UNIVERSITA' DEGLI STUDI DI MESSINA

DIPARTIMENTO DI MEDICINA CLINICA E SPERIMENTALE

Dottorato di Ricerca in

Biotecnologie Mediche e chirurgiche

XXXII ciclo

Coordinatore: Ch.mo Prof.re G. Squadrito

Study of HBV DNA integration in patients with HBV-related HCC by a high-throughput viral integration detection method

Tesi di dottorato di:

Dott.ssa Francesca Casuscelli di Tocco

Tutor:

Prof.ssa Teresa Pollicino

Summary :

I. Introduction

1. Hepatitis B virus

1a. Epidemiology

Hepatitis B virus (HBV) is a small enveloped DNA virus, which is a member of the Hepadnaviridae family and it is responsible for causing acute and chronic hepatitis in humans. It is currently estimated that 250 million of people worldwide are chronically infected with the HBV virus. Immune-mediated liver damage in these individuals may lead to the development of cirrhosis and hepatocellular carcinoma later in life[1]. In highly endemic areas, hepatitis B is most commonly spread from mother to child at birth (perinatal transmission), or through horizontal transmission (exposure to infected blood)[2]. The hepatitis B vaccine is the pillar of hepatitis B prevention. Although, there is important evidence showing that vaccine-induced escape mutants can be selected[3]. Moreover, during chronic HBV infection different selective pressures may cause selection and emergence of HBV variants that can evade prophylactic and therapeutic interventions[4]. Nine HBV different genotypes have been detected and numbered alphabetically from A to I, with approximately 8% intergroup nucleotide difference across the complete genome, futhermore genotypes A–D, F, H, and I are classified into at least 35 subgenotypes[5]. HBV genotypes and subtypes present different geographical distribution (fig.1), genotype A is distributed globally and is the main genotype found in Europe, North America, Africa and India. Genotypes B and C are predominant in East and Southeast Asia[6], genotype D is mainly found in Eastern Europe, Mediterranean countries and in the Middle East. The infections with genotypes E, F, G and H are usually observed in West Africa, Central-South America, Central Europe and Southern USA, respectively[7]. Recently, HBV genotype I was described in Northwest China, Vietnam and Laos[6]. It has been show that a superinfection, due to the presence of two or more genotypes that coinfect the same patient can be associated with a acute intensification of the chronic disease[8]. Futhermore it has been demonstrated that among all the genotypes, genotype C is the one that shows a higher association with the development of cirrhosis and HCC compared with the other genotypes[9].

1b. HBV virology

Hepatitis B virus (HBV) is a DNA virus characterized by liver tropism and species specificity. It belongs to the Hepadnaviridae family, which includes two genera: the Orthohepadnavirus genus, which contains members like HBV that infect mammals and the Avihepadnavirus genus, which infects birds[10]. HBV infectious virion has a spherical structure of 40-42nm and it is also called « Dane particle ». It consists of a lipoproteic envelope which contains the surface antigen (HBsAg) and of internal capsid « core particle » which is formed by the core protein (HBcAg), containing a relaxed circular, partially double stranded DNA genome (rcDNA) of 3.2 kb and the viral polymerase (P) (fig.2).

The HBV genome is composed by two linear strands: one complete negative DNA strand (-) of 3.2 Kb and one positive DNA strand (+), with a variable length (50/100% of the negative strand). The 5' end of (-) DNA strand is covalently linked to P protein, while 5' end of the (+) DNA strand consists of an RNA oligonucleotide, derived from a pre-genomic RNA (pgRNA), which serves as the primer for its synthesis. The complementarities between the 5' regions of both strands create a "cohesive end region" that makes possible the circularization of the genome. In this region are present two direct repeats of 11 nucleotides called DR1 and DR2 that show an important role in the viral replication[11]. The HBV genome shows a compact organization that consists of four overlapping open reading frames (ORFs) covering the entire genome (fig.3). The pre-S/S ORF encodes the three viral surface proteins: Large (LHBs), Medium (MHBs) and Small (SHBs) proteins. It contains the major viral antigenic domains. All three envelope components are glycosylated, type II transmembrane proteins that can form multimers stabilized by disulfide bridges formed by cysteine residues present in the S domain. These proteins are generated by alternative translation initiation from in-frame AUG codons and share the same carboxy-terminus part but have different amino-terminal extensions. The S-protein (226 amino acids long) is the smallest one and defines the S domain (HBsAg), the M-protein contains an extra N-terminal extension of 55 aa, whereas the L-protein has a further N-terminal sequence of 108 or 119 aa – depending on the genotype – compared to the M-protein. The S-protein contains the anti-HBs antibody neutralization domain, known as ''a'' determinant (99–170 aa)[12]. The Pol ORF encodes a multifunctional protein, that posseses DNA-dependent DNA polymerase, reverse transcriptase (RT) and RNaseH activities and it is also include a terminal protein (TP) domain that acts as a primer for HBV DNA synthesis. The preS/S ORF completely overlaps with the Pol ORF. The

precore (PC)/core encodes two different proteins: the structural core protein of the viral nucleocapsid (the hepatitis B core antigen, HBcAg) and the non-structural protein also known as secreted e-antigen (HBeAg). Expression of PC/core and core proteins depends on translation initiation from the precore AUG and core AUG, respectively. The X ORF, encodes the small regulatory X protein (154 amino-acid residues), which is essential for viral replication and functions as transcriptional transactivator of many cellular and/or viral genes. All regulatory elements like enhancers, promoters polyadenylation signal, transcription start sites, necessary for the biological activity of the virus are present within the four coding reagions[13].

1c. Genome replication

HBV infection is charachterized by different steps: firstly, there is an interaction of low affinity between the virus and the cell surface proteoglycans. Glycoproteins, glycolipids, and proteoglycans are constitutive components of the extracellular matrix and the plasma membrane. HBV infection depends on these cell surface structures as primary attachment sites. Secondly, the HBV-cell interaction occurs through a high affinity interaction between the PreS1 domain of HBV and the bile salt transporter receptor of hepatocytes. Indeed, Pres1 and S proteins have a principal role in the viral entry, in contrast with the Pres2 region that it is not involved in the viral infection[14]. In particular, the preS1 protein has been shown to be essentially involved in the HBV infection binding to a hepatocyte-specific receptor. It is recently identified an HBV entry receptor, that can have a role in the HBV infection. It is the sodium taurocholate cotransporting polypeptide (NTCP) and it is a transporter residing in the basolateral membrane of hepatocytes. It is involved in the hepatic uptake of conjugated bile salts. Hepatocytes infection by HBV is followed by a endocytosis-mediated internalization of the virus and its fusion with the cellular membrane compartment, probably in an endosomal compartment[15]. After viral entry, via the nuclear pore complex, the relaxed circular DNA (rcDNA) is release into cell nucleus and it is converted in a circular covalently closed DNA (cccDNA), by the positive DNA strand synthesis and the ligation of the two complete strands. In this step, the polymerase, which is covalently attached to the 5’ end of the negative DNA strand and the short RNA oligomer from the 5’ end of the positive DNA are removed. cccDNA is very stable and can persist into the nucleus of the infected hepatocytes as a

minichromosome[16]. Importantly, cccDNA represents the template for the synthesis of all viral RNAs, including the pregenomic RNA, which are transcribed by the cellular RNA polymerase II. Four unspliced viral RNAs, 3.5, 2.4, 2.1, and 0.7 kb, are transcribed from their respective promoters and two enhancer regions and end at common polyadenylation signal located in the core open reading frame. The 3.5 kb RNA includes precore and pregenomic RNA species. Precore mRNA codes for precore antigen or HBeAg. The pregenomic RNA serves as a template for the synthesis of HBV DNA and also as the mRNA of core antigen (HBcAg) and polymerase. The 2.4 kb RNA codes for the the L envelope protein, the 2.1 kb RNA codes for the viral S antigen (HBsAg) and for the M protein, the 0.8 kb RNA codes for HBx protein (0.8 kb). All the transcripts contain 5' cap structures and all are 3' terminally poly-adenylated at common site[13]. When the transcription is completed, the RNAs are exported to the cytoplasm where they are translated into the different viral proteins. The role of pgRNA has been studied from long time and it is known that the pgRNA is the only viral transcript used as a template for the generation of new DNA genomes by reverse transcription mechanism. The pgRNA is longer than the viral genome, infact, it is characterized by a terminal redundancy (120 nucleotides), which contains a second copy of DR1 and the encapsidation signal ε, plus the poly-A tail[1,17]. In the cytoplasm the pgRNA and P protein are packaged into new subviral core particles, where the reverse transcription of the pgRNA into the new molecule of rcDNA takes place. There are different elements that are involved in this process and two of them are essential for the replication of the virus: the encapsidation signal ε and P protein. The reverse transcription starts when these two elements interact together. Following this step, four nucleotides are synthetized from the 5' end of the pgRNA, hence the nascent (-) DNA strand is transferred in the 3' end of the pgRNA, where it binds to DR1 repeat element. During the synthesis of the (-) strand, the RH domain of the polymerase simultaneously degrades the RNA, except for a capped 5' terminal region of the pgRNA that it is used for the synthesis of the positive DNA strand. When the (-) DNA strand is completed, the RNA oligomer is translocated to a complementary sequence present in the 3' DR2 domain to complete the synthesis of the new (+) DNA strand towards the 5' end of the (-) DNA strand. At the end of these transfer reactions a new rcDNA is synthetized. The new viral capsids containing mature rcDNA are firstly coated with the viral envelope proteins in the endoplasmic reticulum. It has been described that a « matrix » domain (MD), formed by the C-terminal portion of the PreS1 region of L, is required for the envelopment of the capsids[18]. The mature virions are enriched for L and S proteins, but not the M protein. When the HBV

capsids are completed they are released from the cell or can reshuttle to the nucleus to replenish cccDNA pool.

1d. Viral proteins

HBV virus shows a high compact organization of the genome and express a limited number of proteins. Virions contain three distinct but related surface proteins called S, M, and L, respectively. Each of these polypeptides is cotranslationally inserted into the endoplasmic reticulum (ER) membrane and span the bilayer multiple times[19,20,]. The small protein (SHBs) is constituted by 226 aa. It consists of three hydrophobic and two hydrophilic regions. Four transmembrane domains (TM 1-4), which are linked by external and internal loops, allow SHBs to cross the ER membrane. The loop of amino acids linking TM2 and TM3 (99–161aa) represents the 'a' determinant, which is the major antigenic determinant of HBV. The MHBs protein presents the same structure of the SHBs, but compared to the small protein it contains an extra N-terminal extension of 55 aa. The MHBs protein seems to be not necessary for the virus infectivity and its role remain not clear[21]. The Large (LHBs) protein is the longest envelope protein because, compared to the SHBs and MHBs proteins, it contains a further domain of 75aa. It has been demonstrated that LHBs is essential for viral entry, moreover the pre-S1 domain, which contains a myristic acid in the N-terminal extension, is critical for HBV infection. Furthermore LHBs is implicated in the encapsidation of the core, providing the ligand for core particles during the assembly of the viral envelope[22 23].

The core gene has two in-frame translation initiation codons (core and pre-core) and it encodes for two viral proteins, the core and the precore-core proteins, depending on where the translation begins. The core protein (21 kd), which is serologically defined as the hepatitis B core antigen (HBcAg), plays an essential role in the viral replication and in the viral assembly. HBcAg is a structural protein and represents the major constituent of the nucleocapsid, which presents an icosahedral structure[24,25]. The precore-core protein (25p) is transportated into the ER and there, before its secretion, it is N- and C-terminally processed. This mature precore-core protein is known as HBeAg (18 kd). HBeAg is a non-particulate protein and is not required for viral replication, infection, or capsid assembly, but it is implicated in the modulation of the host immune response through multiple pathways, such as by the depletion

of Th1 helpers cells limiting the cytotoxic T-lymphocyte (CTL) response against infected hepatocytes[26].

Among the viral proteins, the polymerase is the bigger HBV protein (832aa). The polymerase is a multifunctional protein and it consists of 4 different domains: the terminal protein (TP) domain, the spacer, the reverse transcriptase (RT) domain and the RNase domain. The TP domain represents the amino-terminal domain of the protein and it is necessary for the packaging of pgRNA and for the priming of minus strand DNA, moreover it allows the binding to the ε loop of the pgRNA[27]. The spacer domain is the less conserved region, its function is not known yet[28]. The RT domain is subdivised in 7 domains (A-G). In the C subdomain there is the tyrosine-metionine-aspartate-aspartate (YMDD). This motif is necessary for the DNA polymerase DNA-dependent-activity. The subdomain E exerts the reverse transcriptase activity, playing a role in the interaction of ε loop and polymerase, and also in the polymerization of DNA[29 30]. The RNase domain is located in the C terminal part of the protein and it permits the pgRNA degradation during the (-) DNA synthesis[30].

The HBx protein is smallest viral protein. It is essential for viral replication and significantly contributes to viral pathogenesis[31,32,33].

2. Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is the most common type of liver cancer accounting for approximately 70%-85% of cases[34]. Many different factors can stimulate the development of HCC: enviromental factors, genetic predisposition and viruses infection. The environmental risk factors include the exposure to the aflatoxin B1 (AFB1), that is a mycotoxin produced by fungi of the Aspergillus species; it grows on foods such as corn and peanuts stored in warm, damp conditions. AFB1 is considered a powerful hepatocarcinogen and it can contributes to the development of the HCC; particularly in patients with chronic HBV infection[35]. Alcohol abuse, tobacco smoking, obesity and diabetes (NASH associated) are other risk factors for HCC development. Have been detected different host parameters that can play an important role in the progress of the cancer such as gender: it has been show that male gender is more often associated with HCC development compared to female gender. It seems that men are at increased risk of HCC because they have much higher prevalence of viral hepatitis and alcoholic cirrhosis. This discrepancy can also be due to the fact that the female sex hormone

estrogen protects against the development and progression of HCC. Hemochromatosis, tyrosinemia, and α1-antitrypsin deficiency are also frequently associated with liver disease and HCC development. However, HCC typically occurs in the setting of chronic hepatitis virus infections. Hepatitis B and C account for approximately 80% of case worldwide. In case of HBV infection, HCC development is also associated with older age and early infection in childhood[36]. HBV infected patients have a higher risk to develop HCC as compared with non-infected subjects. In particular, it has been demonstrated that patients showing high viral load and hepatitis B e antigen (HBeAg) positivity have a stronger risk of developing liver cancer. Although, important evidence indicates that HBsAg-negative patients with occult HBV infection may also develop liver cancer[37]. Chronic HBV infection is a major global cause of HCC. HBV oncogenic role is exterted by its ability to cause liver inflamation, hepatic damage and subsequent cirrhosis, which significantly contribuites to HCC development. Moreover, many studies reported a correlation between the HCC development and the presence of particular mutations in specific regions of HBV genome. It has been described that both the T1762/A1764 double mutation in the Basal Core Promoter (BCP) and the alterations in the preS/S region (as, the mutations in the preS2 start codon, the presence of premature stop codon at position 172 or 182 of the S gene creating a truncated S protein, or deletions in the preS1 and preS2 sequences) are commonly associated with a higher risk of HCC. The mutated envelop proteins are accumulated inside the cells in the endoplasmic reticulum and the huge amount of these proteins within the hepatocytes causes oxidative stress, and can activate oncogenic cellular pathways[38].

2a. HCC and HBV integration

The HBV virus can exerts is oncogenic role in the development of HCC by three molecular mechanisms: 1) the expression of HBV proteins, like X protein (HBx) that can lead to cell proliferation and can activates oncogenic patways, 2) the integration of HBV DNA into the host genome that determines chromosomal instability, and causes an alteration in gene expression and in gene functionality (fig.4), 3) the genetic damage due to the inflammation determined by the chronic infection of hepatocytes. A huge number of studies have investigated the correlation between HBV DNA integration into the host genome and HCC development. It is well known that this integration process can modulate cellular growth and

differentiation by the alteration of the expression of cellular genes that are involved in these pathways. Many studies have been conducted to elucidate the mechanisms of virus integration into the host genome. Moreover, recently, they have been developed whole-genome and exome sequencing approaches that succeeded in obtaining important informations on the genetic changes occurring in HCC, such as mutations, deletions, translocations and copy number variations[36]. Concerning HBV DNA integration, it occurs randomly into the host genome involving multiple sites in different chromosomes[39]. The cells with an accumulation of genetic defects, due to the viral integration, show a selective growth advantage. Moreover, driver mutations can cause a trasformation of hepatocytes, leading to hepatocarcinogenesis. HBV integration can also determine the occurrence of copy number aberration, which may results in chromosomal focal amplifications of some oncogenes and less frequently in homozygous deletions of tumor suppressors[40]. Most of HBV insertion sites occur within or near regulatory regions. Many different human genes have been described as possible sites of virus integration, and the majority of these genes are involved in the regulation of proliferation, in cell survival and immortalization, such as hTERT, which encodes the catalytic subunit of telomerase, and is responsible for restoring telomere length during cellular immortalization[41], MLL4 (mixed lineage leukemia protein 4), CCNE1 (cyclin 1), SENP5 (sentrin-specific protease 5), ROCK1 (Rho-associated coiled-coil containg protein kinase1), TP53 tumor suppressor, and CTNNB1, which encodes the β-catenin[42]. Frequently HBV integrations have been also detected within or near repetitive, non coding sequences, such as long interspersed nuclear elements (LINEs), or near short interspersed nuclear elements (SINEs), and within the long terminal repeats (LTR) of endogenous retroviruses (ERVs)[38]. A recent study has reported a high prevalence of HBx-LINE1 chimeric transcripts in HCC[43]. It has been demonstrated that this chimera acts like a long noncoding RNA (lncRNA) implicated in the regulation of epithelial-to-mesenchymal transition (EMT), a process by which epithelial cells lose their cell polarity and cell-cell adhesion, and gain migratory and invasive properties. The HBx-LINE1 expression positively promotes the invasion of tumor cells lines by the activation of EMT phenomenon[44]. Other sites that have been described as hotspots for HBV DNA integration are the fragile sites of chromosomes. All chromosomes are characterizated by the presence of fragile sites, which are regions of the genome that are prone to break. These elements frequently represent the sites of chromosomal translocation, oncogene amplification and tumor suppressor deletion in cancer. The fragile sites cluster at G-light chromosomal bands, which are rich in Alu sequences. It has been recognized that their propensity for breakage and recombination is due to a tardive replication of these sites in the

cell cycle[39]. HBV may exerts its oncogenic activity also by the synthesis of alterated proteins, as mutated HBx or truncated preS/S proteins that have been associated with HCC development. Integrated HBV sequences frequently include a 3'-deleted X gene, and this may result in the production of truncated HBX proteins. In vitro studies have shown that truncated HBX proteins may lose their pro-apoptotic capacity and acquire the ability to induce uncontrolled cell proliferation. It has been reported that the truncated HBx proteins are able to induce cell proliferation because they lose the p53-dependent transcriptional repression binding site at 1637–1667 nt in the enhancer II region[45]. Another HBV protein that is correlated with HCC development is the Pres2 mutated protein. It has been described that 3' truncated preS/S sequences are implicated in tumor development in HBV-infected hepatocytes[46].

2b. Methods for the detection of HBV integrants

Various technical approaches have been described for the analysis of integrated HBV DNA. Southern blot hybridization, which uses restriction enzymes and DNA probes specific for HBV DNA detection is one of these approaches. This technique is able to discriminate between free viral DNA and HBV DNA sequences integrated in the host genome. HBV fragments are produced by enzymatic digestion, and the chimeric fragments are recognized because they are longer than the free HBV DNA[47]. This method was frequently used because it is relatively cheap and simple to carry out, but it presents some limits: it cannot identify the viral-human sequences or quantify the copy number of individual integrants. Furthermore, it shows a very low sensitivity (lower detection limit: 1–10 pg of HBV DNA). Another approach that have been used for the detection of virus-cell DNA junctions, is the Alu-PCR. This method takes advantage of the presence in the human genome of Alu elements (short DNA stretches of 300-bp) that are the most abundant repetitive elements i with over 1,000,000 copies[48]. This technique uses both HBV and Alu elements specific primers for the detection of the chimera sequences. Many studies described some problems related to the utilization of Alu PCR protocol for the detection of the HBV integrants, such as the the lack of detection of integrations that are not close to the Alu sequences. Inverse nested PCR (invPCR) is another method used to investigate HBV DNA integration. It is a sensitive method[49] used in several fields, such as in molecular biology, virology, genetic research because it is inexpensive and highly specific. Indeed, this technique does not require a huge

amount of DNA and small piece of liver tissues are required for the DNA extraction.[50] Recently new strategies have been developed for the detection of HBV integrants in the human genome. These strategies involve the use of the next generation sequencing (NGS) approch. Whole-genome sequencing (WGS) has been used to describe the HBV integration sites present in tumor and non-tumor samples providing data concerning recurrent somatic genetic alterations, including single nucleotide variations (SNVs), small insertions and deletions (indels), DNA copy number variations (CNVs), and structural variations (SVs)[51]. Although the WGS has made possible the detection of HBV integrants over the entire mappable genome, this technique presents some limitations, such as the request of huge amount of starting DNA (1µg or 10^6 cells), the failure to detect viral and human junctions that show a low frequency, and the excessive costs associated with the relative rarity of HBV integration detection[50]. Due to the limitations of WGS, an alternative methods has been developed to enrich for DNA fragments containing the viral-host chimeric junctions[27,52]. However, most data on HBV integration were generated by methods that favor preferential amplification and bias identification of unique integration sites, and a limitation of the more recent Next Generation Sequencing (NGS) approaches is the low coverage of HBV reads.

AIM

To detect HBV integration sites across the genome and decipher viral-host interaction, by applying a new a high-throughput targeted sequencing of enriched HBV integrants, in tumor (T) and non-tumor (NT) liver tissues from HBsAg-positive patients with HCC, on HBV-positive PLC/PRF/5 cells, and on 3 different HBV-negative tissue samples.

II. Materials and Methods

1. Tissue samples and cell line

Tumor tissue specimens from 7 HBsAg-positive HCC patients and paired non-tumor tissues, availabe only from 6 of them, as well as normal tissue specimens from 3 HBsAg-negative control patients were provided by the Oncology Surgery of the University Hospital of Messina, Italy. The PLC/PRF/5 (Alexander) human hepatoma cell line containing more integrated HBV DNA fragments was obtained from SIGMA, and was used as positive control

2. Library preparation

Integration library construction was performed using genomic DNA after β-globin quantification (Kit LightCycler Control kit DNA, Roche) to establish the number of cells included in each experiment. We analysed for each sample $10x10^7$ cells and we followed a specific workflow (fig.5).

3. Genomic DNA extraction and sonication

Tissue specimens were digested using a lysis buffer (150 mM Nacl, 10 nM tris Hcl pH 8, 10 nM EDTA, 0.5% SDS, PK 1mg/mL) and incubated overnight at 56°C. DNA was extracted with phenol/chloroform and resuspended in TE. Its concentration was determined using the Qubit fluorometer (Qubit, Life Technologies). DNA was then fragmented by sonication at low power for 3 cycles (15″on/15″off) using SONOPLUS HD 2070 (Bandelin) to yield a 500–1000 bp distribution of DNA fragments.

4. Genomic DNA fragment preparation and adaptor ligation

The following reactions were perfomed individually on 5µg aliquots of DNA. Firstly, the DNA was blunted using End-it DNA Repair Kit (Epicentre) according to the manufacturer's instruction. The DNA fragments were then purified using Qiagen PCR purification kit and eluted in 40 µl of water. Blunted DNA was adenosine-tailed at 3' for 1h at 37°C using a New England Biolabs kit (5 µl NEB buffer 10x, 1 µl dATP 10 mM, and 2 µl of Klenow fragment 3 -> 5' exo^- 5000U/mL). The reactions were then purified by Qiagen PCR Clean UP Kit and eluted in 40 µl of water. Each aliquot of blunted, A-tailed DNA fragments was ligated to 200 pmol of Adaptor Top and Adaptor Bottom by addition of 4 µl adaptors, 5 µl T4 DNA ligase buffer (NEB), and 1 µl of 2x 10^6 U/ml T4 DNA ligase (NEB), for 1 h at 25°C, then overnight at 16°C. Ligase was inactivated by incubation at 70°C for 20 min and the reactions were purified using Qiagen PCR Clean Up columns and eluted in 40 µl of water. Finally, all the reactions were pooled.

5. Round I PCR

Aliquots of 1 µg DNA were used to perform a semi-nested ligation-mediated PCR using either forward or reverse biotynilated HBV specific primers to detect HBV sequences that could be integrated into the human genome. Forward and reverse enrichment streams were kept separate for the entire reminder of the protocol. Each DNA aliquot was mixed with 10 µl Top Taq buffer (10x), 1.5 µl dNTPs (10 mM), 0.5 µl biotinylated primer (2.5 pM), 0.5 µl Top Taq (5U/µl, Qiagen) and H_2O to 50 µl of final volume. Single-primer PCR reactions were run: 1x (98°C- 1min) 12x (98°C- 15 s, 65°C- 30 s, 72°C- 45 s) 1x (72°C- 1 min) 1x (4-∞). Each tube was then spiked with 1 µl pAdaptor (2.5pM) and subjected to addictional cycles of PCR: 1x (98°C- 1min) 35x (98°C- 15 s, 65°C- 30 s, 72°C- 45 s) 1x (72°C- 5 min) 1x (4°C-∞). Forward and reverse PCR reactions were pooled separately, and were run on a 2% agarose gel until well separated, and appropriately sized fragments (0,3-1 kb) were excised. After this "size selection" the fragments were run on a 1.5% agarose gel to allow the bands to be re-assembled. The DNA was then purified in Qiagen gel purification columns; gel-based size selection and purification was repeated once. One hundred µl of washed T1 magnetic

streptavidin beads (Dynabeads MyOne streptavidin T1, Invitrogen) was resuspended in 400 µl 2x B&W buffer (10 mM Tris pH 7.5, 1 mM EDTA, 2 M NaCl), 200 µl was added to each forward and reverse PCR I products and the mixture was incubated for 1 h with gentle nutation at room temperature. Beads were magnetically isolated, washed 3x in 500 µl 1x B&W buffer, once in H_2O and resuspended in 50 µl H_2O.

6. Round II PCR, and Library Preparation

A second round of PCR was then performed, using HBV primers containing Illumina adapters (P5 or P7) for the binding to the Illumina MiSeq flow cell. The PCR mix contained: 20 µl Top Taq buffer (10x), 3 µl dNTPs (10mM), 1 µl HBV primer F or R (2.5 pM), including at their 5′ ends Illumina adaptor subsequences, 1 µl pAdaptor Fw or Rev (2.5 pM), 1 µl Top Taq (5 U/µl) and H_2O to obtain a final volume of 100 µl. PCR reactions were run: 1x (98°C- 1min) 35x (98°C- 10s, 65°C- 30s, 72°C- 40sec),1x (72°C- 5min), and 1x (4°C- ∞). The PCR reactions were then magnetically separated from beads, and each reaction was purified with Qiagen PCR Purification kit. The quality and the size of the amplified fragments were evaluated using Agilent 2100 Bioanalyzer. After the purification, an index PCR was conducted using a specific combination of indexes (Nextera XT Index kit, Illumina) for each purified reactions. The index PCR was performed using the Expand High Fidelity PCR System (Roche) as follows: 1x (94°C-3min), 25x (94°C-30sec, 65°C-30sec, 72°C-30sec) and 1x (72°C-10sec). Forward and reverse libraries for the same sample were mixed in equimolar ratios and sequenced by 250 bp paired-end sequencing on Illumina MiSeq.

7. Computational Analysis

Read Alignment

After quality check, raw reads were cleaned and resulting good quality reads were mapped to the hybrid Human-HBV genome (GRCh38.p10 and NC_003977.2).

Integration Determination

Chimeric reads were extracted using a combination of SAMtools, bedtools and in house scripts. HBV integration breakpoints were reconstructed using a combination of CAP3 and cd-hit. Chimeras were remapped with BLAST to the hybrid genome and the presence of microhomology across the breakpoints was evaluated. HBV integration events with a coverage lower than 3 reads were filtered out.

Hotspot Detection

Using the bedtools shuffle option, we performed the Monte Carlo test by shuffling randomly (100 times) each intersection corresponding to the human genome site in which the virus was found to be integrated.

Over-represented motifs

HOMER software was used to determine if specific human or viral sequences were found over-represented in the chimeric sequences. The analysis was performed by joining only human chimeric sequences, only viral chimeric sequences and all the entire chimeras found in this study. For the statistical purpose the software used the whole human and HBV genomes, including the hybrid human-virus sequence.

Statistical analysis

Data were statistically analyzed by means of the $\chi 2$ test for categorical data, and the Student t test for continuous data. Proportion test is the standard test for the difference between proportions, also known as a two-proportion z test. We also used R's implementation. $P < 0.05$ (two-tailed) was considered significant.

III. Results

A total of 5,539 HBV integration breakpoints (covered by a total of 44,141 chimeric reads) were detected in tumor tissue specimens from 7 HBsAg-positive HCC patients, in paired non-tumor tissues, availabe from 6 of them, and in PLC/PRF/5 human hepatoma cell (4,369 in T, 1,139 in NT, and 31 in cells). Among the 5,539 HBV integration sites, 3,211 were located in repeated DNA sequences. Most of the HBV integrations were found in SINEs (23% in T, 10% in NT, and 22,6% in cells), in simple repeats (12,5% in T, 28% in NT, and 0% in cells), in LINE (10,3% in T, 11,5% in NT and 9,7% in cells), and in LTR (8% in T, 4% in NT and 6% in cells) (Table 1A). Among the remaining 2,328 viral integration sites, 289 occurred within exons [242 in T, 47 in NT (P=0.02), and 0 in cells] and 772 within introns (604 in T, 164 in NT, and 4 in cells). Interestingly, 215 breakpoints were found in lncRNA [175 in T, 36 in NT (P=0.03), and 4 in cells]. A total of 1,052 integrations were found in intergenic regions and most of them were located within 100kbp from genes. (Table 1A and 1B). An enrichment of microhomology (MH) sequences between host DNA and integrated HBV DNA was found at the site of integration in a high percentage of chimeras (27%-97%) from each sample (Table 1B). PreS-S genomic sequences and sequences including ENHI/X promoter/3' deleted-X gene were the most frequent HBV integrants detected (Table 1B and Fig.6). In PLC/PRF/5 cells, 11 distinct HBV integration breakpoints were recognized (Fig.7). All HBsAg-negative control patients showed no HBV integration event, confirming the specificity of our HBVsequencing approach.

IV. Discussion

We developed a high-throughput NGS-based viral integration detection method along with a specific bioinformatics pipeline that strongly increased the detection efficiency and characterization of HBV integration events. In fact, this approach enables us to detect a very large number of HBV integrations, averaging at about 847 HBV integration events per individual patient. In concordance with other studies[42,53], we found more (about 4 times) HBV integration events in tumor tissues (86.4%) than in adjacent non-tumor tissues (30.7%). In addition, our high-throughput NGS method revealed a distinct pattern in the preferential sites of integration between tumor and non-tumor tissues. Indeed, HBV insertional sites were significantly enriched in exons and lncRNAs in tumors. The fact that most HBV breakpoints in HCC were found at level of coding genes would suggest that these gene regions are more likely associated with open chromatin where HBV may integrate most efficiently. Moreover, concerning lncRNAs, it is known that HBV integration may modify their production, and that lncRNAs regulate the expression of protein-coding genes[54]. It has been described that lncRNAs are aberrantly expressed in HCC and that they may play a role in modulating malignant phenotypes[55]. According to different studies[51,56,57], which describe the presence of HBV integrants close to repetitive, non-coding human sequences, such as long interspersed nuclear element (LINEs) or short interspersed nuclear elements (SINEs) and long terminal repeats (LTR) we detected recurrent HBV insertion sites within LINEs, SINEs and LTR. In addition, until now, it was not clear whether HBV DNA integration occurs through canonical sequence independent non homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ)[58]. A number of reports underline the role of microhomology (MH) sequences between the cellular and the inserted HBV DNA at level of the integration sites. There is evidence, confirming the importance of these MH sequences, suggesting the potential involvement of MH-mediated mechanism in HBV integration process[53]. In our study we found an enrichment of microhomology sequences between host DNA and integrated HBV DNA at the site of integration, suggesting that MH might have played a major role in HBV integration in the studied cases. Among the different HBV genomic regions, we found that the PreS/S genomic sequences and HBV sequences including ENHI/X promoter/3'-deleted-X gene were the most frequent HBV integrants. Indeed, ENHI is known to be active in the integrated form, and can induce the production of HBX transcripts[59]. Moreover, in HBV integrants, the HBX sequences are frequently truncated[60]. Many experimental studies have described the presence of viral integrants, which can produce proteins with transforming properties, such as HBX and truncated pre-S2/S genomic sequences[60]. These mutated proteins play an important role in the development of HCC: they are able to alter host gene expression

and cellular phenotypes. They may stimulate growth factor-independent proliferation, resistance to growth inhibition, tissue invasion and metastasis, angiogenesis, reprogramming of energy metabolism, and resistance to apoptosis[61].

Many studies reported that several human genes and their regulatory regions are targeted by HBV integration[62]. In this study, we found the presence of an HBV integrant into the regulatory region of the cell division cycle 42 (CDC42) gene, which encodes for a protein involved in regulation of the cell cycle. Another viral integration site was found in the regulatory region of ELL gene, which codes an elongation factor for RNA polymerase II. It has been reported that the E3 ubiquitin ligase activity of ELL is particularly required for its tumour suppressive function. In particular, it can suppress c-Myc-induced proliferation and tumorigenesis. Moreover, it has been demonstred that ELL mutants can also promote metastasis, and this because cells with mutated ELL can acquire invasive capabilities[59]. Among the genomic regulatory regions targeted by HBV, we also found those regulating the expression of *SLC22A7*, *TUBB2B*, *TTYH1*, *TRNP*.

In summary, the developed high-throughput HBV integration sequencing enabled us to perform a large scale and unbiased analysis of HBV DNA integration sites in liver cancer. This new approach allowed the definition of preference sites of integration occurring within regions of the genome prone to DNA mutations or rearrangements and novel genomic elements recurrently affected by HBV integration.

V. Figures and Tables

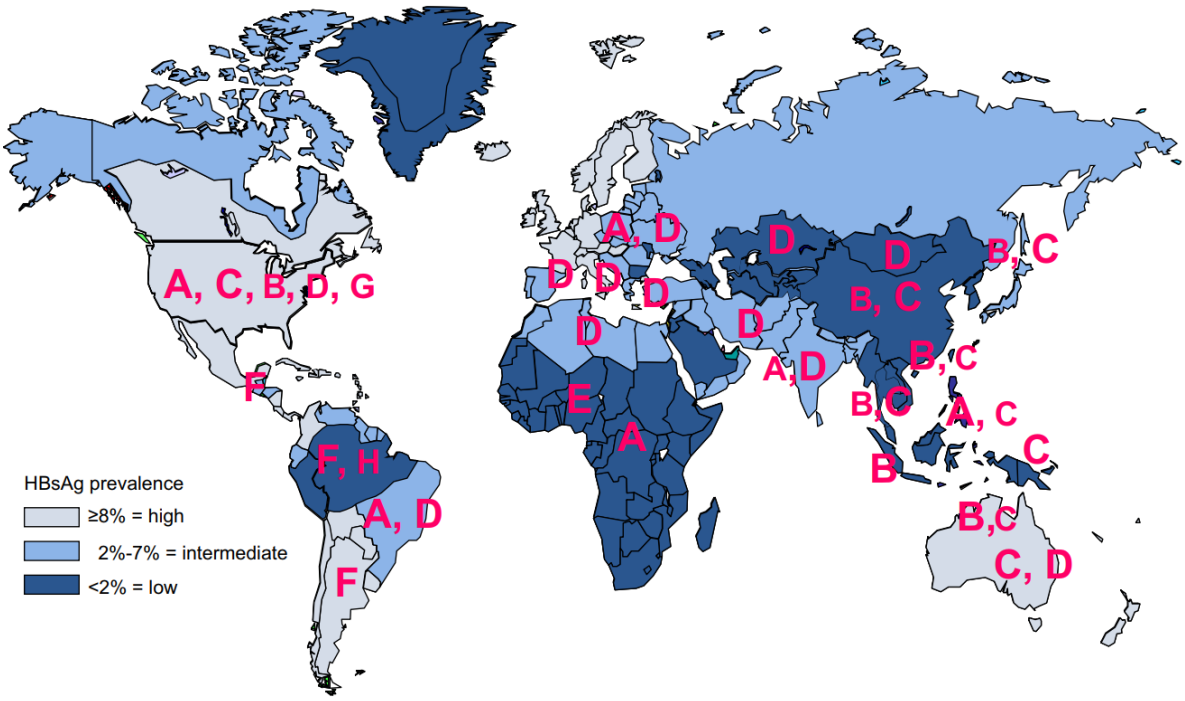

Figure 1 - Geographical distribution of HBV endemicity levels in the world and the areas of predominance of the various HBV genotypes (Stephanos J. Hadziyannis, 2011)[63]

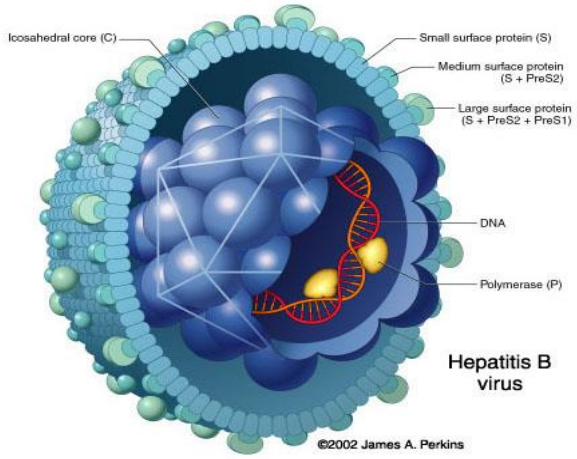

Figure 2 - Hepatitis B virus

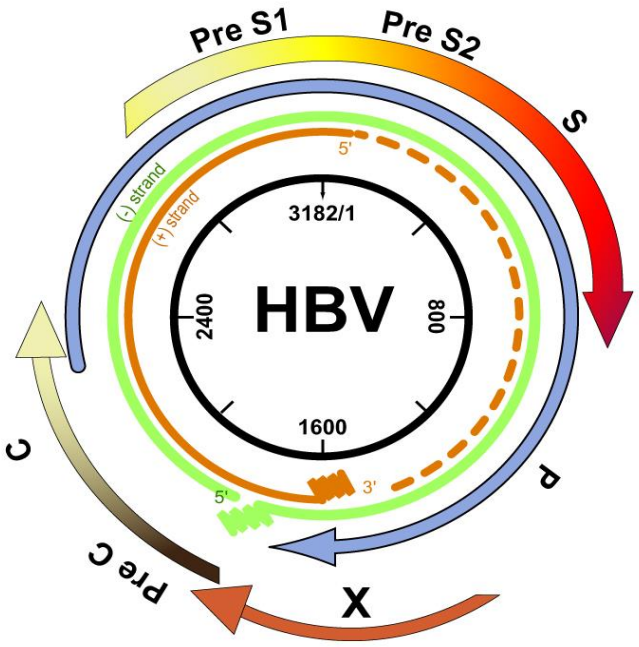

Figure 3 – HBV genome organization (Pollicino at. al, 2014)[12]

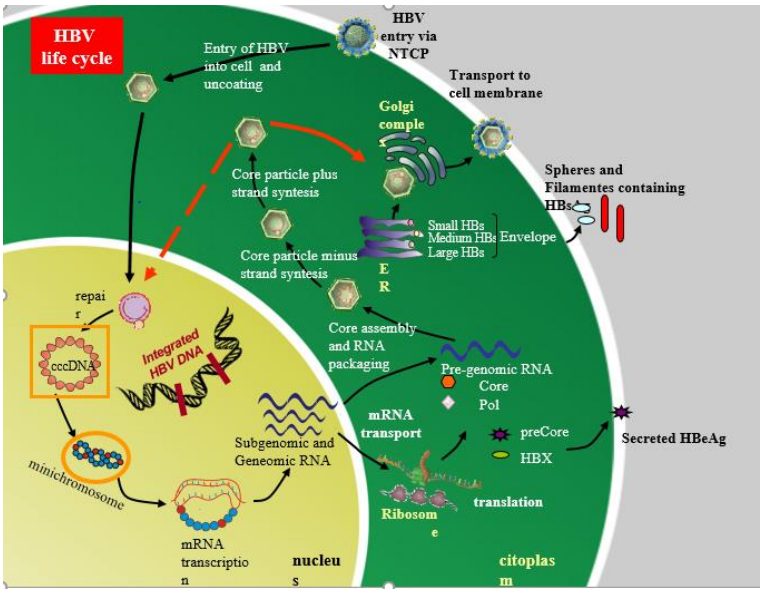

Figure 4 – Schematic representation of HBV virus life cycle and HBV integration into the human hepatocytes (Raimondo et al, 2007)[64]

Figure 5 – Representation of the work flow used for the preparation of HBV samples for the Illumina sequencing

A B

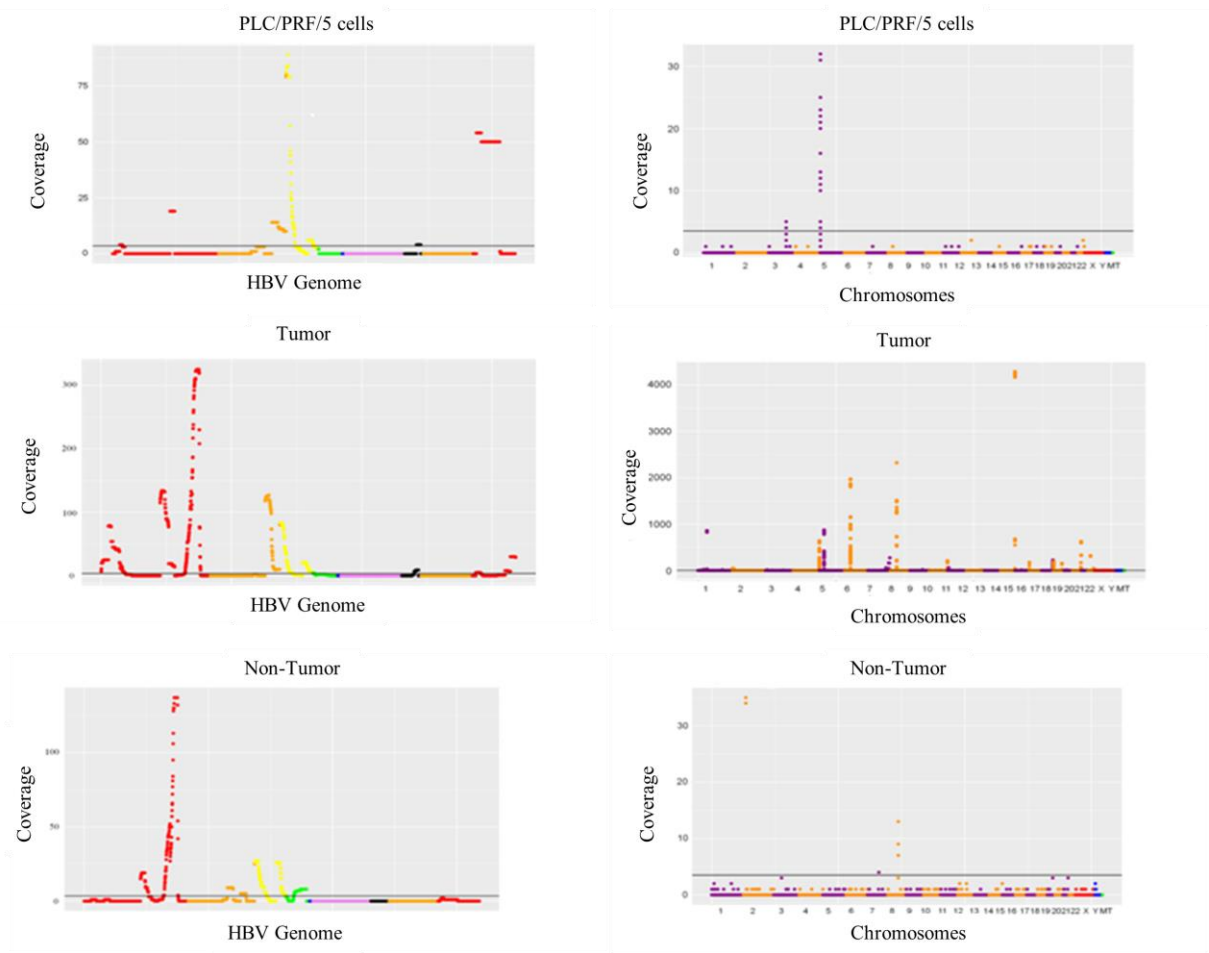

Figure 6 - Coverage of chimeras across HBV DNA (A) and human genome (B), respectively, relative to cells and T and NT liver tissues of 2 patients

- MVK -> preS1/preS2/S
- 3 intergenic -> preS1/preS2/S
- LOC105375660 -> PreCore/Core
- (-8995 bp) LOC105374110 -> ENH1/promoterX

- ENH1/promoterX -> (-1769 bp) hTERT
- PreCore/Core -> UNC5D
- preS1/preS2/S -> CCDC57
- X -> (+5978 bp) CXCL12
- X -> intergenic

Figure 7 - HBV integration breakpoints in PLC/PRF/5 cells

Table 1 A - Summary of HBV targeted host genome elements (numbers and percentage of reads are detailed).

	PLC/PRF/5		TUMOR		NON-TUMOR	
Features	n° reads	%	n° reads	%	n° reads	%
centromere	0	0.000	16	0.047	17	0.163
DNA/hAT	0	0.000	173	0.510	16	0.153
DNA/TcMar	0	0.000	38	0.112	0	0.000
exon	0	0.000	790	2.328	93	0.892
intron	5	7.353	3,362	9.907	658	6.308
intergenic	43	63.235	6,915	20.378	2,882	27.629
LINE	3	4.412	791	2.331	301	2.886
lncRNA	4	5.882	1,475	4.347	1,016	9.740
low complexity	0	0.000	8	0.024	7	0.067
LTR	2	2.941	717	2.113	141	1.352
RC/Helitron	0	0.000	1	0.003	0	0.000
Retroposon/SVA	1	1.471	17	0.050	1	0.010
rRNA	0	0,000	2	0.006	0	0.000
Satellite	2	2.941	181	0.533	98	0.940
Simple repeat	0	0.000	17,260	50.863	4,730	45.346
SINE	8	11.765	2,181	6.427	184	1.764
snRNA	0	0.000	1	0.003	0	0.000
srpRNA	0	0.000	1	0.003	0	0.000

Table 1 B - Integrated HBV sequences in PLC/PRF/5 cells, in tumor and non-tumor tissue specimens

	PLC/PRF/5			TUMOR			NON-TUMOR		
HBV region	coverage	%	MH	coverage	%	MH	coverage	%	MH
S	19	8.962	21	12,777	71.845	19.2	6,285	87.377	15
pol	81	38.208	17	431	2.424	23.4	215	2.989	0.75
ENH1	3	1.415	42	1,124	6.320	16	13	0.181	8.25
promoter X	14	6.604	4	2,235	12.567	16.4	462	6.423	3.75
X	95	44.811	9	291	1.636	26	136	1.891	29
X-ENH2	0	0.000	0	12	0.067	10.2	11	0.153	15.5
X-BCP-Pre-core	0	0.000	0	4	0.022	11.2	4	0.056	39.5
Core	0	0.000	0	4	0.022	5.6	0	0.000	0
Core-pol	4	1.887	0	214	1.203	12.8	4	0.056	8.25
preS1	54	25.472	0	374	2.103	5.8	49	0.681	3.5
preS1-promoter/preS2	0	0.000	0	318	1.788	8.8	14	0.195	16.5

VI. Bibliografy

1. Karayiannis, P. Hepatitis B virus: virology, molecular biology, life cycle and intrahepatic spread. *Hepatol. Int.* **11**, 500–508 (2017).
2. Lavanchy, D. Viral hepatitis: Global goals for vaccination. *J. Clin. Virol.* **55**, 296–302 (2012).
3. Karayiannis, P., Main, J. & Thomas, H. C. Hepatitis vaccines. *Br. Med. Bull.* **70**, 29–49 (2004).
4. Pourkarim, M. R. Molecular identification of hepatitis B virus genotypes/subgenotypes: Revised classification hurdles and updated resolutions. *World J. Gastroenterol.* **20**, 7152 (2014).
5. Kramvis, A. Genotypes and Genetic Variability of Hepatitis B Virus. *Intervirology* **57**, 141–150 (2014).
6. Santos, A. O. *et al.* Characterization of Hepatitis B virus (HBV) genotypes in patients from Rondônia, Brazil. *Virol. J.* **7**, 315 (2010).
7. Schaefer, S. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J. Gastroenterol.* **13**, 14 (2007).
8. Kao, J.-H., Chen, P.-J., Lai, M.-Y. & Chen, D.-S. Acute exacerbations of chronic hepatitis B are rarely associated with superinfection of hepatitis B virus. *Hepatology* **34**, 817–823 (2001).
9. Haga, H. *et al.* Incidence of development of hepatocellular carcinoma in Japanese patients infected with hepatitis B virus is equivalent between genotype B and C in long term. *J. Viral Hepat.* **26**, 866–872 (2019).
10. Seeger, C. & Mason, W. S. Hepatitis B Virus Biology. *Microbiol. Mol. Biol. Rev.* **64**, 51–68 (2000).
11. Wei, Y., Neuveut, C., Tiollais, P. & Buendia, M.-A. Molecular biology of the hepatitis B virus and role of the X gene. *Pathol. Biol.* **58**, 267–272 (2010).
12. Pollicino, T., Cacciola, I., Saffioti, F. & Raimondo, G. Hepatitis B virus PreS/S gene variants: Pathobiology and clinical implications. *J. Hepatol.* **61**, 408–417 (2014).
13. Tong, S. & Revill, P. Overview of hepatitis B viral replication and genetic variability. *J. Hepatol.* **64**, S4–S16 (2016).

14. Schulze, A., Gripon, P. & Urban, S. Hepatitis B virus infection initiates with a large surface protein–dependent binding to heparan sulfate proteoglycans. *Hepatology* **46**, 1759–1768 (2007).

15. Watashi, K., Urban, S., Li, W. & Wakita, T. NTCP and Beyond: Opening the Door to Unveil Hepatitis B Virus Entry. *Int. J. Mol. Sci.* **15**, 2892–2905 (2014).

16. Karayiannis, P. Hepatitis B virus: virology, molecular biology, life cycle and intrahepatic spread. *Hepatol. Int.* **11**, 500–508 (2017).

17. Beck, J. Hepatitis B virus replication. *World J. Gastroenterol.* **13**, 48 (2007).

18. Hu, J. & Liu, K. Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. *Viruses* **9**, 56 (2017).

19. Bruss, V. & Ganem, D. The role of envelope proteins in hepatitis B virus assembly. *Proc. Natl. Acad. Sci.* **88**, 1059–1063 (1991).

20. Pollicino, T., Cacciola, I., Saffioti, F. & Raimondo, G. Hepatitis B virus PreS/S gene variants: Pathobiology and clinical implications. *J. Hepatol.* **61**, 408–417 (2014).

21. Valaydon, Z. S. & Locarnini, S. A. The virological aspects of hepatitis B. *Best Pract. Res. Clin. Gastroenterol.* **31**, 257–264 (2017).

22. Glebe, D. Viral and cellular determinants involved in hepadnaviral entry. *World J. Gastroenterol.* **13**, 22 (2007).

23. Bruss, V., Lu, X., Thomssen, R. & Gerlich, W. H. Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. *EMBO J.* **13**, 2273–2279 (1994).

24. Milich, D. & Liang, T. J. Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection. *Hepatology* **38**, 1075–1086 (2003).

25. Crowther, R. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **77**, 943–950 (1994).

26. Eren, E. *et al.* Structures of Hepatitis B Virus Core- and e-Antigen Immune Complexes Suggest Multi-point Inhibition. *Structure* **26**, 1314-1326.e4 (2018).

27. Bartenschlager, R. & Schaller, H. The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. *EMBO J.* **7**, 4185–4192 (1988).

28. Jones, S. A., Clark, D. N., Cao, F., Tavis, J. E. & Hu, J. Comparative Analysis of Hepatitis B Virus Polymerase Sequences Required for Viral RNA Binding, RNA Packaging, and Protein Priming. *J. Virol.* **88**, 1564–1572 (2014).

29. Jones, S. A., Boregowda, R., Spratt, T. E. & Hu, J. In Vitro Epsilon RNA-Dependent Protein Priming Activity of Human Hepatitis B Virus Polymerase. *J. Virol.* **86**, 5134–5150 (2012).

30. Jones, S. A. & Hu, J. Hepatitis B virus reverse transcriptase: diverse functions as classical and emerging targets for antiviral intervention. *Emerg. Microbes Infect.* **2**, 1–11 (2013).

31. Decorsière, A. *et al.* Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature* **531**, 386–389 (2016).

32. Tu, H. *et al.* Biological Impact of Natural COOH-Terminal Deletions of Hepatitis B Virus X Protein in Hepatocellular Carcinoma Tissues. 9

33. Slagle, B. L. & Bouchard, M. J. Hepatitis B Virus X and Regulation of Viral Gene Expression. *Cold Spring Harb. Perspect. Med.* **6**, a021402 (2016).

34. Hai, H. Role of hepatitis B virus DNA integration in human hepatocarcinogenesis. *World J. Gastroenterol.* **20**, 6236 (2014).

35. El-Serag, H. B. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology* **142**, 1264-1273.e1 (2012).

36. Buendia, M.-A. & Neuveut, C. Hepatocellular Carcinoma. *Cold Spring Harb. Perspect. Med.* **5**, a021444–a021444 (2015).

37. Chen, C.-J. *et al.* Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA* **295**, 65–73 (2006).

38. Levrero, M. & Zucman-Rossi, J. Mechanisms of HBV-induced hepatocellular carcinoma. *J. Hepatol.* **64**, S84–S101 (2016).

39. Bonilla Guerrero, R. & Roberts, L. R. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. *J. Hepatol.* **42**, 760–777 (2005).

40. An, P., Xu, J., Yu, Y. & Winkler, C. A. Host and Viral Genetic Variation in HBV-Related Hepatocellular Carcinoma. *Front. Genet.* **9**, (2018).

41. Horikawa, I. & Barrett, J. C. Transcriptional regulation of the telomerase hTERT gene as a target for cellular and viral oncogenic mechanisms. *Carcinogenesis* **24**, 1167–1176 (2003).

42. Sung, W.-K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).

43. Liang, H.-W. *et al.* Hepatitis B virus-human chimeric transcript HBx-LINE1 promotes hepatic injury via sequestering cellular microRNA-122. *J. Hepatol.* **64**, 278–291 (2016).

44. Lau, C.-C. *et al.* Viral-Human Chimeric Transcript Predisposes Risk to Liver Cancer Development and Progression. *Cancer Cell* **25**, 335–349 (2014).

45. Wang, Y. *et al.* Characterization of HBV integrants in 14 hepatocellular carcinomas: association of truncated X gene and hepatocellular carcinogenesis. *Oncogene* **23**, 142–148 (2004).

46. Caselmann, W. H. *et al.* A trans-activator function is generated by integration of hepatitis B virus preS/S sequences in human hepatocellular carcinoma DNA. *Proc. Natl. Acad. Sci.* **87**, 2970–2974 (1990).

47. Wirth, S. & Zabel, B. Molecular hybridization techniques in current diagnosis of chronic hepatitis B in childhood. *Eur. J. Pediatr.* **151**, 243–249 (1992).

48. Kitada, K., Aikawa, S. & Aida, S. Alu-Alu Fusion Sequences Identified at Junction Sites of Copy Number Amplified Regions in Cancer Cell Lines. *Cytogenet. Genome Res.* **139**, 1–8 (2013).

49. Budzinska, M. A., Shackel, N. A., Urban, S. & Tu, T. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg. Microbes Infect.* **7**, 1–12 (2018).

50. Tu, T. & Jilbert, A. R. Detection of Hepatocyte Clones Containing Integrated Hepatitis B Virus DNA Using Inverse Nested PCR. in *Hepatitis B Virus* (eds. Guo, H. & Cuconati, A.) **1540**, 97–118 (Springer New York, 2017).

51. Kan, Z. *et al.* Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* **23**, 1422–1433 (2013).

52. The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378–384 (2017).

53. Zhao, L.-H. *et al.* Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat. Commun.* **7**, (2016).

54. Yang, X. *et al.* Distinct hepatitis B virus integration patterns in hepatocellular carcinoma and adjacent normal liver tissue. *Int. J. Cancer* **140**, 1324–1330 (2017).

55. Ghidini, M. & Braconi, C. Non-Coding RNAs in Primary Liver Cancer. *Front. Med.* **2**, (2015).

56. Murakami, Y. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. *Gut* **54**, 1162–1168 (2005).

57. Ding, D. *et al.* Recurrent Targeted Genes of Hepatitis B Virus in the Liver Cancer Genomes Identified by a Next-Generation Sequencing–Based Approach. *PLoS Genet.* **8**, e1003065 (2012).

58. Mladenov, E., Magin, S., Soni, A. & Iliakis, G. DNA double-strand-break repair in higher eukaryotes and its role in genomic instability and cancer: Cell cycle and proliferation-dependent regulation. *Semin. Cancer Biol.* **37–38**, 51–64 (2016).

59. Shamay, M., Agami, R. & Shaul, Y. HBV integrants of hepatocellular carcinoma cell lines contain an active enhancer. *Oncogene* **20**, 6811–6819 (2001).

60. Schlüter, V., Meyer, M., Hofschneider, P. H., Koshy, R. & Caselmann, W. H. Integrated hepatitis B virus X and 3' truncated preS/S sequences derived from human hepatomas encode functionally active transactivators. *Oncogene* **9**, 3335–3344 (1994).

61. Pollicino, T. Occult hepatitis B virus and hepatocellular carcinoma. *World J. Gastroenterol.* **20**, 5951 (2014).

62. Paterlini-Bréchot, P. *et al.* Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. *Oncogene* **22**, 3911–3916 (2003).

63. Hadziyannis, S. J. Natural history of chronic hepatitis B in Euro-Mediterranean and African Countries. *J. Hepatol.* **55**, 183–191 (2011).

64. Raimondo, G., Pollicino, T., Cacciola, I. & Squadrito, G. Occult hepatitis B virus infection. *J. Hepatol.* **46**, 160–170 (2007).