Oncogene (2001) 20, 6233–6240

www.nature.com/onc

SHORT REPORTS

Identification of human cancer-related genes by naturally occurring Hepatitis B Virus DNA tagging

Devrim Gozuacik[1,5], Yoshiki Murakami[1,2,5], Kenichi Saigo[1,3,5], Mounia Chami[1,5], Claude Mugnier[4], David Lagorce[1], Takeshi Okanoue[2], Tetsuro Urashima[3], Christian Bréchot[1] and Patrizia Paterlini-Bréchot*[,1]

[1]U370 INSERM, Necker Institute, 75015, Paris, France; [2]Third Department of Internal Medicine, Kyoto Prefectural University of Medicine, Kyoto 602-0841, Japan; [3]Second Department of Surgery, Chiba University, Chiba 263-8852, Japan; [4]Department of Biotechniques and Medical Informatics, Necker-Enfants Malades Hospital, 75015 Paris, France

Proviral tagging has been used in animals as a powerful tool for cancer genetics. We show that a similar approach is possible in patients with hepatocellular carcinoma (HCC) infected by Hepatitis B Virus (HBV), a human pararetrovirus which may act by insertional mutagenesis. In this work, the HBV genome is used as a probe to identify cancer-related genes. By using HBV-Alu-PCR, we obtained 21 HBV/cellular DNA junctions from 18 different patients. In six of 21, we found the HBV DNA integrated into a cellular gene: (1) Sarco/Endoplasmic Reticulum Calcium ATPase1 Gene; (2) Thyroid Hormone Receptor Associated Protein 150 alpha Gene; (3) Human Telomerase Reverse Transcriptase Gene; (4) Minichromosome Maintenance Protein (MCM)-Related Gene; (5) FR7, a new gene expressed in human liver and cancer tissues; and (6) Nuclear Matrix Protein p84 Gene. Seven junctions contained unique cellular sequences. In the remaining eight, the HBV DNA was next to repetitive sequences, five of them of *LINE1* type. The cellular genes targeted by HBV are key regulators of cell proliferation and viability. Our results show that studies on HBV-related HCCs allow to identify cellular genes involved in cancer. We therefore propose this approach as a valuable tool for functional cancer genomic studies in humans. *Oncogene* (2001) **20,** 6233–6240.

Keywords: hepatitis B virus; Alu-PCR; gene trap; cancer; insertional mutagenesis; hepatocellular carcinoma

The use of animal retroviruses as proviral tags has been highly successful in revealing genes and pathways involved in murine leukemia and mammary carcinoma (Cecconi and Meyer, 2000; Evans *et al*., 1997; Jonkers and Berns, 1996). Retroviral insertional mutagenesis induces a high incidence of neoplasia in animals and studies based on amplification and sequencing of proviral integrations have led recently to isolate a large number of candidate leukemia genes (Largaespada, 2000; Li *et al*., 1999). Human oncogenic viruses acting by insertional mutagenesis like animal retroviruses have never been described. However, Hepadnaviruses, which are small, circular, hepatotropic DNA viruses, have been classified as Pararetroviruses since they share partial genomic sequence, replication strategy and genome organization similarities with slow oncogenic retroviruses (Robinson *et al*., 1987). Hepadnaviruses have strict host-specificity and similar genomic organization (Fourel and Tiollais, 1994). They include HBV, infecting humans, GSHV (Ground Squirrel Hepatitis Virus), WHV (Woodchuck Hepatitis Virus), DHV (Duck Hepatitis Virus) and HHBV (Heron Hepatitis Virus). Interestingly, WHV induces liver cancer in 100% of animals infected at birth. Molecular studies have shown that, in 80% of cases or more, it acts by insertional mutagenesis, like a slow oncogenic retrovirus, and cis-activates oncogenes of the Myc family (c-Myc, N-Myc and N-Myc2) (Buendia, 1992).

In humans, epidemiological studies have demonstrated the association of HBV chronic infection and development of hepatocellular carcinoma (HCC) (Johnson, 1994). HCC develops after a long delay, generally of 15 to 30 years, from HBV infection. Long lasting viral multiplication and HBV protein expression are known to stimulate the host immune response and thus liver inflammation and fibrosis (Chisari, 1995). This process modifies cell-to-cell and cell-to-extracellular matrix interactions and induces the synthesis of cytokines which may trigger liver cell proliferation. Therefore, chronic liver disease with persistence of the HBV genome is a major risk factor for HCC. Besides this mechanism, a direct role of HBV in liver carcinogenesis has been suggested since HBV-related HCC may also develop without a background of chronic liver inflammation (Brechot *et al*., 2000). The HBV X protein has a transactivating effect on both cellular and viral genes and its expression has been shown to interfere with the control of cell growth and apoptosis. *In vitro* studies and *in vivo* experiments in

*Correspondence: P Paterlini-Bréchot; E-mail: paterlini@necker.fr
[5]Authors contributing equally to this work
Received 15 January 2001; revised 5 July 2001; accepted 16 July 2001

transgenic mice have suggested its possible role in liver carcinogenesis. However, no studies have demonstrated up to now the transforming or tumorigenic effect of X gene sequences directly isolated from human tumorous liver tissues (Brechot *et al*., 2000). HBV DNA integration into cellular DNA is an early event during chronic HBV infection and the majority of HCC tissues exhibits clonal expansion of tumor cells carrying the same integration site. HBV genome integration and reintegration events allow the persistence of viral sequences and may lead to chromosomal rearrangements, deletion of both viral and cellular DNA and synthesis of truncated, transactivating, $PreS_2/S$ and viral proteins (Brechot *et al*., 2000). In addition, HBV DNA integration may act by insertional mutagenesis. In fact, the HBV genome has been found integrated into the *Retinoic Acid Receptor Beta* (Dejean *et al*., 1986) and the *Cyclin A2* (Wang *et al*., 1990) gene in two isolated HCCs. Interestingly, further analyses have demonstrated that the mutation of these two genes identified in the tumorous tissue has transforming and tumorigenic properties (Berasain *et al*., 1998; Garcia *et al*., 1993). Thus, studies on HBV-related insertional mutagenesis allowed to isolate two previously unknown human genes, which play a major role in cell differentiation and proliferation, and to prove their role in carcinogenesis.

Over the past 20 years, conventional cloning strategies have been applied to HCCs to identify cellular genes at the HBV integration site. A limited number of tumors has been explored by these long and cumbersome methods and only very few genes have been identified (Brechot *et al*., 2000). Chromosomal studies have also shown that, at variance with slow oncogenic retroviruses, HBV does not seem to have common integration sites in the cellular DNA (Matsubara and Tokino, 1990). The general conclusion of these analyses was that HBV-related insertional mutagenesis is probably an anectodical event in HCC.

Despite this discouraging background, we decided to further explore the frequence of HBV related insertional mutagenesis and test the hypothesis that the natural tagging of cellular genes by HBV could be used to identify cancer-related genes.

We have developed a PCR-based assay which enables the rapid screening of HBV/cellular DNA junctions using HBV-X and Alu-specific primers (Alu-PCR) (Minami *et al*., 1995) (Figure 1). The HBV X gene is frequently interrupted by the viral DNA integration (Schluter *et al*., 1994). Circular HBV DNAs coexist in liver tissues with integrated viral genomes and the cellular sequence targeted by the virus is completely unknown. In order to overcome these limitations and avoid undesirable amplifications between Alu sequences, primers are synthetized with dUTP instead of dTTP and destroyed after 10 initial cycles of PCR. Only specifically targeted sequences are then amplified with primers to the HBV X region and to a tag sequence introduced in the Alu-specific primer. By using this approach, we have now isolated 21 HBV integration sites from 18 HCCs. Eight junctions exhibit repetitive sequences, five of them of *LINE1* type. Seven HBV/cellular DNA junctions contain unique cellular sequences which, at present, do not correspond to any coding sequence in data banks.

The HBV genome was found to be integrated into different cellular genes in six independent tumors (Table 1 and Figure 2). Remarkably, five of them belong to gene families involved in cell proliferation and/or viability. The fifth one is expressed in fetal and cancer tissues. Southern blot analyses performed in three tumors (77T, 83T, 86T) and repeated analyses on different DNA samples from the same tumor showed that the HBV-DNA integration in the target gene took place in about all the tumor cells (data not shown).

In tumor 54T, HBV DNA integration occurred into an intronic sequence (position 222267 to 222742 in the supercontig NT011005) of the Nuclear Matrix Protein p84 (NMP p84, code gene: XM 008756). The HBV and the NMP p84 ORFs were found in the same orientation. The Nuclear Matrix Protein p84 gene encodes a nuclear protein which has been shown to bind $p110^{RB}$ in its N-terminal region (Durfee *et al*., 1994). This binding would contribute to, and concentrate the $p110^{RB}$ protein to subnuclear regions where active forms of several target proteins (transcription factors) would be located. NMP p84 is therefore expected to participate in the control of the cell cycle.

In tumor 86T, HBV DNA integration occurred in frame in the third exon of the *Sarco/Endoplasmic Reticulum Calcium ATPase-1 (SERCA1)* gene (SERCA1 genomic sequence accession number U96773, nt 1979–2293). SERCA proteins play a pivotal role in regulating cellular calcium which, in turn, acts as an intracellular messenger involved in a broad range of cellular activities, including cell proliferation and death (Berridge *et al*., 1998). In the tumor, HBV integration cis-activated HBV-X/SERCA1 fusion transcripts and their *in vitro* expression has been shown to induce ER calcium depletion and apoptosis (Chami *et al*., 2000). Although the pleiotropic role of cellular calcium, including on cell growth and apoptosis, has been repeatedly shown in the past, this is the first report implicating the clonal mutation of a *SERCA* gene in a tumor cell proliferation. This study also led to discover new truncated SERCA1 proteins involved in the control of apoptosis (Chami *et al*., 2001).

In tumor 100T, HBV DNA integration was found next to a 258 bp cellular sequence containing a 115 bp stretch identical to the *Thyroid Hormone Receptor Associated Protein-150 alpha (TRAP-150 alpha)* gene (TRAP-150 genomic sequence accession number AL360074, nt 103283–103541 and TRAP150 cDNA sequence accession number AF117756, nt 2119–2233). The HBV and TRAP-150 alpha ORF were found in the same orientation. TRAP proteins are coactivators of nuclear receptors, including thyroid receptor, required for the formation of functional transcriptional preinitiation complexes on target genes (Ito *et al*., 1999). Thyroid hormone is a major regulator of liver cell proliferation and mutant thyroid hormone nuclear receptors have been described in HCC cells

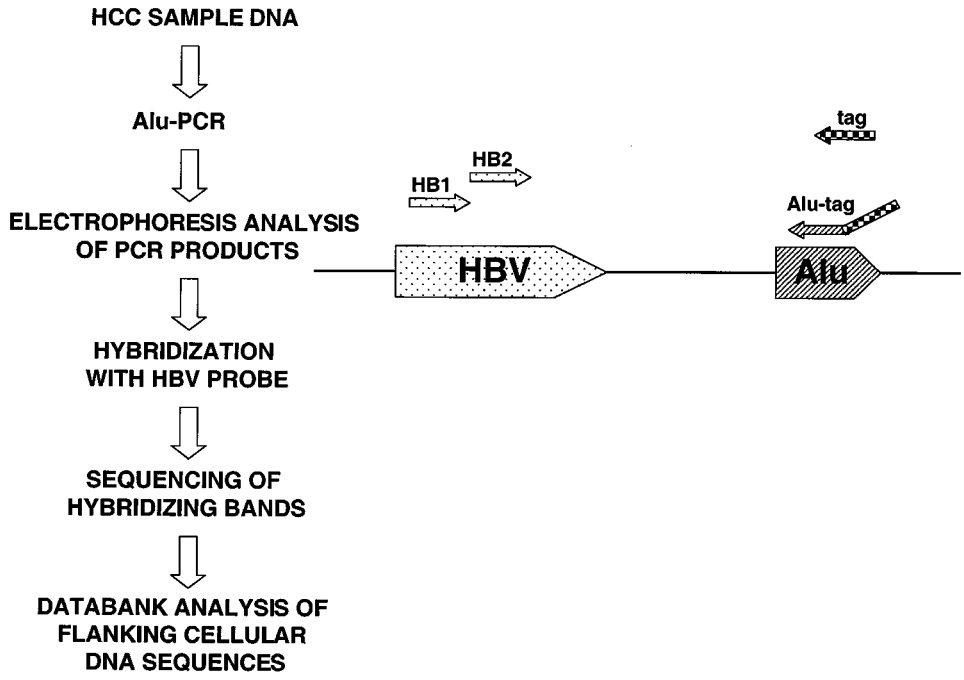

Figure 1 Schematic representation of the technical protocol applied to isolate HBV/cellular DNA junctions (Minami *et al*., 1995)

Table 1 Histological, serological and sequencing data of HCCs

HCC code	*Non-tumor liver histology*	*HBsAg*	*Anti-HBs*	*Anti-HBc*	*Cellular flanking sequence* (bp)*	*Chromosomal localization (type of sequence)*
54T	minimal liver charge	+	–	+	475	18p11.3 (NMP p84 gene)
63T	minimal liver change	+	–	+	606	9p11 (repetitive)
77T	chronic hepatitis	+	–	+	485	20p12.3 (hMCM8 gene)
83T	liver cirrhosis	+	–	+	673	5p15.33 (hTERT gene)
83T					239	2p24.1–24.3 (unique)
86T	chronic hepatitis	–	+	+	314	16p12.1 (SERCA1 gene)
95T	liver cirrhosis	+	–	+	355	9q22.1 (unique)
100T	minimal liver change	+	–	+	258	1p32.3 (TRAP150 gene)
101T	chronic hepatitis	+	–	+	502	N.I.
FR2	liver cirrhosis	+	–	+	1220	14q21.1–24.1 (unique)
FR3	liver cirrhosis	+	–	+	1349	9q21.1 (unique)
FR7	liver cirrhosis	+	–	+	660	12p (FR7 gene)
SA1	chronic hepatitis	+	–	+	366	3q11.2 (unique)
SA2	liver cirrhosis	+	–	+	511	3p26 (unique)
SA2					591	N.I.
SA5	liver cirrhosis	+	+	+	586	N.I.
GR1	liver cirrhosis	–	+	–	282	2p23 (repetitive)
GR2	liver cirrhosis	+	–	+	271	2q11.2 (unique)
GR2					780	N.I.
GR3	liver cirrhosis	+	–	+	886	N.I.
GR10	liver cirrhosis	+	–	+	1880	N.I.

HBsAg, Hepatitis B surface antigen; anti-HBs, antibody against the hepatitis B surface antigen; anti-HBc, antibody against the hepatitis B core antigen. N.I. = Not identified. *Size of the isolated cellular sequence flanking the HBV sequence

(Lin *et al*., 1999). However, the precise biological effect of the human TRAP-150 alpha gene is still unknown (RG Roeder, personal communication). It could participate in gene- or activator-selective cofactor complexes. Our preliminary data indicate that TRAP-150 alpha is involved in the control of apoptosis (Y Murakami *et al*., in preparation). The present finding of its mutation in a human tumorous tissue strongly suggests its implication in liver carcinogenesis.

In tumor 77T, HBV DNA was found next to a 485 bp cellular sequence identical to part of the *Minichromosome Maintenance Protein (MCM) 2/3/5* gene-like sequence on chromosome 20p12.3-13 (*MCM2/3/5* gene-like genomic sequence accession number AL035461, nt 51574–52059). The HBV and MCM8 ORF were found in the opposite orientation. The six members of the MCM protein family (MCM 2–7) form a complex with DNA helicase activity which controls DNA replication (Tye, 1999). MCM

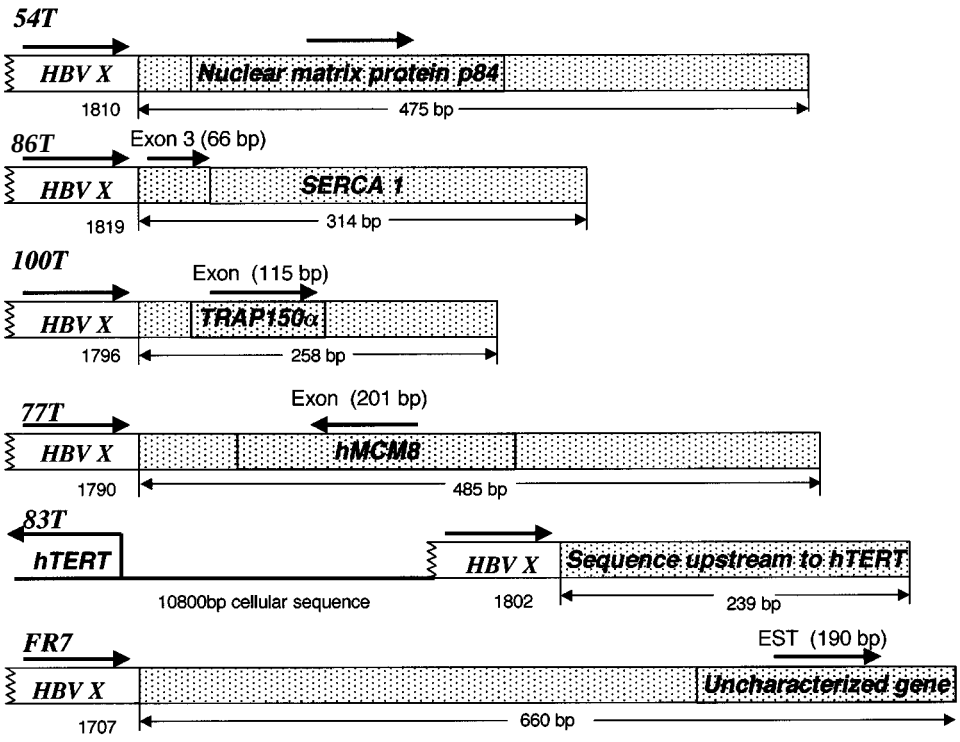

Figure 2 HBV DNA integration in six cellular genes. 54T, 86T, 100T, 77T, 83T and FR7 are HCC codes (Table 1). Open box: HBV X sequence. Number below indicates HBV nucleotide at the HBV/cellular DNA junction (HBV subtype and accession number V00866 was taken as reference sequence.). The double headed arrow marks the size of the cellular sequence obtained by Alu-PCR. Dotted box: cellular region of identity with GenBank sequences. Shaded box: coding sequences. Bold arrowhead: orientation of the ORF

proteins have been reported to be highly expressed in proliferating and neoplastic tissues (Freeman *et al*., 1999; Todorov *et al*., 1998). Our finding allowed us to clone the complete cDNA sequence of this new human MCM gene, that we named human MCM8 (hMCM8) (D Gozuacik *et al*., submitted). In order to study the expression of the new MCM gene in the tumor and adjacent liver, we have generated a polyclonal antibody raised to a N-terminal peptide sequence specific to hMCM8. The Western blot analysis showed that a smaller hMCM8 protein (33 kDa) is selectively expressed in the tumorous (77T) as compared to the non tumorous tissue (Figure 3c). The size of this protein is consistent with a C-terminally truncated hMCM8 protein derived from a premature stop codon introduced by the HBV sequence (in the 'reverse orientation') (predicted size: 32.8 kDa). Work is in progress to clone the cDNA encoding this short hMCM8 protein.

In tumor 83T, HBV DNA integration was found next to a 239 bp cellular sequence identical to a sequence located upstream to the *Human Telomerase Reverse Transcriptase (hTERT)* gene promoter, (hTERT genomic sequence accession number AF128893, nt 240–478). The HBV and hTERT ORF were found in opposite orientations. The implication of hTERT gene in cancer is already well established in the literature. In fact, telomerase is involved in the process of cell immortalization and is expressed in the majority of human malignancies (Hanahan and Weinberg, 2000; Meyerson, 2000). This case is relevant since it further supports the hypothesis that HBV integration targets genes involved in the tumorigenic process. The Western blot analysis of the tumor 83T with the anti-hTERT antibody showed a hyperexpressed band (170 kDa) in addition to the hTERT regular hyperexpressed band (136 kDa) in the tumor as compared to the non tumorous tissue (Figure 3d). It is presently difficult to explain the larger size of the additional TERT protein in the tumor (the HBV integration took place 10.8 kb upstream the hTERT ATG); one may hypothesize that the HBV enhancer could have activated the expression of a normally underexpressed hTERT variant, as previously shown for the SERCA1 gene (Chami *et al*., 2001).

Moreover, it has been recently shown that, in the HuH4 human hepatocellular carcinoma-derived cell line (GenBank, AF325900), the HBV genome integration takes place into the promoter region of hTERT, 313 bp before the ATG (Horikawa and Barrett, 2001) and cis-activates the hTERT gene expression. Together with our results, this is the first demonstration of a gene targeted by the HBV integration in two independent HCCs.

In tumor FR7, HBV DNA was found next to a 660 bp cellular sequence identical to a genomic clone on chromosome 12p (accession number AC009318, nt 112312–112971). A 190 bp stretch of this sequence is identical to human fetal (accession numbers N67205, H16791) and cancer (neuroblastoma and breast cancer)

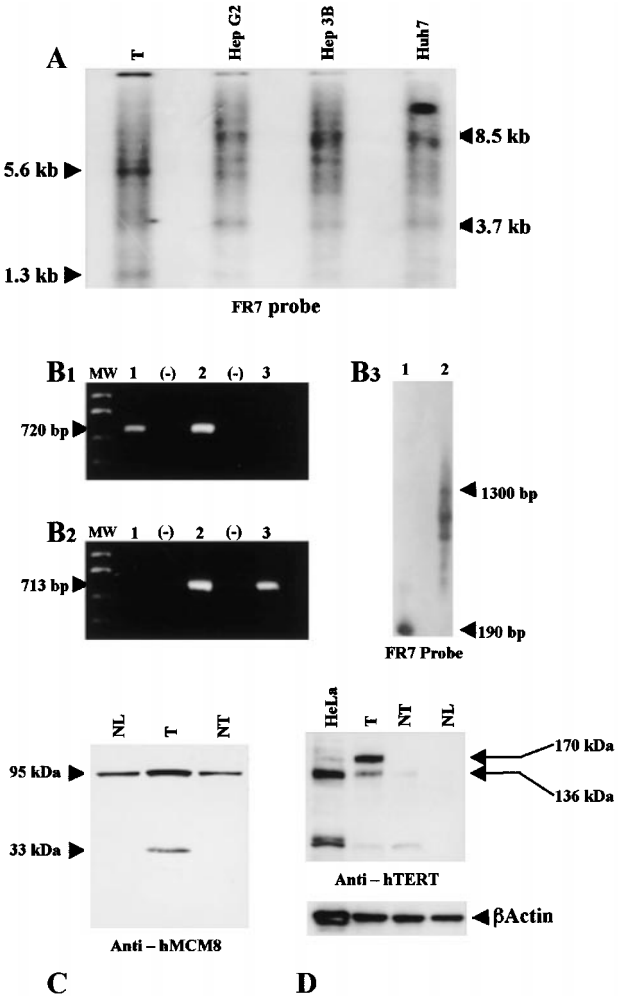

Figure 3 Mutagenic effect of HBV DNA integration into cellular genes. (**A**) Northern blot analysis of poly $(A)^+$ RNA from tumorous (T) tissue FR7, HepG2, Hep3b and Huh7 HCC-derived cell lines. Two bands of 1.3 and 5.6 Kb are clearly visible in the tumor, while transcripts of 3.7 and 8.5 Kb are expressed in HCC-derived cell lines. Northern blot analysis was performed as previously described (Chami *et al*., 2000). FR7 probe was synthetized by amplifying normal liver RNA with primers KP1 (5′-CGTGGTAGTTTAAGTGTACA-3′) and KP2 (5′-GATCAC-CAGGTGCTTAGTTTC-3′) drawn on the FR7 unique sequence. (**B1**, **B2** and **B3**) Amplification of X/FR7 chimeric transcript. (**B1**) Amplification with MX2 primer (on the HB X sequence: 5′-TGCCCAAGGTCTTACATAAGAGGA-3′) and KP2 primer (on the unique FR7 sequence) of cDNA made from FR7 tumor poly $(A)^+$ RNA by using the Marathon kit (Clontech, Palo Alto, CA, USA) protocol according to manufacturer's instructions (lane 1), FR7 tumor DNA (lane 2), and normal liver DNA (lane 3). (−) indicates negative controls. MW: molecular weight marker. The picture shows the amplified chimeric X/FR7 transcript (lane 1). The amplified sequence does not contain intronic sequences, since a band of the same size is also obtained in the tumor DNA. In contrast, no chimeric sequence is detectable by amplifying normal liver DNA. (**B2**). The same samples as in **B1** are amplified with primers on exon 10 and exon 12 of the SERCA1 gene as described (Chami *et al*., 2001). The expected size of the amplified genomic sequences is 713 bp. The picture shows that there is no residual DNA in the cDNA obtained from FR7 tumorous tissue (lane 1). (**B3**) Amplification of the FR7 tumor cDNA with the KP1 and KP2 primers (lane 1), and with a primer on the X ATG (5′-TTTCCATGGCTGCTAGGCTGACTGC-3′) and a primer on the AP1 sequence (5′-CCATCCTAATACGACTCACTA-TAGGGC-3′, present in the adaptor linkers blunt ligated to Marathon cDNA) followed by transfer and Southern blot hybridization with the FR7 probe. The picture shows that the probe is specific for the FR7 unique sequence (expected size: 190 bp, lane 1) and that the chimeric transcript amplified has the size of 1300 bp (lane 2). With the Marathon protocol only the higher band (1300 bp) has to be taken into account, the lower bands being due to incomplete amplifications. (**C**) Western Blot analysis (Musahl *et al*., 1998) of tumorous (T) and non tumorous (NT) tissue from patient 77 (see Table 1) and normal liver (NL) by using the anti-hMCM8 antibody (D Gozuacik *et al*., submitted). In addition to the expected hMCM8 size (95 kDa), a shorter band is clearly visible in the tumor (33 kDa). (**D**) Western blot analysis (Liu *et al*., 2001) of tumorous (T) and non tumorous tissue (NT) from patient 83 (see Table 1), normal liver (NL) and HeLa cells (positive control for hTERT expression) using anti-hTERT goat polyclonal antibody (Santa Cruz Biotechnology). The hTERT expression (136 kDa) is increased in the tumor tissue as compared to the non-tumorous tissue (see ß actin normalization using anti-actin goat polyclonal antibody (Santa Cruz Biotechnology)). In addition, a higher band (170 kDa) is present in the tumor lane, indicating the hyper-expression of an unusual hTERT protein

derived ESTs (Cancer Genome Anatomy Project, accession numbers AI361463, AA996057). The Gen-scan analysis showed that the FR7 gene is an 'unnamed gene' (supercontig NT_009622, position 2129217, AK001927, EST BG741742 and BF572281) without any homology with known genes or domains. This analysis also predicted a FR7 mRNA of 3847 bp. According to this study, the HBV integration took place in the 3′ part of the gene corresponding to the EST BG741742. The HBV and FR7 ORF were found in the same orientation. This is an example of a new gene, expressed in fetal and tumor tissues, identified through the HBV-DNA integration analysis. The low amount of tumorous tissue allowed us to perform only one Northern blot analysis using poly$(A)^+$ RNA. This test showed abnormal bands in the tumorous FR7 tissue as compared to the FR7 sizes identified in tumor cell lines (Figure 3A). The hybridization of the membrane with the HBV probe gave a very weak signal in the tumorous tissue consistent with one of the two bands found with the FR7 probe (data not shown). In order to confirm the detection of a chimeric transcript, we synthetized a cDNA from poly$(A)^+$ RNA using the Marathon kit protocol, which includes ligation to the cDNA of linkers carrying the AP1 sequence. By using this approach we confirmed the evidence for a chimeric transcript. In fact we amplified the tumor derived cDNA with a primer on the HBV X sequence (MX2) and a primer on the cellular FR7 sequence (KP2). In repeated experiments we obtained a band of 720 bp (Figure 3, B1, lane 1). This band was consistent with a 1300 bp band obtained on the FR7 cDNA by using a primer on the X ATG (X ATG) and a primer on the cDNA adaptor sequence (AP1) (Figure 3, B3) and corresponding to the size of the shorter transcript (1.3 Kb) detected by Northern blot in the tumor (Figure 3A). Since the primers MX2 and KP2 amplified a 720 bp band also from the tumor DNA (Figure 3, B1, lane 2), we finally tested the Marathon cDNA for the presence of genomic DNA with SERCA

```
GR1   TGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGAACAGGTTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGCAA
      1719                                                                                          1818
GR10  AGACTGGGAGGAGTTGGGGGAGGAGAACAGGTTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGCAACTTTTTC
      1726                                                                                          1825
54T   AGACTGTGTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAATGATCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCA
      1711                                                                                          1810
GR3   GTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGACTAGGTTAATGATCTTTGTATTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATG
      1716                                                                                          1815
GR2   TTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGNTTAATGGTCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGTAACCT
      1722                                                                                          1821
GR2s  AGACTGTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGATTAATGATCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCA
      1711                                                                                          1810
63T   GAGTGGGAGGAGCTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGTTCCCCACCACCAGGCATCTTTTTCA
      1728                                                                                          1827
SA5   GAGGCTTACTTCAAAGACTGTGTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTATTAGGAGGCTGTAGGCATAAATTGG
      1697                                                                                          1796
FR3   GACTGGGAGGAGCTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTATTAGGAGGCTGTAGGCATAAAGTGGTCTGCGCACCAGCACCATGCAACTTTTTCA
      1728                                                                                          1827
86T   TGTTTAAGGACTGGGAGGAGTTGGGGGAGGAGACTAGGTTAATGATCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCACCATGCAA
      1720                                                                                          1819
FR2   CGACCGACCTTGAGGCATACTTCAAAGACTGTGTATTTAAAGACTGGGAGGAGTTGGGGGAGGAGACTAGGTTAATGATCTTTGTACTGGGAGGCTGGAG
      1686                                                                                          1785
101T  TAGACTGGGAGGAATTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCACCATGCAACTTTTT
      1725                                                                                          1824
95T   AGACTGTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAATGATCTTTGTACTAGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCA
      1711                                                                                          1810
83T   GTCACGACCGACCTTGAGGCGTACTTCAAAGACTGTTTGTTTAAGGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAATGATCTTTGTACTAGGAGGCT
      1682                                                                                          1781
83T   TCAGCAATGTCAACGACCGACCTTGAGGCGTACTTCAAAGACTGTTTGTTTAAGGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTAC
      1673                                                                                          1772
SA1   CAGATCCTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCCCAGCAATGTCAACGACCGACC
      1640                                                  1693
SA2s  ATCAGATCCTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCCCAGCAATGTCAACGACCGACCTTGAGGCATACTT
      1630                                                                    1708
FR7   CCCACAAGTCTTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCTCTGCAATGTCAACGACCGACCTTGAGGCAAACT
      1628                                                                     1707
100T  TGAGGCATACTTCAAAGACTGTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGAGTAGGCTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAATTG
      1697                                                                                          1796
SA2   CTCTTGGACTCCCAGCAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTGTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAAT
      1661                                                                                          1760
77T   GCATACTTCAAAGACTGTGTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAATGTTTTC
      1691                                                                                          1790
```

Figure 4 Alignment of the last 100 nucleotides of the HBV sequences found at the integration site in the isolated 21 HBV/cellular DNA junctions. Nucleotides position is calculated according to the HBV adw sequence (accession number V00866). The same short sequence AGCA (in bold) is found at the junction in tumors 54 T, GR2s, and 95T

```
100T  GCAGCCCCAGCCTCACCGGCACCTTCGCCGGCGCCCTCGCCCACCCCAGCCCCGCCCCAGAAGGAGCAGCCCCCCGCGGAGACCCCTACAGGTAGGAATGA
54T   CATTATAATAGAGTTGATCCTTATTATTCATGACTATTCCAGAACAATTCCACAGAATGCAATATAGAATAATTTATAATTCTGTTGGAAGCACAATGTA
63T   GTGTGGCTTTCAGGGATTACCAAAGACAGAATGTTGAATGCATTGTATGGTTTCTGTCTTCAGTCATAGGAACCATTCTTTGCTCTAGGAGTAGCAATGAG
83T   CTGAGGCTTAAAGAGATTAAGTAATACACCTCATCAAATGGCTACAGATTAGTAGAACTGAGATTACTAAGGAGATGTCAGAAAATGCCCAAAGTACCCTG
95T   AGACACAGTCCAAAGTTCTAAAACCCCCAATGGAGGGGCTGAGAAGTAAACATCAGAGTAATAAAAATTGATGCCTTTCACTGAATCTTTGTGAATGCCAG
FR2   CTAACTTGCTGTAACCTTAGCATTAGGCTAAGGTAAGTATTATTAGTCCCAGTGTGCAAATGAAGTTTACCCGCAGACATTTAACCTTACCAAAGTCACTC
FR3   AGATTGAGCCACTGCATTCCAGCCTGGGTGACACAGCAAGAACAAGACTCCCATCTCCCAAAAAAAAAATAAAAAAAAAGACGTTCACTGTCAAAAAAAAC
GR1   TTGATGAGTAGCTTCACATGTAAAAGCATGGATGCAGTTACTGGTGGCCTTAAAAACAAAGTCACCTAAGTTTGAGGCATTCTGGGGAGATGGAAAGGGCT
GR2   TGGCCTCCGAAAGTGCTGGGATTATAGGCATGAGGCACTGTGCCAGGTCCTTTTAATTTTTTTTAAATTATATAATATTGCACATAATCACATTAAAATAA
GR2s  TTGTATATACAAGAGCTATTTTTGTTCAACAATTTCAGCTTTTGGAAATATCTGGGAAAGGATTGCTTTGTGTGATAATGAAAGTGCACCTGCTCCATGTA
SA1   ACTAAAAATACCAAAAAAAAAAAAAAAATTAGCTCAGCATGGTGGCATGTGCCTGTAATCCCAGCTACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATCA
SA2   ATAGACCATGTATTTCCGAATGTTATAGTAAAAATATGAAAGTAATTGTTTTAGAAACATAAAGTTATGGCTTTGTCAAAGATAAAGACTTACTCTGACTC
SA2s  CCAACCCATATACTTTTATATCACTGTAAAACTGTGGATGGGATTCATCAGCCTAAAGATTAACTTGGTTATTCAGGTATTATCTCACCCACTTTGTGAAAC
SA5   TACATTTTGTTTTTAAAATCTACATCACTTAAAACCCTCTGCTATTATAAAGTCTGTTAGGATGTCTATTCACTCTTGTGAATAGACATTCACAAGACATT
```

Figure 5 Alignment of the 100 nucleotides of the cellular sequence preceding the integration site, and recovered from the data banks for 14 junctions. Highly homologous sequences (in bold) are present at the integration site in four tumors (ATGA in 100T/ATGTA in 54T and GR2s/ATGAG in 63T). Short recurrent sequences are preceding (5 to 87 bp) the integration site: AAAGT (in pink) in eight tumors: 83T, 95T, FR2, GR1, GR2, GR2s, SA2 (twice), and SA5; AAATT (in orange) in three tumors: 95T, GR2, and SA1; AAAGA (in blue) in five tumors: 63T, 83T, FR3, SA2 (twice) and SA2s; TAAGT (in dark green) in three tumors: 83T, FR2, and GR1; and AGAGT (in light green) in two tumors: 54T and 95T

specific primers on exon 10 and 12. By using these primers we obtained the expected genomic 713 bp band in the tumor DNA and normal liver DNA (Figure 3, B2, lanes 2 and 3), but not in the tumor cDNA, therefore ruling out contamination of the tumor derived cDNA by cellular DNA. Taken overall, our results are therefore consistent with a chimeric HBV-FR7 transcript expressed in the FR7 tumor.

Our work also provides further information about the DNA motifs at the integration site both in the viral and cellular genome (Figures 4 and 5). In fact, we found that, in three different tumors, the HBV sequence was interrupted at the same nucleotide position (Figure 4). Furthermore, the cellular sequence at the integration site (last nucleotides) was highly homologous (ATGA in 100T/ATGTA in 54T and GR2s/ATGAG in 63T) in four tumors. Other highly homologous repeated sequences (AAAGA/ AAAGT) were found to precede (5 to 87 bp) the integration site in several tumors. Finally the AAATT sequence was found in three different tumors, 32 (in two tumors) and 71 bp before the HBV/cellular DNA junction (Figure 5). In addition to that, the viral (AGCA) and the cellular (ATGTA) sequence at the integration site was the same in two different tumors: 54T and GR2s (Figures 4 and 5). These new data may have implications for further studies focused on the molecular mechanisms of HBV DNA integration into the cellular genome.

We have provided evidence that HBV genome insertion in a cellular gene is not an anecdotical event since it occurred in 30% of cases. This breakthrough is due to a combination of our efficient PCR-based approach, allowing to study several tumors simultaneously, and the exponentially growing bulk of genomic sequences now available in data banks. We detected HBV integration in cellular genes in HCC developing in both cirrhotic and non-cirrhotic livers and in both HBsAg-positive and negative patients (Table 1), thus establishing that this is one of the relevant synergistic mechanisms of liver carcinogenesis.

Strikingly, all the genes we have found at the HBV integration site are involved in cancer and/or are key regulators of cell proliferation and death. We did not identify common HBV integration sites at the chromosomal level (Table 1). This means that HBV acts by similar but different mechanisms, as compared to those reported for slow oncogenic retroviruses and WHV. However, the recent demonstration, by an independent group, of another case of HBV integration into the hTERT gene in an independent HCC, does suggest that common integration sites do exist and may predictably be identified by a large scale screening of HBV DNA integration sites.

In conclusion, our results provide evidence that integration of HBV into the cellular genome of HCC cells frequently implicates cellular genes and that it can be used as a unique approach, in humans, to identify cancer-related genes.

Acknowledgments

This work was supported by grants from INSERM, (Institut National de Santé et Recherche Médicale), LNC (Ligue Nationale contre le Cancer), ARC, (Association pour la Recherche contre le Cancer).

References

Berasain C, Patil D, Perara E, Huang SM, Mouly H and Brechot C. (1998). *Oncogene,* **16,** 1277 – 1288.

Berridge MJ, Bootman MD and Lipp P. (1998). *Nature,* **395,** 645 – 648.

Brechot C, Gozuacik D, Murakami Y and Paterlini-Brechot P. (2000). *Semin. Cancer Biol.,* **10,** 211 – 231.

Buendia MA. (1992). *Semin. Cancer Biol.,* **3,** 309 – 320.

Cecconi F and Meyer BI. (2000). *FEBS Lett.,* **480,** 63 – 71.

Chami M, Gozuacik D, Lagorce D, Brini M, Falson P, Peaucellier G, Pinton P, Lecoeur H, Gougeon ML, le Maire M, Rizzuto R, Brechot C and Paterlini-Brechot P. (2001). *J. Cell Biol.,* **153,** 1301 – 1314.

Chami M, Gozuacik D, Saigo K, Capiod T, Falson P, Lecoeur H, Urashima T, Beckmann J, Gougeon ML, Claret M, Maire M, Brechot C and Paterlini-Brechot P. (2000). *Oncogene,* **19,** 2877 – 2886.

Chisari F. (1995). *Hepatology,* **22,** 1316 – 1325.

Dejean A, Bougueleret L, Grzeschik KH and Tiollais P. (1986). *Nature,* **322,** 70 – 72.

Durfee T, Mancini MA, Jones D, Elledge SJ and Lee WH. (1994). *J. Cell Biol.,* **127,** 609 – 622.

Evans MJ, Carlton MB and Russ AP. (1997). *Trends. Genet.,* **13,** 370 – 374.

Fourel G and Tiollais P. (1994). *Primary liver cancer: Etiological and progression factors.* Bréchot C (ed). CRC Press: London, pp. 89 – 124.

Freeman A, Morris LS, Mills AD, Stoeber K, Laskey RA, Williams GH and Coleman N. (1999). *Clin. Cancer Res.,* **5,** 2121 – 2132.

Garcia M, De Thé H, Tiollais P, Samarut J and Dejean A. (1993). *Cell Biology,* **90,** 89 – 93.

Hanahan D and Weinberg RA. (2000). *Cell,* **100,** 57 – 70.

Ito M, Yuan CX, Malik S, Gu W, Fondell JD, Yamamura S, Fu ZY, Zhang X, Qin J and Roeder RG. (1999). *Mol. Cell,* **3,** 361 – 370.

Horikawa I and Barrett JC. (2001). cis. Activation of the human telomerase gene (htert) by the hepatitis b virus genome. *J. Natl. Cancer Inst.,* **93,** 1171 – 1173.

Johnson P. (1994). *Primary liver cancer: Etiological and progression factors.* Bréchot C (ed). CRC Press: London, pp. 31 – 40.

Jonkers J and Berns A. (1996). *Biochim. Biophys. Acta.,* **1287,** 29 – 57.

Largaespada DA. (2000). *Leukemia,* **14,** 1174 – 1184.

Li J, Shen H, Himmel KL, Dupuy AJ, Largaespada DA, Nakamura T, Shaughnessy Jr JD, Jenkins NA and Copeland NG. (1999). *Nat. Genet.,* **23,** 348 – 353.

Lin KH, Shieh HY, Chen SL and Hsu HC. (1999). *Mol. Carcinog.,* **26,** 53 – 61.

Liu K, Hodes RJ and Weng N. (2001). *J. Immunol.,* **166,** 4826 – 4830.

Matsubara K and Tokino T. (1990). *Mol. Biol. Med.,* **7,** 243 – 260.

Meyerson M. (2000). *J. Clin. Oncol.,* **18,** 2626 – 2634.

Minami M, Poussin K, Brechot C and Paterlini P. (1995). *Genomics,* **29,** 403 – 408.

Musahl C, Holthoff HP, Lesch R and Knippers R. (1998). *Exp. Cell Res.,* **241,** 260 – 264.

Robinson WS, Miller RH and Marion PL. (1987). *Hepatology,* **7,** 64S – 73S.

Schluter V, Meyer M and Hofschneider P. (1994). *Oncogene,* **9,** 1 – 10.

Todorov IT, Werness BA, Wang HQ, Buddharaju LN, Todorova PD, Slocum HK, Brooks JS and Huberman JA. (1998). *Lab. Invest.,* **78,** 73–78.

Tye BK. (1999). *Annu. Rev. Biochem.,* **68,** 649–686.

Wang J, Chenivesse X, Henglein B and Brechot C. (1990). *Nature,* **343,** 555–557.