

Comparison Study on the Complete Sequence of Hepatitis B Virus Identifies New Mutations in Core Gene Associated with Hepatocellular Carcinoma

Yu Zhu^{1,2}, Yan Jin², Xia Guo², Xin Bai², Taoyang Chen³, Jinbing Wang³, Gengsun Qian², John D. Groopman⁴, Jianren Gu², Jinjun Li², and Hong Tu²

Abstract

Background: Mutations in the hepatitis B virus (HBV) genome may influence the activity of liver disease. The aim of this study was to identify new viral variations associated with hepatocellular carcinoma (HCC).

Methods: We carried out a comparison study on the complete sequence of HBV isolated from 20 HCC and 35 non-HCC patients in Qidong, China, an area with a high incidence of HCC. We compared the HBV sequences in a consecutive series of plasma samples from four HCC cases before and after the occurrence of HCC. In addition, we selected four mutations in the HBV core (C) gene to verify their relationships to HCC in an independent set of 103 HCC cases and 103 sex- and age-matched non-HCC controls.

Results: The pre-S deletion and 12 point mutations, namely, the pre-S2 start codon mutation, T53C in the pre-S2 gene, T766A in the S gene, G1613A, C1653T, A1762T, G1764A in the X gene, and G1899A, C2002T, A2159G, A2189C, and G2203W (A or T) in the pre-C/C gene, showed close associations with HCC. In the validation study, A2159G, A2189C, and G2203W showed consistent associations with HCC by univariate analysis. Multivariate analysis showed that A2189C and G2203W were independent risk factors for HCC. The odds ratios (95% confidence interval) were 3.99 (1.61-9.92) and 9.70 (1.17-80.58), respectively, for A2189C and G2203W.

Conclusions: These results implicate A2189C and G2203W as new predictive markers for HCC.

Impact: The complete genome analysis of HBV provided pilot data for the identification of novel mutations that could serve as markers for HCC. *Cancer Epidemiol Biomarkers Prev*; 19(10); 2623-30. ©2010 AACR.

Introduction

Hepatitis B virus (HBV) infection is one of the most serious and prevalent global health problems. Although an effective vaccine has been used for two decades, >350 million people worldwide are chronically infected with HBV and are at increased risk for developing hepatocellular carcinoma (HCC; ref. 1). In addition to host factors, viral factors per se could also play an important role in determining clinical outcomes (2).

HBV is a small enveloped DNA virus of the Hepadnaviridae family. The virus has a partially double-stranded DNA genome of about 3.2 kb with four overlapping open

reading frames that encode the envelope protein, X protein, DNA polymerase, and nucleocapsid. HBV replicates through RNA-intermediated reverse transcription. Because reverse transcriptase lacks proofreading activity, errors in HBV replication occur at a much higher rate than in other DNA viruses. Hence, various mutations may be observed in the HBV genome during long-term infection, and some of them could serve as viral markers for predicting the development of HBV-associated HCC. Although many studies have indicated that HBV carriers with a basal core promoter (BCP) mutation (3, 4) and a pre-S deletion (5) are at increased risk for HCC, it remains unclear whether other predictive markers might be found by comparative analysis of the complete HBV genomes at different stages of liver disease. To screen for new risk variants for HCC, we carried out a study comparing the complete sequences of HBV isolated from 20 HCC and 35 non-HCC patients in Qidong, China, an area with a high incidence of HCC. Consecutive plasma samples from four HCC patients were used to observe the evolution of mutations during the development of HCC. In addition, an independent case-control study was carried out to verify the association between the newly identified mutations in the C gene and HCC.

Authors' Affiliations: ¹Shanghai Medical College, Fudan University; ²State Key Laboratory of Oncogene and Related Genes, Cancer Institute of Shanghai Jiao-Tong University, School of Medicine, Shanghai, China; ³Qidong Liver Cancer Institute/Qidong Tumor Hospital, Jiangsu Province, China; and ⁴Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

Corresponding Author: Hong Tu, State Key Laboratory of Oncogene and Related Genes, Cancer Institute of Shanghai Jiao-Tong University, 2200-25 Xietu Road, Shanghai 200032, China. Phone: 86-21-64437301; Fax: 86-21-64437196. E-mail: tuhong@shsci.org

doi: 10.1158/1055-9965.EPI-10-0469

©2010 American Association for Cancer Research.

Materials and Methods

Patients and samples

Plasma samples were collected from the Qidong Liver Cancer Institute/Qidong Tumor Hospital between 1996 and 2006. All participants were positive for hepatitis B surface antigen (HBsAg) and HBV DNA. Patients with HCC were diagnosed on the basis of pathologic findings or an elevated serum α -fetoprotein level (≥ 400 ng/mL) combined with positive images on either computerized tomography or ultrasonography. Diagnosis of chronic hepatitis was based on the current Chinese diagnostic criterion for viral hepatitis (6). Patients with hepatitis C virus coinfection or cirrhosis were excluded from the study. For full-length genome analysis, 20 HCC patients and 35 non-HCC patients (21 chronic hepatitis patients and 14 chronic HBV carriers) were recruited. For validation of mutations in the C gene, an independent set of 103 HCC patients and 103 age- and sex-matched non-HCC patients (all chronic hepatitis patients) were recruited. For the longitudinal study, serial plasma samples from four HCC patients were obtained from an ongoing prospective cohort investigation of liver disease started in 1992 (3), in which 852 HbsAg-seropositive individuals and 786 HbsAg-seronegative individuals residing in the Qidong high-risk area were recruited. The plasma samples of each individual were collected annually. For each of the four cases, at least one PCR-amplifiable DNA sample was available before the onset of HCC. Written informed consent was obtained from all patients, and the study protocol was approved by the local ethical committee at the Qidong Liver Cancer Institute/Qidong Tumor Hospital and Shanghai Cancer Institute. The study was done in accordance with the principles of the Declaration of Helsinki.

Amplification and sequencing of the complete HBV genome and the C gene

HBV DNA was extracted from 100 μ L of plasma using a QIAamp DNA blood mini kit (QIAGEN) according to the manufacturer's instructions or from 50 μ L plasma by boiling in 5 μ L DNA extraction buffer (PG Biotech Co.) for 10 min. The HBV full-length sequence was amplified by PCR using FLP1 [5'-TTTTTACCTCTGCTAAT-CATC-3' [nucleotides (nt) 1821-1843], forward] and

FLP2 [5'-AAAAAGTTGCATGGTGCTGGTG-3' (nts 1825-1804), reverse] as primers. The amplification was carried out in a 50 μ L reaction mixture containing 5 μ L 10 \times buffer, 4 μ L 2.5 mmol/L dNTPs, 2 μ L 10 μ mol/L forward and reverse primers, and 1 U LA Taq (TaKaRa Bio). PCR was done under the following conditions: 94°C for 3 minutes, followed by 94°C for 30 seconds, 58°C for 30 seconds, and 72°C for 3 minutes for 35 cycles, with a final extension at 72°C for 7 minutes. PCR products were purified (Axygen Scientific, Inc.) and cloned into the pUCm-T vector (Shanghai Shenergy Biocolor BioScience and Technology Co., Ltd.) for sequencing. Sequencing was done with the BigDye terminator cycle-sequencing reaction kit and Prism 3700 DNA analyzer (Applied Biosystems) using pUCm-T vector universal primers and HBV-specific primers. The HBV C gene from nts 1901 to 2275 was amplified by seminested PCR using pre-C F1 [5'-TTCACCTCTGCCTAATCATCTC-3' (nts 1824-1845), forward] and HBV2433R [5'-GATTGAGATCTTCTGCGACGC-3' (nts 2433-2413), reverse] as the first-round primers and pre-C F1 and pre-C R2 [5'-CCACACTCCAAAAGACACCAAAA-3' (nts 2275-2254), reverse] as the second-round primers. PCR was done under the conditions described above except that the elongation time was changed to 1 minute. The PCR products were gel purified and were then used as templates for automated sequencing. Sequences of the complete genome or C gene were compared using MEGA4.1 (7).

Serologic markers

HBsAg and hepatitis B e antigen were tested by commercially available assay (Kehua, Inc.).

HBV genotyping

HBV genotypes were determined by comparing the sequence of the complete genome or X gene with a set of database-derived standard sequences. Standard sequences were retrieved from GenBank/DBJ/EMBL. A phylogenetic tree was constructed with MEGA4.1 software (7).

Statistical analysis

The Student's *t* test was used for continuous variables with normal distributions, and Pearson's χ^2 test or Fisher's exact test was applied to analyze categorical variables.

Table 1. Number of substitutions per 1,000 nts in the full genome and in various regions of HBV in HCC and non-HCC patients

	Full genome (nts 1-3215)	Pre-S1 (nts 2848-3204)	Pre-S2 (nts 3205-154)	S (nts 155-835)	X (nts 1374-1838)	Pre-C/C (nts 1814-2452)	P (nts 2307-1623)
HCC (<i>n</i> = 20)	15.0 \pm 3.7*	8.6 \pm 7.4	13.6 \pm 9.6	4.5 \pm 2.6	18.5 \pm 5.7	14.3 \pm 6.2	14.1 \pm 3.7
Non-HCC (<i>n</i> = 35)	11.0 \pm 4.7	6.5 \pm 5.0	8.5 \pm 5.9	3.5 \pm 2.7	9.8 \pm 5.5	8.8 \pm 5.6	11.1 \pm 4.5
<i>P</i>	0.002	0.222	0.017	0.208	<0.001	0.001	0.013

*Number of substitutions per 1,000 nts (mean \pm SD).

Table 2. Prevalence of HBV mutations throughout the complete genome in HCC and non-HCC patients

Mutation	Amino acid	Total (n = 55)	HCC (n = 20)	Non-HCC (n = 35)	P
Pre-S2					
Pre-S2 start codon	M 1 V/T/I	4 (7.3%)	4 (20.0%)	0 (0.0%)	0.014
T53C	F 22 L	5 (9.1%)	5 (25.0%)	0 (0.0%)	0.004
S					
T766A	S 204 R	3 (5.5%)	3 (15.0%)	0 (0.0%)	0.043
X					
G1613A	NC	13 (23.6%)	10 (50.0%)	3 (8.6%)	0.001
C1653T	H 94 Y	12 (21.8%)	11 (55.5%)	1 (2.9%)	<0.001
T1753C	I 127 T	5 (9.1%)	4 (20.0%)	1 (2.9%)	0.053
A1762T	K 130 M	25 (45.5%)	16 (80.0%)	9 (25.7%)	<0.001
G1764A	V 131 I	30 (54.5%)	19 (95.0%)	11 (31.4%)	<0.001
C1766T	NC	6 (10.9%)	4 (20.0%)	2 (5.7%)	0.175
Pre-C					
G1896A	W 28 Stop	5 (9.1%)	3 (15.0%)	2 (5.7%)	0.342
G1899A	G 29 D	6 (10.9%)	6 (30.0%)	0 (0.0%)	0.001
C					
C2002T	NC	3 (5.5%)	3 (15.0%)	0 (0.0%)	0.043
A2159G	S 87 G	6 (10.9%)	5 (25.0%)	1 (2.9%)	0.020
A2189C	I 97 L	9 (16.4%)	8 (40.0%)	1 (2.9%)	0.001
G2203W	NC	6 (10.9%)	6 (30.0%)	0 (0.0%)	0.001

Abbreviations: W, A or T; NC, no change.

Multivariate analyses with logistic regression were used to determine the independent factors that correlated with HCC. All of the tests were two-tailed, and $P < 0.05$ was considered statistically significant. SPSS (SPSS, Inc.) version 12.0 was used for statistical analysis.

Results

Comparison of HBV mutation rates between HCC and non-HCC patients

The complete sequences of HBV from 20 HCC and 35 non-HCC control patients were determined by PCR direct sequencing. There were no significant differences in age or in the distribution of HBV genotypes between HCC and non-HCC patients (43.6 ± 9.9 versus 37.2 ± 10.8 , $P = 0.074$ for age; 2:18 versus 6:29, $P = 0.696$ for the genotype B to genotype C ratio). The number of substitutions per nt was calculated after comparing with each corresponding prototype sequence (GenBank Accession No. GU434374 for genotype C and GU434373 for genotype B, both from an HBV carrier in Qidong, China). The average rate of nt substitutions within the whole HBV genome was 15.0 ± 3.7 per 1,000 nts for HCC patients and 11.0 ± 4.7 per 1,000 nts for non-HCC patients ($P = 0.002$). Table 1 shows the nt substitution rates in various regions of HBV. The HCC group had significantly more nt substitutions in the pre-S2 ($P = 0.017$), X ($P < 0.001$), pre-C/C ($P = 0.001$), and C ($P = 0.013$) regions. The pre-S1 and S genes only showed slightly increased

nt substitutions in the HCC compared with the non-HCC group ($P = 0.222$ and $P = 0.208$, respectively).

Identification of HCC-related mutations within the HBV genome

Table 2 lists all the mutations within the complete genome of HBV that tended to occur more frequently in HCC patients than in non-HCC control patients. These mutations were not genotype-specific polymorphisms and could emerge in both genotype B and C viruses. A total of 12 mutations showed statistically significant differences between HCC and non-HCC groups. These included well-studied mutations (e.g., the pre-S2 start codon mutation, C1653T, A1762T/G1764A in X) and less well-defined mutations [e.g., T53C in pre-S2, T766A in S, G1613A in X, G1899A in pre-C, and C2002T, A2159G, A2189C, and G2203W (A or T) in C]. Among these 12 point mutations, four (33.3%) were located in the X gene and four (33.3%) were in the C gene. Although the S gene constitutes 21.2% of the entire HBV genome, there was only one mutation (8.3%) in the S gene that showed a significantly higher frequency in the HCC group. These data suggest that HCC-related mutations were not likely to distribute evenly throughout the HBV genome.

There were three types of deletion mutations in the HBV genome (Table 3). The common type was the deletion in the pre-S gene, which was detected in 5 of the 20 HCC patients and in one of the 35 non-HCC patients (25.0% versus 2.9%; $P < 0.05$). The C gene deletion was

Table 3. Deletions in HBV genomes isolated from 55 patients

Sample	Related region	Range	Size (bp)
HCC			
A0271	Pre-S2	nts 23-55	33
F0294	Pre-S2	nts 26-55	30
B0206	Pre-S2	nts 3214-55	57
616-93	Pre-S1	nts 2848-2865	18
		nts 2890-2976	87
S169	Pre-S1/S2	nts 2964-3215	252
A0527	CP/X, pre-C	nts 1793-1819	27
A0583	C	nts 2155-2229	75
Non-HCC			
B0949	Pre-S2	nts 23-55	33
E0113	C	nts 2141-2227	87

found in one patient in each group. In addition, one HBV isolate from a HCC patient showed a deletion spanning the X and pre-C region.

Longitudinal observation of HBV mutations during the development of HCC

We retrieved serial plasma samples from four HCC patients and determined the complete HBV DNA sequences in the samples taken before and after the diagnosis of HCC. Analysis was focused on those putative HCC-related mutations identified from the above cross-sectional study. As illustrated in Table 4, HCC-related mutations showed a

gradual accumulation during the development of HCC. It is noteworthy that, in patients 252, 371, and 416, the mutation profiles of HBV in the plasma 1 to 2 years before HCC were identical to those in the HCC stage, suggesting that most HCC-related mutations took place early on before the occurrence of HCC. Indeed, G2203W had existed in the circulating HBV at least 8 years before HCC onset in patient 416, and G1613A, the A1762T/G1764A double mutation, the pre-S deletion, and C2002T were detectable in the plasma samples 5 to 6 years before HCC in patients 99 and 252. However, T1753C in the X gene occurred relatively closer to the HCC stage. Although it was found in patients 99, 252, and 416 at the time of diagnosis of HCC, there was no such mutation found in the plasma samples taken 3, 5, or 8 years before HCC onset.

Validation of the associations between the HBV C gene mutations and HCC

To confirm the results that C2002T, A2159G, A2189C, G2003W, and the deletion in the C gene (not overlapped with the pre-C region) could indeed increase the risk for HCC, we did an independent case-control study by using plasma samples from 103 HCC patients and 103 non-HCC control patients. The age and gender of the patients were matched, and there was no difference in the genotype distribution of HBV between the groups (Table 5). The frequencies of mutation increased in HCC patients, from 3.9% (C2002T), 23.3% (A2159G), 22.3% (A2189C), and 1.0% (G2003W) in non-HCC patients to 9.7%, 37.9%, 48.5%, and 10.7% in HCC patients, respectively. Consistent with the observation from the full-length HBV DNA analysis, the frequencies of A2159G,

Table 4. Longitudinal observation of HBV mutations during HCC development

Sample	Sex/age	Pre-S/S (nts 2848-835)				X (nts 1374-1838)						Pre-C/C (nts 1814-2452)					
		Pre-S2 start codon	53 T→C	766 T→A	Pre-S deletion	1613 G→A	1653 C→T	1753 T→C	1762 A→T	1764 G→A	1766 C→T	1896 G→A	1899 G→A	2002 C→T	2159 A→G	2189 A→C	2203 G→W
Patient 99																	
6 y before	F/49	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	○
3 y before		○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	●
Diagnosis		○	○	○	○	○	○	●	●	●	○	○	○	○	○	○	●
Patient 252																	
5 y before	M/39	○	○	○	●	●	○	○	●	●	○	○	○	●	○	○	○
2 y before		○	○	○	●	●	○	○	●	●	○	○	○	●	○	○	○
Diagnosis		○	○	○	●	●	○	○	●	●	○	○	○	●	○	○	○
Patient 371																	
1 y before	M/35	●	○	○	○	●	●	○	●	●	○	●	○	●	●	●	●
Diagnosis		●	○	○	○	●	●	○	●	●	○	●	○	●	●	●	●
Patient 416																	
8 y before	M/49	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
2 y before		○	○	○	○	●	●	●	●	●	○	○	○	○	○	○	○
Diagnosis		○	○	○	○	●	●	●	●	●	○	○	○	○	○	○	○

Abbreviations: F, female; M, male; ○, wild type; ●, mutation.

Table 5. Clinical and virological characteristics of 103 HCC and 103 non-HCC patients

Characteristics	HCC (n = 103)	Non-HCC (n = 103)	P	OR (95% CI)
Age (mean ± SD)	50.6 ± 10.4	48.5 ± 10.1	Matched	—
Sex (male)	89 (86.4%)	89 (86.4%)	Matched	—
HBeAg positive	34 (33.0%)	21 (20.4%)	0.041	1.92 (1.02-3.62)
Genotype				
B	7 (6.8%)	5 (4.9%)	0.552	1.43 (0.44-4.66)
C	90 (87.4%)	88 (85.4%)	0.684	1.18 (0.53-2.62)
Others	6 (5.8%)	10 (9.7%)	0.298	0.58 (0.20-1.65)
Mutations in C genes				
Deletion	6 (5.8%)	1 (1.0%)	0.119	6.31 (0.75-53.37)
C2002T	10 (9.7%)	4 (3.9%)	0.097	2.66 (0.81-8.78)
A2159G	39 (37.9%)	24 (23.3%)	0.023	2.01 (1.09-3.68)
A2189C	50 (48.5%)	23 (22.3%)	<0.001	3.28 (1.79-6.00)
G2203W	11 (10.7%)	1 (1.0%)	0.003	12.20 (1.54-96.30)

Abbreviations: HBeAg, hepatitis B e antigen; OR, odds ratio.

A2189C, and G2003W were significantly higher in HCC patients compared with those in non-HCC controls ($P = 0.023$, $P < 0.001$, and $P = 0.003$, respectively). However, the frequency of the C2002T mutation did not show a statistically significant difference between the groups ($P = 0.097$). Interestingly, A2159G seemed to have a correla-

tion with A2189C. In 63 cases with the A2159G mutation, 53 (84.1%) were coupled with A2189C. Therefore, a multivariate analysis indicated that A2189C [odds ratio, 3.99; 95% confidence interval (95% CI), 1.61-9.92] and G2003W (odds ratio, 9.70; 95% CI, 1.17-80.58), but not A2159G, were independent predictive factors for HCC (Table 6).

Table 6. Multivariate analysis of risk factors for HCC

Factor	OR (95% CI)	P
Age*		
<49 y	1	0.852
≥49 y	0.94 (0.50-1.76)	
Sex		
Female	1	0.437
Male	1.43 (0.58-3.51)	
HBV genotype		
B	1	0.262
C	0.48 (0.14-1.72)	
HBeAg		
Negative	1	0.109
Positive	1.76 (0.88-3.53)	
A2159G mutation		
Absence	1	0.685
Presence	0.82 (0.32-2.11)	
A2189C mutation		
Absence	1	0.003
Presence	3.99 (1.61-9.92)	
G2203W mutation		
Absence	1	0.035
Presence	9.70 (1.17-80.58)	

*Two groups were divided by the median value.

Discussion

HCC is the leading cause of cancer mortality and accounts for almost one third of the malignancies in Qidong, China (8). The high incidence of HCC is the consequence of a high prevalence of HBV infection (9) and of exposure to aflatoxin B1 (10). We and others have reported that the mutations in the HBV pre-S gene and BCP region were closely associated with HCC in Qidong (3, 11, 12). However, the mutations in other regions of HBV that may also play a role in the development of HCC have not yet been explored in Qidong. To this aim, we compared the full-length sequences of HBV isolated from 20 HCC and 35 non-HCC patients. The HCC patients had a higher frequency of nt substitutions in the HBV genome, with an average mutation rate of 15.0 ± 3.7 per 1,000 nts. The regions with significant differences in the mutation rate between HCC and non-HCC patients were, in rank order, X ($P < 0.001$), pre-C/C ($P = 0.001$), P ($P = 0.013$), pre-S2 ($P = 0.017$), S ($P = 0.208$), and pre-S1 ($P = 0.222$). Although a large number of sporadic mutations were observed in individual HCC patient, there were only a few bona fide mutations associated with HCC. These HCC-related mutations were found to be clustered rather than evenly distributed throughout the HBV genome. Although the region nts 1613 to 1764 in the X gene and nts 1899 to 2203 in the pre-C/C gene constitute only 14.2% of the HBV genome, they contained 75.0% (9 of 12) of the mutations showing a significantly

higher frequency in HCC patients from the full-length HBV sequence analysis (Table 2). Among the four HCC-related mutations in the X gene, C1653T, A1762T, and G1764A have been studied extensively (3, 12-16), whereas G1613A is less well defined. G1613A was first reported by Takahashi et al., who noted that, of 40 HCC tissue samples tested, 15 contained this type of mutation (13). Recent studies have suggested that it may be a molecular marker for HCC in genotype C-infected patients (17, 18). In the present study, although G1613A mutation could emerge in genotypes B as well as genotype C viruses, its association with HCC was only significant in the genotype C-infected patients (50% for HCC versus 10.3% for non-HCC; $P = 0.005$). Because the 1613 G-to-A mutation is a synonymous mutation for the X protein, its impact on viral pathogenesis may be exerted through the overlapping negative regulatory element of HBV (nts 1613-1636; ref. 19). Other rare documented or novel HCC-related mutations identified from this study include T53C in the pre-S2 gene, T766A in the S gene, G1899A in the pre-C gene, and C2002T, A2159G, A2189C, and G2203W in the C gene. These data provided potential targets for early diagnosis and treatment of HCC.

Comparing the complete genome sequence of HBV, we found that the pre-C/C gene was the region second to that of the X gene that exhibited the most significant difference in mutation frequency between the HCC and non-HCC groups ($P = 0.001$). Of the 12 HCC-related mutations within the HBV genome, four were located in the middle part of the C gene. Compared with the hot-spot mutations in the pre-S and X/BCP regions, the effect of C gene variability on HCC progression is less well delineated. Hence, we carried out a case-control study on 103 HCC patients and 103 sex- and age-matched non-HCC control patients to confirm our findings from the full-length HBV DNA comparison study. To our knowledge, this is the only investigation from mainland China that has focused on the relationship between C gene mutations and HCC. Univariate analysis indicated that the A2159G (S87G), A2189C (I97L), and G2203W (synonymous) mutations were closely associated with the development of HCC. Multivariate analysis revealed that the A2189C and G2203W mutations were independent predictive factors for HCC. The core protein of HBV is the major target for the antiviral immune response (20). It contains CTL epitopes, T-helper cell epitopes, and B-cell recognition sites (21-23). Although a variety of mutations may emerge in C genes during the immune clearance phase, only a few mutations within or flanking the HBcAg epitopes have been reported to be of clinical relevance (24-26). About the association of the C gene mutations and HCC, Sung et al. (25) reported that mutations at nts 1961, 1938, 2045, 2136, 2239, and 2441 were associated with decreased risk for HCC, whereas no mutation in the C gene was found to be related to an increased risk for HCC in Taiwan. Such reverse associations were not observed in the present study, probably be-

cause the samples analyzed were collected at the baseline of a prospective cohort, whereas our experiment was based on samples collected at the time of diagnosis of HCC. Alternatively, it may be due to the different genotypes or subgenotypes circulating in Taiwan and Qidong. In Taiwan, >60% of patients were infected with genotype B (27); however, in Qidong, as shown in Table 5, around 85% patients were infected with genotype C. The A2159G and A2189C mutations were noted in an early study based on 15 tissue samples of HCC patients (28). The A2159G mutant was later isolated in 33% (4 of 12) of children with HCC and in 0% (0 of 23) of non-HCC control children (24). However, there has been a lack of large-scale confirmatory studies conducted in adult HCC patients. Because 2159 A to G and 2189 A to C are missense mutations resulting in an amino acid change of HBcAg codon 87 and 97, respectively, it is possible that the mutants could enhance hepatocarcinogenesis through the altered function of HBcAg. Because codon 87 is located within a known B-cell epitope (29) and codon 97 within a potent T-cell epitope (30), these two mutations may change the immunodominant epitopes of HBcAg and permit HBV escape from immune clearance. G2203W is a synonymous mutation. Its biological consequence is an enigma at present. It is likely that G2203W does not enhance the virulence of the virus. It may be accompanied by other critical mutations in the HBV genome, thus being selected from viral quasispecies during the development of HCC. Nonetheless, this intergenotypic polymorphism could serve as a useful signature for HCC prediction.

Most previous studies on the relationship of HBV mutations and HCC with full-length genome analysis were conducted by using the samples taken after a diagnosis of cancer (18, 31). Because most HBV mutations are acquired during the course of chronic infection rather than being obtained from an initial infection (12, 15), it is important to know when or at which stage of the disease the mutations developed. This study was facilitated by the availability of prospectively collected plasma samples from Qidong. Our longitudinal observation showed that hot-spot mutations accumulated in different combinations during the development of HCC. In three patients (99, 252, and 416) who had plasma samples taken 5 to 8 years before developing HCC, the mutation numbers all increased at the stage of HCC onset. Recently, increasing evidence have shown that an HBV strain with a complex mutation pattern rather than a single mutation was associated with a higher risk for advanced liver disease. These combinations included C1653T plus A1762T/G1764A (14), A1762T/G1764A plus C1766T and/or T1768A (12), pre-S deletion plus A1762T/G1764A (5), and deletions in BCP plus C and/or pre-S (32). Those cross-sectional studies provided little information on the evolution of the HBV sequence during HCC development. Our longitudinal study allowed us to see that the HBV mutation profile remained consistent for at least 2 years before HCC onset, indicating that HBV mutations could be served as early markers for the detection of

HCC. Our study also suggested that, during the development of HCC, HBV mutations may occur in a certain order. Consistent with the earlier observation that A1762T/G1764A was detectable up to ≥ 8 before the diagnosis of HCC (3, 15), we also found that the A1762T/G1764A mutation existed in the plasma of patients 99 and 252 for 5 to 6 years before HCC. The early events may also include G2203W, G1613A, C2002T, and the pre-S deletion. However, the T1753C mutation emerged relatively late. Whereas patients 99, 252, and 416 had this mutation at the time of diagnosis of HCC, none of them had it 3 to 8 years before HCC onset. These data lead us to speculate that HCC-related mutations might have early and late types. They may play different roles at different steps of liver carcinogenesis. Many studies have been conducted about the pathologic effects of HBV mutants. The pre-S2 deletion mutants were found to induce the formation of "ground glass hepatocyte" (33, 34). It was also found that the shortened large envelope protein accumulated in endoplasmic reticulum and initiated endoplasmic reticulum stress to induce oxidative DNA damage and genomic instability (35). In the X/BCP region, the A1762T/G1764A/C1766T/T1768A clustering mutations could modify the biological functions of HBx by controlling cell proliferation and viability, thus enhancing the carcinogenesis potential of HBV (12). The X/BCP mutations, as well as nt2189 mutation in the C gene, have been shown to confer significantly higher replication capacity on wild-type viruses (36-38). It is noted that most of the findings about viral replication were based on the results from cell culture system. Whether these mutations have impacts on viral life cycle *in vivo* in chronic hepatitis patients is largely unknown. In this study, we have analyzed the relationship between the core mutations and the peripheral HBV DNA levels in 206 patients, but no correlation was found for any type of mutation (data not shown). Indeed, it is well established that the level of viremia declines over the course of HBV infection, especially during the period of cirrhosis and HCC. Because HBV core protein is a principle target for immune response, immune-mediated pathogenesis is

likely to play a key role in the progression of liver diseases by the core mutants. It is generally thought that the core mutations were a result of selection under the pressure of immune response. Mutations in the major epitopes may allow immune escape and lead to the persistence of HBV infection. The prolonged viral persistence cause continuous hepatocyte injury and subsequent regeneration, which significantly increases the risk for HCC.

The limitation of this investigation is that we only used a case-control study to validate the associations between HBV core mutations and HCC. A prospective cohort study with a large number of HBV core mutant-infected patients and a long period of follow-up will better assess the interplay between HBV mutations and HCC. Such a longitudinal investigation is being carried out in Qidong to confirm our findings from a cross-sectional study.

Our study highlights the influence of genetic variants in the HBV C gene on the progression of HCC. The complete genome analysis of HBV provided pilot data for the identification of other novel mutations related to HCC. A combined examination of different viral mutations could predict the progression of liver disease more precisely, thus helping those who are at high risk for HCC to benefit from early diagnosis and treatment.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

Chinese State Key Project Specialized for Infectious Diseases (2008ZX10002-015; H. Tu) and National Institute of Environmental Health Sciences grant PO I (ES06052; J.D. Groopman).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 05/05/2010; revised 07/05/2010; accepted 08/03/2010; published OnlineFirst 08/10/2010.

References

- Liaw YF, Chu CM. Hepatitis B virus infection. *Lancet* 2009;373:582-92.
- Lin CL, Kao JH. Hepatitis B viral factors and clinical outcomes of chronic hepatitis B. *J Biomed Sci* 2008;15:137-45.
- Kuang SY, Jackson PE, Wang JB, et al. Specific mutations of hepatitis B virus in plasma predict liver cancer development. *Proc Natl Acad Sci U S A* 2004;101:3575-80.
- Yuen MF, Tanaka Y, Shinkai N, et al. Risk for hepatocellular carcinoma with respect to hepatitis B virus genotypes B/C, specific mutations of enhancer II/core promoter/precore regions and HBV DNA levels. *Gut* 2008;57:98-102.
- Chen CH, Hung CH, Lee CM, et al. Pre-S deletion and complex mutations of hepatitis B virus related to advanced liver disease in HBeAg-negative patients. *Gastroenterology* 2007;133:1466-74.
- Guideline on prevention and treatment of chronic hepatitis B in China (2005). *Chin Med J (Engl)* 2007;120:2159-73.
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596-9.
- Chen JG, Zhu J, Parkin DM, et al. Trends in the incidence of cancer in Qidong, China, 1978-2002. *Int J Cancer* 2006;119:1447-54.
- Sun Z, Ming L, Zhu X, Lu J. Prevention and control of hepatitis B in China. *J Med Virol* 2002;67:447-50.
- Ming L, Thorgeirsson SS, Gail MH, et al. Dominant role of hepatitis B virus and cofactor role of aflatoxin in hepatocarcinogenesis in Qidong, China. *Hepatology* 2002;36:1214-20.
- Cao Z, Bai X, Guo X, Jin Y, Qian G, Tu H. High prevalence of hepatitis B virus pre-S mutation and its association with hepatocellular carcinoma in Qidong, China. *Arch Virol* 2008;153:1807-12.
- Guo X, Jin Y, Qian G, Tu H. Sequential accumulation of the mutations in core promoter of hepatitis B virus is associated with the development of hepatocellular carcinoma in Qidong, China. *J Hepatol* 2008;49:718-25.

13. Takahashi K, Akahane Y, Hino K, Ohta Y, Mishiro S. Hepatitis B virus genomic sequence in the circulation of hepatocellular carcinoma patients: comparative analysis of 40 full-length isolates. *Arch Virol* 1998;143:2313–26.
14. Kim JK, Chang HY, Lee JM, et al. Specific mutations in the enhancer II/core promoter/precore regions of hepatitis B virus subgenotype C2 in Korean patients with hepatocellular carcinoma. *J Med Virol* 2009; 81:1002–8.
15. Chou YC, Yu MW, Wu CF, et al. Temporal relationship between hepatitis B virus enhancer II/basal core promoter sequence variation and risk of hepatocellular carcinoma. *Gut* 2008;57:91–7.
16. Ito K, Tanaka Y, Orito E, et al. T1653 mutation in the box α increases the risk of hepatocellular carcinoma in patients with chronic hepatitis B virus genotype C infection. *Clin Infect Dis* 2006;42:1–7.
17. Shinkai N, Tanaka Y, Ito K, et al. Influence of hepatitis B virus X and core promoter mutations on hepatocellular carcinoma among patients infected with subgenotype C2. *J Clin Microbiol* 2007;45: 3191–7.
18. Sung JJ, Tsui SK, Tse CH, et al. Genotype-specific genomic markers associated with primary hepatomas, based on complete genomic sequencing of hepatitis B virus. *J Virol* 2008;82:3604–11.
19. Kramvis A, Kew MC. The core promoter of hepatitis B virus. *J Viral Hepat* 1999;6:415–27.
20. La Torre G, Nicolotti N, de Waure C, et al. An assessment of the effect of hepatitis B vaccine in decreasing the amount of hepatitis B disease in Italy. *Virol J* 2008;5:84.
21. Pumpens P, Grens E. HBV core particles as a carrier for B cell/T cell epitopes. *Intervirology* 2001;44:98–114.
22. Bozkaya H, Ayola B, Lok AS. High rate of mutations in the hepatitis B core gene during the immune clearance phase of chronic hepatitis B virus infection. *Hepatology* 1996;24:32–7.
23. Khakoo SI, Ling R, Scott I, et al. Cytotoxic T lymphocyte responses and CTL epitope escape mutation in HBsAg, anti-HBe positive individuals. *Gut* 2000;47:137–43.
24. Ni YH, Chang MH, Hsu HY, Tsuei DJ. Different hepatitis B virus core gene mutations in children with chronic infection and hepatocellular carcinoma. *Gut* 2003;52:122–5.
25. Sung FY, Jung CM, Wu CF, et al. Hepatitis B virus core variants modify natural course of viral infection and hepatocellular carcinoma progression. *Gastroenterology* 2009;137:1687–97.
26. Kim HJ, Lee DH, Gwak GY, et al. Analysis of the core gene of hepatitis B virus in Korean patients. *Liver Int* 2007;27:633–8.
27. Yang HI, Yeh SH, Chen PJ, et al. Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma. *J Natl Cancer Inst* 2008;100:1134–43.
28. Hosono S, Tai PC, Wang W, et al. Core antigen mutations of human hepatitis B virus in hepatomas accumulate in MHC class II-restricted T cell epitopes. *Virology* 1995;212:151–62.
29. Salfeld J, Pfaff E, Noah M, Schaller H. Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. *J Virol* 1989;63:798–808.
30. Menne S, Maschke J, Tolle TK, Lu M, Roggendorf M. Characterization of T-cell response to woodchuck hepatitis virus core protein and protection of woodchucks from infection by immunization with peptides containing a T-cell epitope. *J Virol* 1997;71:65–74.
31. Blackberg J, Kidd-Ljunggren K. Mutations within the hepatitis B virus genome among chronic hepatitis B patients with hepatocellular carcinoma. *J Med Virol* 2003;71:18–23.
32. Preikschat P, Gunther S, Reinhold S, et al. Complex HBV populations with mutations in core promoter, C gene, and pre-S region are associated with development of cirrhosis in long-term renal transplant recipients. *Hepatology* 2002;35:466–77.
33. Choi MS, Kim DY, Lee DH, et al. Clinical significance of pre-S mutations in patients with genotype C hepatitis B virus infection. *J Viral Hepat* 2007;14:161–8.
34. Tong S. Mechanism of HBV genome variability and replication of HBV mutants. *J Clin Virol* 2005;34:134–8.
35. Hartmann-Stuhler C, Prange R. Hepatitis B virus large envelope protein interacts with γ 2-adaptin, a clathrin adaptor-related protein. *J Virol* 2001;75:5343–51.
36. Baumert TF, Rogers SA, Hasegawa K, Liang TJ. Two core promoter mutations identified in a hepatitis B virus strain associated with fulminant hepatitis result in enhanced viral replication. *J Clin Invest* 1996;98:2268–76.
37. Parekh S, Zoulim F, Ahn SH, et al. Genome replication, virion secretion, and e antigen expression of naturally occurring hepatitis B virus core promoter mutants. *J Virol* 2003;77:6601–12.
38. Suk FM, Lin MH, Newman M, et al. Replication advantage and host factor-independent phenotypes attributable to a common naturally occurring capsid mutation (I97L) in human hepatitis B virus. *J Virol* 2002;76:12069–77.