 76% and a specificity of 78% (non-viral HCC) and 75% (viral HCC) for detecting early-stage HCC. Performances of the two methods were not affected by HCC ethiology and demonstrated similar numerical values.

Mutations of the *TERT* promoter and other driver gene regions have been studied for liquid biopsy. Qu et al. [65] have studied an assay using WGS followed by target-gene amplification using primers covering *TERT* promoter, *TP53*, *CTNNB1*, *AXIN1*, and HBV integration breakpoint in plasma cfDNA. They constructed an assay combining these gene mutations with serum levels of AFP and DCP (designated HCCscan). This assay showed a good sensitivity and a specificity for detecting HCC. Although the low positive predictive value for the occurrence of HCC (17%) warrants improvement, this assay poses a trigger point for visualizing mutations in driver gene regions in cfDNA.

Variations of α Sat and other repetitive sequence have been investigated for liquid biopsy. Annapragada et al. [66] have developed a method called an alignment-free, genome-wide approach for analyzing repeat elements in diseases (ARTEMIS), which uses short k-mer sequences in plasma cfDNA. This method successfully identified tumor-specific changes in repeat elements, including centromeric satellite DNAs, LINEs, SINEs, long terminal repeats, and

transposable elements in cfDNA in early-stage HCC and lung cancer among more than ten kinds of cancers investigated. This method is a good example of emerging αSat and other repetitive sequences in the clinical fields of cancers in cfDNA.

Alterations of chromosomes have been examined for liquid biopsy. Zhang et al. [67] have improved an approach called DNA evaluation of fragments for early interception (DELFI), which constructs WGS and analyzes the degree of DNA fragmentation in plasma cfDNA. This approach was revealed to be useful for distinguishing chromosome lengths in cfDNA in early-stage HCC. This assay is a representative breakthrough toward identifying chromosomal gains and losses in cancers in cfDNA.

MicroRNAs (miRNAs) are essentially 18 to 22-nucleotide-long endogenous noncoding RNAs. The effects of miRNAs on the regulation of expression of various genes are very broad. Several reports have shown aberrant expression of specific miRNAs. Circulating blood miRNAs, which are highly stable in a cell-free form, have shown promise as novel potential biomarkers for the early detection of HCC and there have been several miRNAs that have been demonstrated to be useful in the HCC diagnosis [68), 69)].

Among several miRNAs, miR-122 is a liver-specific, highly expressed miRNA [70)]. A meta-analysis performed by Zhao et al. [71)] has shown that miR-122 confered moderate efficacy for discriminating patients with HCC from healthy controls or patients with HBV infection. Furthermore, Trung et al. [72)] reported that combination assays of circulating miR-122 and *TERT* promoter exhibited satisfactory diagnostic performance in discriminating patients with HCC from other patients.

3. Therapy using anti-*TERT*

Therapies that target telomerase have been viewed as a highly attractive means of cancer treatment. The disruption of the telomerase maintenance mechanism is important in achieving replicative immortality, which is fundamental characteristics in various cancers, including HCC. Recent main telomerase-targeting therapies are antisense oligonucleotides (ASO), nucleoside analogues (NA), small molecule inhibitors, and stabilizers of G-quadruplexs (see Refs 73, 74 for revews)

Among these, ASO and NA have recently been reported as promising candidates for HCC therapy.

Ningarhari et al.[75)] found that ASO was efficient in highly proliferative and poorly differentiated cells. This revealed oncogenic addiction to *TERT* in HCC, providing a rationale for anti-*TERT* ASO treatment in HCC.

Mender et al. [76), 77)] reported that NA, 6-thio-dG (THIO), which was concerted into the corresponding 5'-triphosphate and incorporated into telomeres by telomerase, was effective in controlling tumor growth, especially when combined with immune check-point inhibitors in a T cell-dependent manner, providing a rationale for combining anti-*TERT* NA therapy and immunotherapy in HCC.

X. Perspectives

Robust HBV replication activity measured by peripheral serum HBeAg and HBV DNA levels has been significantly correlated with increased frequency of HBV integration events. Because such a virological state can persist for decades (e.g., the immune-tolerant phase of chronic HBV infection following neonatal infection), individuals with persistent HBV infection may experience a long window period during which hepatocarcinogenesis is initiated.

It is likely that the initiation of HBV-related HCC, one of the cancers driven by viral genome-integration, differs fundamentally from cancers of non-viral genome-integration etiology, including other forms of HCC. HBV may act as a molecular trigger that directly interacts with critical host genomic elements in hepatocytes [6), 8)]. Depending on the context, these interactions may induce genomic or proteomic alterations that activate carcinogenic signaling pathways, ultimately driving tumorigenesis in ways that are similar to other cancer types.

Fortunately, current preventive strategies such as vaccination [78), 79)] and antiviral therapy [80)] can prevent HCC occurrence. Antiviral therapy in the early life of HBV-infected individuals, together with further development in molecular and biological technologies supported by the complete human genome sequence, including αSat, is expected to eliminate HBV-related HCC.

Conflicts of interest

The author declares no conflict of interest.

Funding

None

References

1) Maupas P, Melnick JL. Hepatitis B infection and primary liver cancer. Prog Med Virol. 1981; 27: 1-5.

2) Beasly RP. Hwang LY, Lin CC, et al. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. Lancet. 1981; 2 (8256): 1129-33. doi: 10.1016/s0140-6736 (81) 90585-7.

3) Chen C-J, Yang H-I, Su J, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA. 2006; 295 (1): 65-73. doi: 10.1001/jama 295.1.65.

4) Zapatka M, Borozan I, Brewer DS, et al. The landscape of viral associations in human cancers. Nat Genet. 2020; 52 (3) : 320-30. doi: 10.1038/s41588-019-0558-9. Epub 2020 Feb 5.
5) Ogata N, Ichida F, Hamada C, et al. Mode of integration of hepatitis B virus DNA in chronically infected lives with indications of multiclonal growth of hepatocytes and some hepatoma cells. In: Zuckermann AJ, editor. Viral Hepatitis and Liver Disease. New York: Ala R. Liss, Inc; 1988. p. 746-51.
6) Ogata N. Mechanisms of hepatocarcinogenesis induced by hepatitis viruses: HBV DNA integration into hepatocytes and development of hepatocellular carcinoma. In: Ichida F, editor. New Development of Research of Liver Disease. Vol. III3 Tokyo: Medical View; 1998. p. 211-26. (In Japanese)
7) Ogata N, Tokino T, Kamimura T, et al. A comparison of the molecular structure of integrated hepatitis B virus genomes in hepatocellular carcinoma cells and hepatocytes derived from the same patient. Hepatology. 1990; 11 (6) : 1017-23. doi: 10.1002/hep.1840110617.
8) Matsubara K, Tokino T. Integration of hepatitis B virus DNA and its implications for hepatocarcinogenesis. Mol Biol Med. 1990; 7 (3) : 243-60.
9) Fujiyama A, Miyanohara A, Nozaki C, et al. Cloning and structural analyses of hepatitis B virus DNAs, subtype adr. Nucleic Acids Res. 1983; 11 (13) : 4601-10. doi: 10.1093/nar/11.13.4601.
10) Dejean A, Bougueleret L, Grzeschik KH, et al. Hepatitis B virus DNA integration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma. Nature. 1986; 322 (6074) : 70-2. doi: 10.1038/322070a0.
11) Zhang XK, Egan JO, Huang D, et al. Hepatitis B virus DNA integration and expression of an erb B-like gene in human hepatocellular carcinoma. Biochem Biophys Res Commun. 1992; 188 (1) : 344-51. doi: 10.1016/0006-291x (92) 92391-a.
12) Wang J, Chenivesse X, Henglein B, et al. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. Nature. 1990; 343 (6258) : 555-7. doi: 10.1038/343555a0.
13) Tsuei D-J, Chang M-H, Chen P-J, et al. Characterization of integration patterns and flanking cellular sequences of hepatitis B virus in childhood hepatocellular carcinomas. J Med Virol. 2002; 68 (4) : 513-21. doi: 10.1002/jmv.10240.
14) Zhou YZ, Slagle BL, Donehower LA, et al. Structural analysis of a hepatitis B virus genome integrated into chromosome 17p of a human hepatocellular carcinoma. J Virol. 1988; 62 (11) : 4224-31. doi: 10.1128/JVI.62.11.4224-4231.1988
15) Slagle BL, Zhou YZ, Butel JS. Hepatitis B virus integration event in human chromosome 17p near the p53 gene identifies the region of the chromosome commonly deleted in virus-positive hepatocellular carcinomas. Cancer Res. 1991; 51 (1) ; 49-54.
16) Berger I, Shaul Y. Integration of hepatitis B virus: analysis of unoccupied sites. J Virol. 1987; 61 (4) : 1180-6. doi: 10.1128/JVI.61.4.1180-1186.1987.
17) Shaul Y, Garcia PD, Schonberg S, et al. Integration of hepatitis B virus DNA in chromosome-specific satellite sequences. J Virol. 1986; 59 (3) : 731-4. doi: 10.1128/JVI.59.3.731-734.1986.
18) 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 (8) : 773-82. doi: 10.1101/gad.1.8.773.
19) Meyer M, Wiedorn KH, Hofschnerider PH, et al. A chcomosome 17:7 translocation is associated with a hepatitis B virus DNA integration in human hepatocellular carcinoma DNA. Hepatology. 1992; 15 (4) : 665-71. doi: 10.1002/hep.1840150419.
20) Tsuei DJ, Hsu TY, Chen JY, et al. Analysis of integrated hepatitis B virus DNA and flanking cellular sequences in a childhood hepatocellular carcinoma. J Med Virol. 1994; 42 (3) : 287-93. doi: 10.1002/jmv.1890420316.
21) Quade K, Saldanha J, Thomas H, et al. Integration of hepatitis B virus DNA through a mutational hot spot within the cohesive region in a case of hepatocellular carcinoma. J Gen Virol. 1992; 73 (Pt 1) : 179-82. doi: 10.1099/0022-1317-73-1-179.
22) Fowler MJ, Thomas HC, Monjardino J. Cloning and analysis of integrated hepatitis B virus DNA of the adr subtype derived from a human primary liver cell carcinoma. J Gen Virol. 1986; 67 (Pt 4) : 771-5. doi: 10.1099/0022-1317-67-4-771.
23) Matsumoto H, Yoneyama T, Mitamura K, et al. Analysis of integrated hepatitis B virus DNA and cellular flanking sequences cloned from a hepatocellular carcinoma. Int J Cancer. 1988; 42 (1) : 1-6. doi: 10.1002/ijc.2910420102.
24) Chen JY, Harrison TJ, Tsuei DJ, et al. Analysis of integrated hepatitis B virus DNA and flanking cellular sequences in the hepatocellular carcinoma cell line HCC36. Intervirology. 1994; 37 (1) : 41-6. doi: 10.1159/000150355.
25) Mizusawa H, Taira M, Yaginuma K, et al. Inversely repeating integrated hepatitis B virus DNA and cellular flanking sequences in the human hepatoma-derived cell line huSP. Proc Natl Acad Sci USA. 1985; 82 (1) : 208-12. doi: 10.1073/pnas.82.1.208.
26) Yagimuma K, Kobayashi M, Yoshida E, et al. Hepatitis B virus integration in hepatocellular carcinoma DNA: duplication of cellular flanking sequences at the integration site. Proc Natl Acad Sci USA. 1985; 82 (13) : 4458-62. doi: 10.1073/pnas.82.13.4458.
27) Pineau P, Marchio A, Mattei MG, et al. Extensive analysis of duplicated-inverted hepatitis B virus integration in human hepatocellular carcinoma. J Gen Virol. 1998; 79 (Pt 3) : 591-600. doi: 10.1099/0022-1317-79-3-591.
28) Rogler CE, Sherman M, Su CY, et al. Deletion in chromosome 11p associated with a hepatitis B integration site in hepatocellular carcinoma. Science. 1985; 230 (4723) : 319-22. doi: 10.1126/science.2996131.
29) Hino O, Shows TB, Rogler CE. Hepatitis B virus integration site in hepatocellular carcinoma at chromosome 17;18 translocation. Proc Natl Acad Sci USA. 1986: 83 (21) ; 8338-42. doi: 10.1073/pnas.83.21.8338.

30) Tokino T, Fukushige S, Nakamura T, et al. Chromosomal translocation and inverted duplication associated with integrated hepatitis B virus in hepatocellular carcinomas. J Virol. 1987; 61(12): 3843-54. doi: 10.1128/JVI.61.12.3843-3854.1987.

31) Pineau P, Marchio A, Terris B, et al. A t(3;8) chromosomal translocation associated with hepatitis B virus integration involves the carboxypeptidase N locus. J Virol. 1996; 70(10): 7280-4. doi: 10.1128/JVI.70.10.7280-7284.1996.

32) Becker SA, Zhou YZ, Slagle BL. Frequent loss of chromosome 8p in hepatitis B virus-positive hepatocellular carcinomas form China. Cancer Res. 1996; 56(21): 5092-7.

33) Totoki Y, Tatsuno K, Covington KR, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet. 2014; 46(12); 1267-73. doi: 10.1038/ng.3126. Epub 2014 Nov 2.

34) Fujimoto A, Furuta M, Totoki Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet. 2016; 48(5): 500-9. doi: 10.1038/ng.3547. Epub 2016 Apr 11.

35) Shibata T, Arai Y, Totoki Y. Molecular genomic landscapes of hepatobiliary cancer. Cancer Sci. 2018; 109(5): 1282-91. doi: 10.1111/cas.13582. Epub 2018 Apr 28.

36) Schulze K, Imbeaud S, Letouzé E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet. 2015; 47(5): 505-11. doi: 10.1038/ng.3252. Epub 2015 Mar 30.

37) Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell. 2017; 169(7): 1327-41. e23. doi: 10.1016/j.cell.2017.05.046.

38) Zhao L-H, Liu X, Yan H-X, et al. Genomic and oncogene preference of HBV integration in hepatocellular carcinoma. Nat Commun. 2016; 7: 12992. doi: 10.1038/ncomms12992.

39) Li W, Zeng X, Lee NP, et al. HIVID: an efficient method to detect HBV integration using low coverage sequencing. Genomics. 2013; 102(4): 338-44. doi: 10.1016/j.ygeno.2013.07.002. Epub 2013 Jul 15.

40) Rajaby R, Zhou Y, Meng Y, et al. SurVirus: a repeat-aware virus integration caller. Nucleic Acids Res. 2021; 49(6): e33. doi: 10.1093/nar/gkaa1237.

41) Sung W-K, Zheng H, Li S, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet. 2012; 44(7): 765-9. doi: 10.1038/ng.2295.

42) Rajaby R, Liu D-X, Au CH, et al. INSurVeyor: improving insertion calling from short read sequence data. Nat Commun. 2023; 14(1): 3243. doi: 10.1038/s41467-023-38870-2.

43) Rajaby R, Sung W-K. SurVIndel2: improving copy number variant calling from next-generation sequencing using hidden split reads. Nat Commun. 2024; 15(1): 10473. doi: 10.1038/s41467-024-53087-7.

44) Zhuo Z, Rong W, Li H, et al. Long-read sequencing reveals the structural complexity of genomic integration of HBV DNA in hepatocellular carcinoma. NPJ Genom Med. 2021; 6(1): 84. doi: 10.1038/s41525-021-00245-1.

45) Álvarez EG, Demeulemeester J, Otero P, et al. Aberrant integration of hepatitis B virus DNA promotes major restructuring of human hepatocellular carcinoma genome architecture. Nat Commun. 2021; 12(1): 6910. doi: 10.1038/s41467-021-26805-8.

46) Péneau C, Imbeaud S, Bella TL, et al. Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma. Gut. 2022; 71(3): 616-26. doi: 10.1136/gutjnl-2020-323153. Epub 2021 Feb 9.

47) Qian Z, Liang J, Huang R, et al. HBV integrations reshaping genomic structures promote hepatocellular carcinoma. Gut. 2024; 73(7): 1169-82. doi: 10.1136/gutjnl-2023-330414.

48) Bill CA, Summers J. Genomic DNA double-strand breaks are targets of hepadnaviral DNA integration. Proc Natl Acad Sci USA. 2004; 101(30): 11135-40. doi: 10.1073/pnas.0403925101.

49) Li W, Wei W, Hou F, et al. The integration model of hepatitis B virus genome in hepatocellular carcinoma cells based on high-throughput long-read sequencing. Genomics. 2022; 114(1): 23-30. doi: 10.1016/j.ygeno.2021.11.025. Epub 2021 Nov 26.

50) Ogata N, Miller RH, Ishak KG, et al. The complete nucleotide sequence of a pre-core mutant of hepatitis B virus implicated in fulminant hepatitis and its biological characterization in chimpanzees. Virology. 1993; 194(1): 263-76. doi: 10.1006/viro.1993.1257.

51) Ogata N, Miller RH, Ishak KG, et al. Genomic and biological characterization of two hepatitis B virus variants: A pre-core mutant implicated in fulminant hepatitis and a surface mutant resistant to immunoprophylaxis. In: Nishioka K, Suzuki H, Mishiro S, et al., editors. Viral Hepatitis and Liver Disease. Tokyo: Springer-Verlag; 1994. p. 238-42.

52) Ogata N, Fujii K, Takigawa S, et al. Novel patterns of amino acid mutations in the hepatitis B virus polymerase in association with resistance to lamivudine therapy in Japanese patients with chronic hepatitis B. J Med Virol. 1999; 59(3): 270-6.

53) Gao Q, Zhu H, Dong L, et al. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. Cell. 2019; 179(2): 561-77. e22. doi: 10.1016/j.cell.2019.08.052.

54) Altemose N, Logsdon GA, Bzikadze AV, et al. Complete genomic and epigenetic maps of human centromeres. Science. 2022; 376(6588): eabl4178. doi 10.1126/science.abl4178. Epub 2022 Apr 1.

55) Altemose N. A classical revival: Human satellite DNAs enter the genomics era. Semin Cell Dev Biol. 2022; 128: 2-14. doi: 10.1016/j.semcdb.2022.04.012. Epub 2022 Apr 27.

56) Nassar R, Thompson L, Fouquerel E. Molecular mechanisms protecting centromeres from self-sabotage and implications for cancer therapy. NAR Cancer. 2023; 5(2): zcad019. doi: 10.1093/narcan/zcad019. eCollection 2023 Jun.

57) Zheng Y, Ahmad K, Heinkoff S. Total whole-arm chromosome losses predict malignancy in human cancer. Proc Natl

Acad Sci USA. 2025; 122(18): e2505385122. doi: 10.1073/pnas.2505385122. Epub 2025 May 2.

58) Lee Y-T, Fujiwara N, Yang JD, et al. Risk stratification and early detection biomarkers for precision HCC screening. Hepatology. 2023; 78(1): 319-62. doi: 10.1002/hep.32779. Epub 2022 Oct 11.

59) Suehiro Y, Hashimoto S, Higaki S, et al. Blood free-circulating DNA testing by highly sensitive methylation assay to diagnose colorectal neoplasias. Oncotarget. 2018; 9(24): 16974-87. doi: 10.18632/oncotarget.24768.

60) Kotoh Y, Suehiro Y, Saeki I, et al. Novel liquid biopsy test based on a sensitive methylated *SEPT9* assay for diagnosing hepatocellular carcinoma. Hepatol Commun. 2020; 4(3): 461-70. doi: 10.1002/hep4.1469. eCollection 2020 Mar.

61) Chandrapalan S, Bannaga A, Weidner A, et al. A systematic review and meta-analysis: the diagnostic accuracy of methylated *SEPTIN9* for the detection of hepatocellular carcinoma and the clinical evaluation of its use in combination with other surveillance modalities. Scand J Gastroenterol. 2022; 57(4): 473-80. doi: 10.1080/00365521.2021.2020331. Epub 2021 Dec 25.

62) Chalasani NP, Ramasubramanian TS, Bhattacharya A, et al. A novel blood-based panel of methylated DNA and protein markers for detection of early-stage hepatocellular carcinoma. Clin Gastroenterol Hepatol. 2021; 19(12): 2597-605.e4. doi: 10.1016/j.cgh.2020.08.065. Epub 2020 Sep 2.

63) Chalasani NP, Porter K, Bhattacharya A, et al. Validation of a novel multitarget blood test shows high sensitivity to detect early stage hepatocellular carcinoma. Clin Gastroenterol Hepatol. 2022; 20(1): 173-82.e7. doi: 10.1016/j.cgh.2021.08.010. Epub 2021 Aug 13.

64) Kunimune Y, Suehiro Y, Saeki I, et al. Combination assay of methylated *HOXA1* with tumor markers shows high sensitivity for detection of early-stage hepatocellular carcinoma. Liver Cancer. 2024; 13(5): 487-97. doi: 10.1159/000536211. eCollection 2024 Oct.

65) Qu C, Wang Y, Wang P, et al. Detection of early-stage hepatocellular carcinoma in asymptomatic HBsAg-seropositive individuals by liquid biopsy. Proc Natl Acad Sci USA. 2019; 116(13): 6308-12. doi: 10.1073/pnas.1819799116. Epub 2019 Mar 11.

66) Annapragada AV, Niknafs N, White J, et al. Genome-wide repeat landscapes in cancer and cell-free DNA. Sci Transl Med. 2024; 16(738): eadj9283. doi: 10.1126/scitranslmed.adj9283. Epub 2024 Mar 13.

67) Zhang X, Wang Z, Tang W, et al. Ultrasensitive and affordable assay for early detection of primary liver cancer using plasma cell-free DNA fragmentomics. Hepatology. 2022; 76(2): 317-29. doi: 10.1002/hep.32308. Epub 2022 Jan 26.

68) Yamamoto Y, Kondo S, Matsuzaki J, et al. Highly sensitive circulating microRNA panel for accurate detection of hepatocellular carcinoma in patients with liver disease. Heptol Commun. 2019; 4(2): 284-97. doi: 10.1002/hep4.1451. eCollection 2020 Feb.

69) Jin Y, Wong YS, Goh BKP, et al. Circulating microRNAs as potential diagnostic and prognostic biomarkers in hepatocellular carcinoma. Sci Rep. 2019; 9(1): 10464. doi: 10.1038/s41598-019-46872-8.

70) Masaki T, Arend KC, Li Y, et al. miR-122 stimulate hepatitis C virus RNA synthesis by altering the balance of viral RNAs engaged in replication versus translation. Cell Host Microbe. 2015; 17(2): 217-28. doi: 10.1016/j.chom.2014.12.014. Epub 2015 Feb 5.

71) Zhao X-F, Li N, Lin D-D, et al. Circulating microRNA-122 for the diagnosis of hepatocellular carcinoma: A meta-analysis. Biomed Res Int. 2020: 5353695. doi: 10.1155/2020/5353695. eCollection 2020.

72) Trung NT, Hoan NX, Trung PQ, et al. Clinical significance of combined circulating TERT promoter mutations and miR-122 expression for screening HBV-related hepatocellular carcinoma. Sci Rep. 2020; 10(1): 8181. doi: 10.1038/s41598-020-65213-8.

73) Suzuki H, Mishra S, Paul S, et al. Molecular and immune landscape of hepatocellular carcinoma for therapeutic development. J Liver Cancer. 2025; 25(1): 9-18. doi: 10.17998/jlc.2024.12.02. Epub 2024 Dec 6.

74) Gao J, Pickett HA. Targeting telomerases: Advances in telomere maintenance mechanism-specific cancer therapies. Nat Rev Cancer. 2022; 22(9): 515-32. doi: 10.1038/s41568-022-00490-1. Epub 2022 Jul 5.

75) Ningarhari M, Caruso S, Hirsh TZ, et al. Telomere length is key to hepatocellular carcinoma diversity and telomerase addiction is an actionable therapeutic target. J Hepatol. 2021; 74(5): 1155-66. doi: 10.1016/j.jhep.2020.11.052. Epub 2020 Dec 15.

76) Mender I, Siteni S, Barron S, et al. Activating an adaptive immune response with a telomerase-mediated telomere targeting therapeutic in hepatocellular carcinoma. Mol Cancer Ther. 2023; 22(6): 737-50. doi: 10.1158/1535-7163.MCT-23-0039.

77) Mender I, Girotti R, Gryaznov S. Nobel telomere-targeting dual-pharmacophore dinucleotide prodrugs for anticancer therapy. Nucleic Acids Res. 2025; 53(12): gkaf591. doi: 10.1093/nar/gkaf591.

78) Wong G L-H, Hui V W-K, Yip T C-F, et al. Universal HBV vaccination dramatically reduces the prevalence of HBV infection and incidence of hepatocellular carcinoma. Aliment Pharmacol Ther. 2022; 56(5): 869-77. doi: 10.1111/apt.17120. Epub 2022 Jul 21.

79) Ogata N, Cote PJ, Zanetti AR, et al. Licensed recombinant hepatitis B vaccines protect chimpanzees against infection with the prototype surface gene mutant of hepatitis B virus. Hepatology. 1999; 30(3): 779-86. doi: 10.1002/hep.510300309.

80) Udompap P, Kim WR. Development of hepatocellular carcinoma in patients with suppressed viral replication: changes in risk over time. Clin Liver Dis (Hoboken). 2020; 15(2): 85-90. doi: 10.1002/cld.904. eCollection 2020 Feb.