

# Comparison of the Characteristics of Hepatitis B Virus Whole Genomes Derived From Patients With Hepatocellular Carcinoma and Chronic Hepatitis B Infection

**Shuang Zhang**

National Institute for Viral Disease Control and Prevention, Chinese Disease Control and Prevention

**Feng Wang**

National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention

**He Liu**

National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention

**Qiudong Su**

National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention

**Feng Qiu**

Shanxi Center for Disease Control and Prevention

**Guomin Zhang**

National Immunization Program, Chinese Center for Disease Control and Prevention

**Fuzhen Wang**

National Immunization Program, Chinese Center for Disease Control and Prevention

**Hui Zheng**

National Immunization Program, Chinese Center for Disease Control and Prevention

**Liping Shen** (✉ [shenlp@ivdc.chinacdc.cn](mailto:shenlp@ivdc.chinacdc.cn))

Chinese Center for Disease Control and Prevention

**Shengli Bi**

National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention

---

**Research**

**Keywords:** Hepatocellular carcinoma, Chronic hepatitis B infection, Substitution, Deletion, Pre-S, Hepatitis B virus X protein, Pre-Core/Core

**Posted Date:** July 13th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-41025/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Hepatocellular carcinoma (HCC) caused by hepatitis B virus (HBV) infection is a serious public health problem in China. The aims of the present study was to report HCC prevalence in China and characterize the whole gene sequences of HBV derived from patients diagnosed with HCC as well as those with chronic hepatitis B (CHB) infections.

**Methods:** Patients in the HCC group and the CHB group were recruited from national HBV surveillance sites, which were matched by age, gender, and region. All serum samples were tested for serological markers. Polymerase chain reaction (PCR) was used to amplify the HBV complete genome sequences. Then the analysis of the full-length HBV genomes was performed with bioinformatics and statistical software.

**Results:** Serum samples were collected from 51 patients with HBV-related HCC and 76 patients with CHB. All patients were from six provinces (Guangxi, Hebei, Henan, Hunan, Qinghai, and Shanghai) in China. Sequencing and analysis of the full-length HBV genomes revealed the presence of four genotypes (B, C, CD, and I). The distribution of HBV genotypes and the positivity rate of the hepatitis B e antigen differed between the two groups. A total of 148 substitution sites deemed statistically significant were identified between the HCC and CHB groups. In addition, three mutational sites associated with HCC, (F22I/L/P in the pre-S2 region, P33S/T and S144A/T/V in the X region) were identified. Deletions to the pre-S and X regions were found in both HCC and CHB patients. However, deletions to the X region were more common in the HCC group than the CHB group.

**Conclusions:** In this study the hotspot mutations associated with high risk of HCC mostly occurred in the sequences and some substitutions (C1470A/T, T1803A/G, and C1804T) that have not been previously reported. It was implicated that aa33 and aa144 substitution in X region may be new predictive markers for HCC. The results of our study would provide important basic information for HCC prevention.

## Background

Hepatitis B Virus (HBV) is considered to be a major global health problem with more than 250 million chronic HBV (CHB) carriers and more than one million HBV-associated human deaths per year [1]. Hepatocellular carcinoma (HCC) is the most common malignancy of the liver. In 2018, HCC was reported as the fourth leading cause of cancer-related death worldwide, with more than 841,000 new diagnoses and 782,000 deaths globally per year [2]. HCC is a serious worldwide health burden due to the high mortality and low resectability rates [3]. Among all the risk factors, HBV infection is the most important one for HCC, accounting for approximately 80% of all cases [4].

Immunization against HBV was implemented in China 30 years ago and was officially integrated into the China National Immunization Plan in 2002. Nonetheless, there are approximately 93 million people currently infected with HBV in China, including about 28 million chronic HBV cases, that are at increased risk for the development of HCC[5].

Various types of mutations to the HBV genome have been reported, resulting amino acid substitutions during long-term infection, some of which could serve as markers to predict the development of HBV-associated HCC [6, 7], such as mutations in the pre-surface antigen (pre-S) (i.e., deletion to the preS1 [W4P/R] and pre-S2 start codon deletion, pre-core (i.e., nucleotide 1896(G1896A), or a double mutation to the basal core promoter (BCP) and X regions (V5M) [8-10].

In order to investigate the effects of various HBV mutations on the development of HCC, this cases-control study will report current HCC prevalence in representative regions and analyze the complete HBV genomes isolated from HCC and CHB patients.

## Methods

### Field background

Patients with HCC in this study were recruited from six hepatitis B surveillance sites in China (Guangxi, Hebei, Henan, Hunan, Qinghai, and Shanghai), where the incidence of HBV infection were surveillanced since the 1980s. Patients with HCC were diagnosed in provincial hospitals or tumor specialist hospitals. Basic information was recorded on the questionnaire prepared beforehand, including name, gender, birth date, address, ethnicity, immunization history and medical information, and 5 mL of venous blood was taken from each participant. All the HCC subjects were collected between January, 2012 and December, 2015.

### **Serological markers**

All sera samples were tested for three HBV markers (HBsAg, anti-HBe, and HBeAg), as well as anti-HCV antibodies and HDV-IgG. HBV markers were detected with the use of MEIA kits and an ARCHITECT i2000SR immunoassay analyzer (Abbott Diagnostics, Chicago, IL, USA). Other markers were detected using enzyme linked immune-sorbent assay (Beijing Wantai Biological Pharmacy Enterprise Co., Ltd., Beijing, China).

### **Amplification and sequencing of the complete HBV genome from HCC patients and CHB patients**

First, representative HCC patients were selected for whole genome sequencing as case group. CHB patients who were matched case group by age, sex and region were selected as control group. These sera were all positive for Hepatitis B surface antigen (HBsAg). Patients in the CHB group were HBsAg-positive for more than 6 months.

In sum, one hundred and twenty-seven patients (51 in the HCC group and 76 in the CHB group) were selected to obtain complete genome sequences. HBV DNA was extracted from 200 $\mu$ L of sera using a QIAamp DNA blood mini-kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's instructions. The full-length HBV genome was divided into two parts, which were amplified using the nested polymerase chain reaction (PCR) method as we previously described [11]. The whole-genome nucleotide sequences of the 127 HBV isolates reported in this article have been deposited in the National Center for Biotechnology Information GenBank database under accession numbers MT644995 to MT645072.

### **Genotyping of HBV**

HBV genotypes were compared with a set of 88 reference sequences (genotype A–I) retrieved from the GenBank database [12]. Phylogenetic trees were reconstructed by the maximum likelihood (ML) method implemented in IQ-tree package under the GTR+I+G nucleotide substitution model, which was selected by ModelFinder [13, 14]. Support for the ML tree was inferred by bootstrapping with 1000 replicates. To detect recombination in genome sequences of HBV, sequences were investigated using SimPlot v3.5.1 software and JPHMM (jumping profile Hidden Markov Model) method [15, 16].

### **Nucleotide and amino acid mutation analysis**

The substitution proportion of every nucleotide site was calculated for all of the 127 HBV whole genome sequences. First, map the nucleotide site frequency distribution of 127 nucleotides for every site among 3215 nucleotide sites of HBV genome. The nucleotide type with the largest proportion were taken as the prototype, other types of nucleotide in this position were defined as a substitution. If there was a deletion for the nucleotide site, it was not included in the calculation. Second, every nucleotide site substitution proportion was calculated for the HCC group and the CHB group respectively. Subsequently, the sites with the nucleotide substitution rate which was more than 5% in either HCC group or CHB group compared between the two groups. The bar chart for the sites with the significant difference for nucleotide substitution rate was drawn.

In previous reports [17, 18], some amino acid substitutions in the pre-S, pre-C/C, and X region was proved associated with HCC. In present study, all of these nucleotide sites were deduced into corresponding amino acid sites within the pre-S, pre-C/C and X regions. And then the amino acid substitution rate was calculated for the HCC group and the CHB group respectively. At last, the amino acid substitution rate was compared between the two groups.

The sequences with pre-S deletion or X region deletion were aligned with the same genotype reference HBV sequences. All of the Nucleotide sequences of datasets in this study and reference HBV sequences were analyzed using the MEGA 7.0 software.

### **Statistical analysis**

Pearson's  $\chi^2$ -test or Fisher's exact test were used for analysis, as appropriate. All statistical tests were two-tailed and a probability ( $p$ ) value of  $<0.05$  was considered statistically significant. All analyses were using IBM SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, NY, USA).

### **Ethical approval**

The study protocol was approved by the Ethics Committee of the Chinese Center for Disease Control and Prevention and conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. The purpose of the study and the right to information were explained to the participants by the research staff. Written informed consent was obtained from each participant before the interview and venous blood collection.

## **Results**

### **Baseline features of the study population**

In total, there were 3262 HCC patients detected during 2012 to 2015 in six surveillance regions. The people with age  $>40$  have significant higher HCC prevalence than age  $<40$  ( $P=0.000$ ). Of all the six regions, Guangxi, which located in southwest of China, was with the highest HCC prevalence. The detail of HCC patients in six regions were listed in Table 1.

From 2012 to 2015, the average age-specific prevalence rate of HCC in the surveillance regions ranged from 0.07/100000 to 31.29/100000 (Table 1). The prevalence of HCC in population  $< 70$  years old significantly increased with age (Trend  $\chi^2=4451.26$ ,  $P=0.000$ ).

A total of 127 whole genome sequences of HBV DNA were successfully amplified from serum samples of 51 patients with HCC and 76 with CHB. Of these 127 serum samples, no anti-HCV antibody or HDV-IgG was detected and none of these patients was alcoholism.

The gender, patient ages, regions, HBV genotypes, and HBeAg status of the 51 HCC and 76 CHB patients are shown in Table 2. There was no obvious difference in sex, patient ages, and regions between the two groups. The HBeAg-positive rate was significantly higher in the CHB group than in the HCC group. Of the 51 patients in the HCC group, two (3.92%) were positive for the HBV/B genotype, 43 (84.31%) for the HBV/C genotype, and six (11.76%) for other genotypes (HBV/CD, HBV/I). Of the 76 patients in the CHB group, 14 (18.42%) were positive for the HBV/B genotype, 55 (72.37%) for the HBV/C genotype, and seven (9.21%) for other genotypes (HBV/CD, HBV/I). The HBV/C genotype was predominant in both groups. About the proportions of genotypes, there was no significant difference between the two groups ( $p > 0.05$ ; Table 1). The geographical location and genotype distribution were shown in Fig 1.

### **Comparison of HBV DNA nucleotide substitutions between the HCC and CHB groups**

Nucleotide substitutions of the complete genome with frequencies of  $>5\%$  derived from HCC and CHB patients were compared and those with significant differences ( $p < 0.05$ ) are visualized in Fig 2. There were 148 (4.60%) sites with significant differences between the two groups. The rates of 53 substitution sites, which were significantly greater in the HCC group than the CHB group, were mainly located in the C-terminus of HBx (T1802C, T1803A/G, C1804T, G1896A, G1899A, C1969T) and C regions (T2263C, A2269G, T2278C/G, C2281T, T2284A/C, T2287C, etc.) C region (T53A/C, C502A/G, A929T, G1896A, G1899A, etc.), while the rates of 95 substitution sites, which were significantly greater in the CHB group than the HCC group, were mainly located in the pre-S, S, X and P regions.

## Comparative analysis of HCC-related nucleotide and amino acid substitutions

The deduced amino acid corresponding to the nucleotide sites in the pre-S, pre-C/C, and X regions were analyzed. The genotype-specific amino acid substitutions were removed. The amino acid substitution rates with significant differences between the HCC and CHB groups are shown in Table 3. All the nucleotide sites shown in Table 3 were associated to HCC, as previously reported [10, 19, 20]. The substitution rates of three sites (aa22, aa33, and aa144) were greater in the HCC group than the CHB group (47.83% vs. 28.38%, 29.41% vs. 9.21%, and 28.57% vs. 9.46%, respectively). The amino acid substitutions at position aa22 were located in the pre-S2, while those at positions aa33 and aa144 were located in the X region, distributed among different genotypes.

## Association of different pre-S deletion types with HCC

As there were few sequences in the genotype B and CD subgroup, only genotype C were analyzed. Most of the pre-S deletion mutations were in the genotype C subgroup, pre-S1 deletions were more frequent in the CHB group (11.63% vs. 25.45%), while pre-S2 deletions were more frequent in the HCC group (25.58% vs. 20.00%). Epitope mapping revealed frequent deletions in five epitopes. Deletions to the sequences in pS1-B1, pS1-B2, pS2-B2, pS2-B3, and pS1-T1 were more common in the HCC group than the CHB group. Details were shown in Table 4.

Functional mapping revealed that the frequencies of deletions in three functional domains (CBF, NBS, and pHSA) were higher in the HCC group than the CHB group, in contrast to other domains (L start codon, HBS, S promoter, HSC70, CAD, M start codon, and VS start codon) which have no significant differences between two groups. As shown in Table 4, the frequency of HBS deletion was significantly higher in the CHB group than the HCC group ( $p < 0.05$ ). Notably, there were no deletion mutants to the functional domains of the L and HBS start codons in the HCC group.

## Characteristics of X region deletions

The 16.5-kDa HBx protein is composed of 155 amino acids encoded by 465 base pairs (nt 1374-1838). A map of deletions to the HBx protein region of the CHB and HCC groups is shown in Fig. 3. C-terminal deletions are among the most frequently reported mutations to HBx and were detected in four and seven patients in the CHB and HCC groups, respectively. Most patients were genotype C, with the exception of one with genotype CD (the X ORF were in the genotype C fragment). The deletion rate was higher in the HCC group than the CHB group (13.73% vs. 5.26%, respectively), although this difference was not significant. All four deletions of the CHB group were located in the C terminus of HBx (aa 126–135). Meanwhile, there was a larger range of the C-terminus deletions for seven patients in the HCC group, two were truncation to the C-terminus of HBx, and the other four were concentrated at the C-terminus of HBx (aa104–154).

## Discussion

The etiology and pathogenesis of HCC is complex, due to many related risk factors, such as hepatitis B virus (HBV) infection, aflatoxin exposure, as well as physical and chemical factors, especially alcoholism and other unhealthy lifestyle habits. HBV can promote carcinogenesis through chromosomal instability, numerous mutations, and the interaction between HBx protein and host proteins [21]. Besides, the indirect effects of HBV infection include chronic inflammation and oxidative stress, which can subsequently lead to varying degrees of hepatic injury [22]. Chronic hepatitis infection (CHB) is a strong risk factor for the development of HCC mostly due to HBV nucleotide level, HBV genetic mutations, positivity for the hepatitis B e antigen (HBeAg), HBV genotypes, and co-infection with hepatitis C virus [23].

The six sampling and surveillance regions selected in this study were HepB long-term immunization regions. Previous HBV epidemiological surveys proved that Guangxi and Qinghai were in high HBV prevalence; Hunan and Henan were in the middle. Hebei and Shanghai were low HBV endemic regions [24]. So these six surveillance sites represent high, medium and low level of HBV infection regions in China. In this study, the prevalence of HCC was much higher in Guangxi and Hunan than in other regions, and it could be presumed that the higher local HBsAg positivity rate was the main cause. However, Qinghai is also a

highly endemic area of HBV, in which more than 95% of the resident population is Tibetan. It was reported that HBV infection rates of Tibetans were significantly higher than any other region of China [25]. However, in this study, the HCC prevalence is not as high as Guangxi and Hunan. This result may be due to the main prevalent genotype in Qinghai-Tibet plateau were CD recombinant [11]. It is reported that the incidence of HCC in HepB vaccinated population group was significantly lower than that in non-vaccinated population control group, and the risk of HCC was reduced by 84% [26]. Therefore, in the high prevalence areas of HCC HBV infection, especially in the members with family history of hepatitis B, it is very important to surveillance the status and progress of liver disease for early diagnosis of HCC [27].

Integration of the HBV genome is currently believed to be an early event in HBV chronic infection. Notably, mutations and deletions to the HBV genome are associated with an increased risk for the development of HCC and the clinical severity of other hepatic diseases [28]. Most mutations to the HBV genome are generated due to the lack of proofreading capacity of HBV polymerase or host immune pressure [19].

In the present study, HBV complete genome sequences were obtained from samples of 51 HCC patients and 76 CHB patients. To determine whether there was any difference in HBV sequences between the HCC and CHB groups, the nucleotide substitutions of HBV whole genome sequences were compared in four overlapping ORF.

The substitution rates of the nucleotide sites located in the pre-C and C regions were mostly higher in the HCC group than the CHB group. Most pre-C/C mutations are generated during HBeAg seroconversion. Several types of HBV pre-C/C mutations, such as G1896A, A1762T, and G1764A, were reportedly related to disease severity [9, 19, 29]. Also, the HBeAg positive rate was significantly lower in the HCC group than the CHB group. However, it remains unclear whether this phenomenon is associated with the nucleotide substitution to C gene, thus further studies are warranted.

It has been reported that the nucleotide substitution rate in HCC tends to be greater in the X and pre-C/C regions. Most substitution mutations, such as G1653 and G1896, have been associated with an increased risk for the development of HCC [10, 19]. However, the impact of these substitutions on disease severity remains unclear, thus further analysis is needed [10, 19]. In the present study, the substitution rates of many sites were proved greater in the CHB group than the HCC group. Though there were lots of studies about nucleotide of HCC [10, 19]. However, substitutions identified in this study of 29 sites have been rarely reported previously. Nonetheless, the rates of substitutions to 13 of these sites were significantly greater in the HCC group than the CHB group (Fig. 1). Among them, nt1470 and other six sites were located at the B cell epitope. The sites nt1726 and nt1730 were located at the T cell epitope. Previous study reported almost 40% of the integrated HBV genomes were cleaved at approximately nt1800 [10]. Therefore, the sites (nt1799, nt1802, nt1803, and nt1804) may play a potential role in HBV genome integration for HCC development.

In this study, the genotype-specific amino acid substitution rates, as deduced from the nucleotide sites in Figure 1, were compared between the HCC and CHB groups. The mutation rates of F221/L/P in the pre-S2 region, as well as P33S/T and S144A/T/V in the X region, were significantly higher in HCC group. F221/L/P is reportedly associated with immune nonreactivity [30, 31]. P33S/T and S144A/T/V occurring in the X region have not been previously reported and, thus, are novel mutations possibly associated with a greater risk for the development of HCC. The amino acid at position aa33 was located in the negative regulation domain of HBx (aa 1–50) which formed a B cell epitope (aa 29–48). The HBx region partially overlaps with the RNase H part of HBV polymerase at the C-terminus, and also contains several critical cis-elements. Genetic alterations in this region may not only affect the reading frame of HBx, but also the overlapping cis-elements and the possible binding affinities of this protein to its targets [32]. The amino acid at position aa144 is located in the core promoter of the C-terminus of HBx, which plays a key role in controlling cell proliferation, viability, and transformation [33-35].

Deletions to the pre-S and X regions are reportedly associated with the development of HCC [8, 36, 37]. The pre-S1 and pre-S2 regions contain several epitopes of T or B cells and play essential roles in the immune response [38, 39]. Pre-S deletion decreases the expression of the surface proteins of HBV, resulting in intracellular accumulation of HBV envelope proteins and viral particles, which induce endoplasmic reticulum stress and oxidative DNA damage, eventually leading to the development of HCC [40, 41]. Truncated pre-S2/S sequences are often found in HBV DNA integration sites of HCC patients, and truncated

pre-S2/S proteins could specifically activate the MAPK signaling pathway to activate transcription factors such as AP-1 and NF- $\kappa$ B, and thus promote abnormal proliferation of liver cells [42]. In the present study, there was no significant difference in deletions pre-S1 and pre-S2 between HCC and CHB groups, while the frequency of pre-S2 deletions was higher than pre-S1 deletions among HCC patients, which was consistent with the findings of other studies [43-45].

Deletions to the C-terminus of HBx are frequently reported in HCC patients [46, 47]. Deletions or insertions to the C-terminus of HBx reportedly impair transactivation activity, thereby inhibiting cell proliferation, which may contribute to the development of HCC [48]. In the present study, all 11 deletions (ten for genotype C and one for genotype CD) were located in the C-terminus of HBx. In the CHB patients, there were deletions to codons 125 to 136 of HBx, but involved more codons in HCC patients especially truncations to the C-terminus of HBx. Previous studies have frequently reported deletions to the 3'-end of the X gene, which leads to truncations of the HBx C-terminus [36]. Reportedly, truncations of the HBx C-terminus occur in nearly 80% of HCC tissues, which may contribute to hepatocarcinogenesis via loss of pro-apoptotic capability of full genes, activation of cell transformation, and subsequent tumor promotion [49]. In the present study, the HBx deletion rate was 13.73% (7/51) in HCC patients, which is lower than previous reports, but higher than in CHB patients 5.26% (4/76).

There were some limitations in this study should be acknowledged. First, there was no significant difference in most of the substitution rates between the HCC and CHB groups. Second, as large fragments of deletion in the genome of HBV is the characteristic of HCC samples, the chance of obtaining full sequencing maybe insufficient. To better understand the clinical relevance of HBV gene substitutions, further prospective investigations of HBV-infected patients are required.

## Conclusion

In this study, a total of 127 full sequences were isolated from HCC and CHB patients. The hotspot mutations associated with an increased risk of HCC mostly occurred in these sequences and included some substitutions that have not been reported previously. Our results implicate P33S/T and S144A/T/V in X region as new predictive markers for HCC. A larger range of the C-terminus deletions is common occurring HCC patients. For the control and prevention of HCC in China, further managements are required in the high HBV prevalence population to reduce the progression to poor clinical outcomes.

## Abbreviations

HBV: Hepatitis B virus; HCC: Hepatocellular carcinoma; CHB: Chronic hepatitis; HBsAg: Hepatitis B surface antigen; HBeAg: Hepatitis B e antigen; HBx: Hepatitis B x region; Pre-S Pre-surface antigen; PCR: Polymerase chain reaction; BCP: Basal core promote; ORF: Open reading frame; aa: Amino acid; HBS: Hepatocyte binding site; pHSA: Polymerized human serum albumin; NRE: negative regulatory element; CP: Core promoter; C/EBP: CCAAT/enhancing binding protein; EnhII: Enhancer II; MHC: Major histocompatibility complex; HBS:Hepatocyte binding site; HSC70: Heat shock protein 70; CAD: Cytosolic anchorage determinant; NBS: Nucleocapsid binding site; VS: Viral secretion; pHSA: Polymerized human serum albumin; nt: nucleotide

## Declarations

### Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of the Chinese Center for Disease Control and Prevention and conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. The purpose of the study and the right to information were explained to the participants by the research staff. Written informed consent was obtained from each participant before the interview and venous blood collection.

### Availability of data and materials

The datasets generated in the current study are based on complete genome sequences which are uploaded by the author in the National Center for Biotechnology Information GenBank database with accession numbers MT644995 to MT645072.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This study was supported by funds from the Chinese National Science and Technology Major Project (Grant No. 2017ZX10105015-001-002; Recipient: Liping Shen and Grant No. 2018ZX10721-202-002-002; Recipient: Shengli Bi). The funders had no role in study design, data collection and analysis and preparation of the manuscript.

## Author's contributions

ZS designed and carried out the study, analyzed data, and wrote the manuscript. WF and QF carried out the serological experiments, LH and SQ carried out the molecular biology experiments, ZG, WF and ZH received funding and supervised, SL and BS designed study and reviewed manuscript. All authors read and approved the final manuscript.

## Acknowledgments

We thank the staff of the Centers of Disease Control and Prevention of Shanghai, Guangxi province, Qinghai province, Xiangtan city of Hunan Province, and Kaifeng city of Henan province, for their support and collaboration. We also thank the village health care workers for their invaluable assistance in performing this study.

## Authors' information

1NHC Key Laboratory of Medical Virology and Viral Diseases, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changbai Rd 155#, Changping District, Beijing, China.

2National Immunization Program, Chinese Center for Disease Control and Prevention, Nanwei Rd 27#, Xicheng District, Beijing, China.

## References

1. World Health Organization. **Global hepatitis report 2017**. Geneva 2017. Licence: CC BY-NC-SA 3.0 IGO. 2017.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA Cancer J Clin* 2018, **68**:394-424.
3. A O: **Global epidemiology of Hepatocellular carcinoma**. *J Gastrointest Canc* 2017, **48**:238-240.
4. Petruzzello A: **Epidemiology of Hepatitis B Virus (HBV) and Hepatitis C Virus (HCV) Related Hepatocellular Carcinoma**. *Open Virol J* 2018, **12**:26-32.
5. Zhang S, Ma Q, Liang S, Xiao H, Zhuang G, Zou Y, Tan H, Liu J, Zhang Y, Zhang L: **Annual economic burden of hepatitis B virus-related diseases among hospitalized patients in twelve cities in China**. *Journal of Viral Hepatitis* 2016, **23**:202-210.
6. Park YM: **Clinical utility of complex mutations in the core promoter and proximal precore regions of the hepatitis B virus genome**. *World J Hepatol* 2015, **7**:113-120.
7. Kuang SY, Jackson PE, Wang JB, Lu PX, Munoz A, Qian GS, Kensler TW, Groopman JD: **Specific mutations of hepatitis B virus in plasma predict liver cancer development**. *Proc Natl Acad Sci U S A* 2004, **101**:3575-3580.
8. Chen BF: **Different pre-S deletion patterns and their association with hepatitis B virus genotypes**. *World J Gastroenterol* 2016, **22**:8041-8049.

9. Al-Qahtani AA, Al-Anazi MR, Nazir N, Abdo AA, Sanai FM, Al-Hamoudi WK, Alswat KA, Al-Ashgar HI, Khan MQ, Albenmoussa A, et al: **The Correlation Between Hepatitis B Virus Precore/Core Mutations and the Progression of Severe Liver Disease.***Front Cell Infect Microbiol* 2018, **8**:355.
10. Kim H, Lee SA, Kim BJ: **X region mutations of hepatitis B virus related to clinical severity.***World J Gastroenterol* 2016, **22**:5467-5478.
11. Liu H, Shen L, Zhang S, Wang F, Zhang G, Yin Z, Qiu F, Liang X, Wang F, Bi S: **Complete genome analysis of hepatitis B virus in Qinghai-Tibet plateau: the geographical distribution, genetic diversity, and co-existence of HBsAg and anti-HBs antibodies.***Viol J* 2020, **17**:75.
12. Kramvis A: **Genotypes and genetic variability of hepatitis B virus.***Intervirology* 2014, **57**:141-150.
13. Lam-Tung N, Schmidt HA, Arndt VH, Bui Quang M: **IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.***Mol Biol Evol* 2015, **32**:268-274.
14. Kalyanamoorthy S, Minh BQ, Tkf W, Von HA, Jermiin LS: **ModelFinder: fast model selection for accurate phylogenetic estimates.***Nat Methods* 2017, **14**.
15. Lole K, Bollinger R, Rs, Gadkari D, Kulkarni S, Novak N, Ingersoll R, Sheppard H, Ray S: **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.***J Virol* 1999, **73**:152-160.
16. Anne-Kathrin S, Ingo B, Mariama AC, Emmanuel G, Burkhard M, Fabien Z, Paul D, Mario S: **jpHMM: recombination analysis in viruses with circular genomes such as the hepatitis B virus.***Nucleic Acids Res* 2012, **40**:193-198.
17. Jiao F, Long L, Ding S, Xie X, Jia L, Lu F: **Complete genome sequencing and clinical analysis of intrahepatic hepatitis B virus cccDNA from HCC.***Microb Pathog* 2017, **109**:49-55.
18. Tan TL, Chen WN: **A proteomics analysis of cellular proteins associated with HBV genotype-specific HBX: potential in identification of early diagnostic markers for HCC.***J Clin Virol* 2005, **33**:293-298.
19. Kim H, Lee SA, Do SY, Kim BJ: **Precore/core region mutations of hepatitis B virus related to clinical severity.***World J Gastroenterol* 2016, **22**:4287-4296.
20. Gil-Garcia AI, Madejon A, Francisco-Recuero I, Lopez-Lopez A, Villafranca E, Romero M, Garcia A, Oliveira A, Mena R, Larrubia JR, Garcia-Samaniego J: **Prevalence of hepatocarcinoma-related hepatitis B virus mutants in patients in grey zone of treatment.***World J Gastroenterol* 2019, **25**:5883-5896.
21. Aoki H, Kajino K, Arakawa Y, Hino O: **Molecular Cloning of a Rat Chromosome Putative Recombinogenic Sequence Homologous to the Hepatitis B Virus Encapsidation Signal.***Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**:7300.
22. Xu W, Yu J, Wong VW: **Mechanism and prediction of HCC development in HBV infection.***Best Pract Res Clin Gastroenterol* 2017, **31**:291-298.
23. Wong VW, Janssen HL: **Can we use HCC risk scores to individualize surveillance in chronic hepatitis B infection?***J Hepatol* 2015, **63**:722-732.
24. Wang F, Shen L, Cui F, Zhang S, Zheng H, Zhang Y, Liang X, Wang F, Bi S: **The long-term efficacy, 13-23 years, of a plasma-derived hepatitis B vaccine in highly endemic areas in China.***Vaccine* 2015, **33**:2704-2709.
25. Shen L, Yin W, Zheng H, Cui F, Bi S: **Molecular Epidemiological Study of Hepatitis B Virus Genotypes in Southwest, China.***Journal of Medical Virology* 2014, **86**:1307-1313.
26. Qu C, Chen T, Fan C, Zhan Q, Wang Y, Lu J, Lu L, Ni Z, Huang F, Yao H: **Efficacy of Neonatal HBV Vaccination on Liver Cancer and Other Liver Diseases over 30-Year Follow-up of the Qidong Hepatitis B Intervention Study: A Cluster Randomized Controlled Trial.***Plos Medicine* 2014, **11**:e1001774.
27. Zhang XF, Liu XM, Wei T, Liu C, Li MX, Long ZD, Lv Y: **Clinical characteristics and outcome of hepatocellular carcinoma in children and adolescents.***Pediatric Surgery International* 2013, **29**:763-770.

28. Li WW, Goto K, Matsubara Y, Ito S, Muroyama R, Li Q, Kato N: **The Characteristic Changes in Hepatitis B Virus X Region for Hepatocellular Carcinoma: A Comprehensive Analysis Based on Global Data.***Plos One* 2015, **10**.
29. Wang XY, Harrison TJ, Chen QY, Hai L, Li GJ, Liu MH, Hu LP, Chao T, Yang QL, Fang ZL: **The incidence rate over 10 years of naturally occurring, cancer related mutations in the basal core promoter of hepatitis B virus.***Infection Genetics & Evolution Journal of Molecular Epidemiology & Evolutionary Genetics in Infectious Diseases* 2015, **34**:75-80.
30. Qu LS, Liu JX, Liu TT, Shen XZ, Chen TY, Ni ZP, Lu CH: **Association of hepatitis B virus pre-S deletions with the development of hepatocellular carcinoma in Qidong, China.***PLoS One* 2014, **9**:e98257.
31. Wang T, Dai Y, Zhang M, Cui D, Xu X, Sun C, Cheng J: **Sequence analysis of the Pre-S gene in chronic asymptomatic HBV carriers with low-level HBsAg.***Int J Mol Med* 2018, **42**:2689-2699.
32. Jegal M-E, Jung S-Y, Han Y-S, Kim Y-J: **C-terminal truncated HBx reduces doxorubicin cytotoxicity via ABCB1 upregulation in Huh-7 hepatocellular carcinoma cells.***BMB Reports* 2019, **52**:330-335.
33. Brechot C: **Pathogenesis of hepatitis B virus-related hepatocellular carcinoma: old and new paradigms.***Gastroenterology* 2004, **127**:S56-61.
34. Feitelson MA, Bonamassa B, Arzumanyan A: **The roles of hepatitis B virus-encoded X protein in virus replication and the pathogenesis of chronic liver disease.***Expert Opin Ther Targets* 2014, **18**:293-306.
35. Zhang H, Shan CL, Li N, Zhang X, Zhang XZ, Xu FQ, Zhang S, Qiu LY, Ye LH, Zhang XD: **Identification of a natural mutant of HBV X protein truncated 27 amino acids at the COOH terminal and its effect on liver cell proliferation.***Acta Pharmacol Sin* 2008, **29**:473-480.
36. Lin CL, Kao JH: **Natural history of acute and chronic hepatitis B: The role of HBV genotypes and mutants.***Best Pract Res Clin Gastroenterol* 2017, **31**:249-255.
37. Wang M, Xi D, Ning Q: **Virus-induced hepatocellular carcinoma with special emphasis on HBV.***Hepatol Int* 2017, **11**:171-180.
38. Peeridogaheh H, Meshkat Z, Habibzadeh S, Arzanlou M, Shahi JM, Rostami S, Gerayli S, Teimourpour R: **Current concepts on immunopathogenesis of hepatitis B virus infection.***Virus Research* 2018, **245**:29-43.
39. Ferrari C, Cavalli A, Penna A, Valli A, Bertoletti A, Pedretti G, Pilli M, Vitali P, Neri TM, Giuberti T, Fiaccadori F: **Fine Specificity of the Human T-Cell Response to the Hepatitis-B Virus Pres1 Antigen.***Gastroenterology* 1992, **103**:255-263.
40. Fan YF, Lu CC, Chen WC, Yao WJ, Wang HC, Chang TT, Lei HY, Shiau AL, Su IJ: **Prevalence and significance of hepatitis B virus (HBV) pre-S mutants in serum and liver at different replicative stages of chronic HBV infection.***Hepatology* 2001, **33**:277-286.
41. Chen BF, Liu CJ, Jow GM, Chen PJ, Kao JH, Chen DS: **High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases.***Gastroenterology* 2006, **130**:1153-1168.
42. Levrero M, Zucman-Rossi J: **Mechanisms of HBV-induced hepatocellular carcinoma.***Journal of Hepatology* 2016, **64**:S84-S101.
43. Wang C, Teng ZW, Zhu Y, Zhao AZ, Sun CH: **Associations between Pre-S Deletion Mutation of Hepatitis B Virus and Risk of Hepatocellular Carcinoma in the Asian Population: A Meta-Analysis.***Medical Science Monitor* 2015, **21**:1072-1077.
44. Zhao ZM, Jin Y, Gan Y, Zhu Y, Chen TY, Wang JB, Sun Y, Cao ZG, Qian GS, Tu H: **Novel approach to identifying the hepatitis B virus pre-S deletions associated with hepatocellular carcinoma.***World Journal of Gastroenterology* 2014, **20**:13573-13581.
45. Lin CM, Wang GM, Jow GM, Chen BF: **Functional analysis of hepatitis B virus pre-s deletion variants associated with hepatocellular carcinoma.***Journal of Biomedical Science* 2012, **19**.
46. Kim H, Jee Y, Mun HS, Park JH, Yoon JH, Kim YJ, Lee HS, Hyun JW, Hwang ES, Cha CY, et al: **Characterization of two hepatitis B virus populations in a single Korean hepatocellular carcinoma patient with an HBeAg-negative serostatus: A novel X-gene-deleted strain with inverted duplication sequences of upstream enhancer site II.***Intervirology* 2007, **50**:273-280.

47. Chen GG, Li MY, Ho RLK, Chak ECW, Lau WY, Lai PBS: **Identification of hepatitis B virus X gene mutation in Hong Kong patients with hepatocellular carcinoma.***Journal of Clinical Virology* 2005, **34**:7-12.
48. Ali A, Abdel-Hafiz H, Suhail M, Al-Mars A, Zakaria MK, Fatima K, Ahmad S, Azhar E, Chaudhary A, Qadri I: **Hepatitis B virus, HBx mutants and their role in hepatocellular carcinoma.***World Journal of Gastroenterology* 2014, **20**:10238-10248.
49. Gounder PP, Bulkow LR, Snowball M, Negus S, Spradling PR, McMahon BJ: **Hepatocellular Carcinoma Risk in Alaska Native Children and Young Adults with Hepatitis B Virus: Retrospective Cohort Analysis.***Journal of Pediatrics* 2016, **178**:206-213.

## Tables

**Table 1:** Distribution of the prevalence rates (per 100,000 populations) of HCC by age in pilot regions, 2012-2015

Age (years)	No.( per 100,000)						
	Guangxi	Hebei	Henan	Hunan	Qinghai	Shanghai	Sum
0-19	2 (0.60)	0 (0)	1 (0.02)	3 (0.14)	0 (0)	0 (0)	6 (0.07)
20-	9 (5.90)	0 (0)	6 (0.18)	7 (0.36)	7 (2.27)	1 (0.18)	30 (0.45)
30-	76 (46.92)	2 (0.96)	18 (0.87)	46 (3.46)	19 (5.98)	9 (2.47)	170 (3.82)
40-	235 (103.18)	33 (11.05)	130 (4.73)	177 (7.25)	45 (14.30)	19 (4.84)	639 (9.95)
50-	226 (154.28)	51 (20.05)	187 (8.86)	203 (15.14)	62 (49.18)	83 (13.87)	812 (17.74)
60-	254 (242.70)	47 (30.76)	181 (13.26)	256 (22.93)	78 (89.95)	60 (21.70)	876 (28.24)
70-	268 (269.35)	14 (15.52)	107 (11.69)	187 (21.92)	44 (71.43)	109 (35.15)	729 (31.29)
Sum	1070 (87.30)	147 (8.00)	630 (3.65)	879 (7.84)	255 (12.20)	281 (10.06)	3262 (8.95)

**Table 2:** Comparisons of baseline features between the HCC and CHB groups

Character		HCC (51 isolates) No. & %	CHB (76 isolates) No. & %	<i>P</i>
Sex	Male	41 (80.39)	59 (77.63)	0.709
	Female	10 (19.61)	17 (22.37)	
Age group (years)	40~50	12 (23.53)	19 (25.00)	0.672
	51~60	17 (33.33)	30 (39.47)	
	>60	22 (43.14)	27 (35.53)	
Region	Guangxi	18 (35.29)	24 (31.58)	0.302
	Hebei	10 (19.61)	15 (19.74)	
	Henan	8 (15.69)	20 (26.32)	
	Hunan	3 (5.88)	9 (11.84)	
	Qinghai	9 (17.65)	6 (7.89)	
	Shanghai	3 (5.88)	2 (2.63)	
Genotype	B	2 (3.92)	14 (18.42)	0.054
	C	43 (84.31)	55 (72.37)	
	Others	6 (11.76)	7 (9.21)	
HBeAg	Positive	14 (27.45)	37 (48.68)	0.017
	Negative	37 (72.55)	39 (51.32)	

CHB: Chronic hepatitis B; HCC: Hepatocellular carcinoma.

**Table 3:** HBV substitutions related to HCC throughout the complete genomes of the HCC and CHB groups

Region	Substitution		HBV genotype	Functional region affected	Substitution rate of aa n & %		P
	Nucleotide	Amino acid (aa)			HCC	CHB	
pre-S1	G3009C/A	E54A/Q	C, CD, I	Immunoinactive	8(15.69)	15(19.74)	0.595
pre-S2	A7C/G/T	T6S/A/P/K	B, C, CD	pHSA	12(24.49)	20(28.57)	0.621
	T53A/C	F22I/L/P	B, C, CD	Immunoinactive	22(47.83)	21(28.38)	0.031
X	A1383C/G	R4G	C	miRNA binding site	0(0)	1(1.32)	1.000#
	C1470A/T	P33S/T	C, CD, I	B-cell epitope	15(29.41)	7(9.21)	0.003
	T1500C/G	S43A/P	B, C	B-cell epitope	3(5.88)	12(15.79)	0.090
	C1504T,C1505T	A44I/T/V	B, C	B-cell epitope	3(5.88)	10(13.16)	0.185
	G1613A	80 NC	/	NRE, CP	/	/	/
	C1632A/G/T, A1633G/T	Q87G/L/R/W	B, C, CD	NRE, CP	6(11.76)	12(16.00)	0.505
	C1653T	H94R	B, C, CD, I	Box $\alpha$ , C/EBP, CP, EnhII	12(23.53)	16(21.33)	0.771
	C1673G/T	100 NC		EnhII, CP	/	/	/
	A1703G/T	110 NC		EnhII, CP	/	/	/
	G1764/T	I131F/V	B, C, CD	BH3-like motif, EnhII, CP	11(21.57)	15(20.83)	0.922
	T1803A/G,C1804T	S144A/T/V	B, C, CD	CP	14(28.57)	7(9.46)	0.006
pre-C	T1858A/C	P15S/L	C, CD	T-cell epitope	5(10.00)	4(5.26)	0.312
	G1896A	W28*	B, C, CD, I	Expression e antigen	19(37.25)	30(39.47)	0.801
	G1899A	G29D/S	B, C, CD	Expression e antigen	7(14.00)	14(19.44)	0.433
C	T2044C	C48V	C	Immunoinactive	0(0)	4(5.33)	0.146
	G2251A	117NC	/	T-cell epitope	/	/	/
	T2278C/G	I126M	C	T-cell epitope	0(0)	1(1.33)	1.000#
	C2289A/T	P130A/I/L/N/Q/S/T	B, C, CD, I	T-cell epitope	16(31.37)	22(29.33)	0.807
	T2441C	S181P/H	B, C, CD	Immunoinactive	4(7.84)	7(9.21)	0.788

\*Termination condon; # Fisher's exact test;

Abbreviations: C/EBP, CCAAT/enhancing binding protein; CHB, Chronic hepatitis B; CP, core promoter, region (nt1613-1849); EnhII, Enhancer II; HCC, Hepatocellular carcinoma; miRNA binding site, MicroRNA binding site, region (nt1362-1383); NC, no change; NRE, negative regulatory element, region (nt1611-1634); pHSA: Polymerized human serum albumin;

B-cell epitope, region (aa positions 29-48); Box  $\alpha$ , region (nt1646-1668); BH3-like motif, region (nt1719-1769); T-cell epitope, region (nt1719-1754);

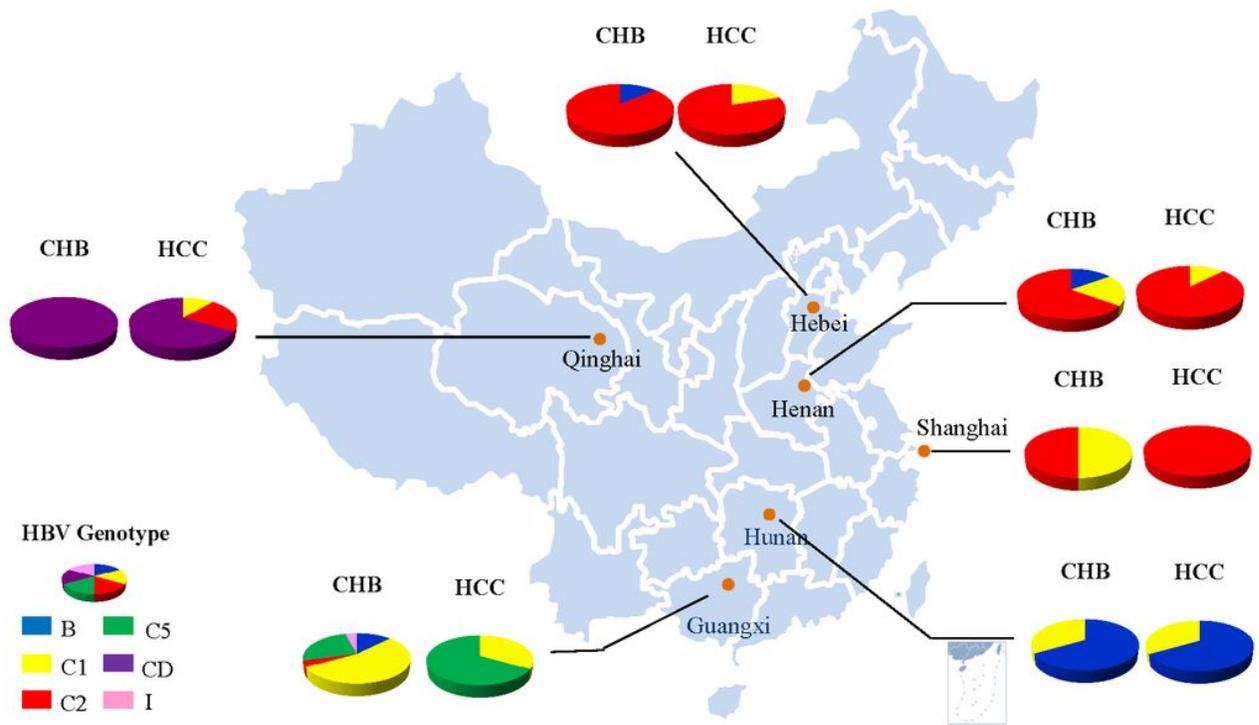
**Table 4:** Frequencies of genotype C sequences with different Pre-S deletion types in the HCC and CHB groups

Section of Pre-S/Group	HCC (43 isolates) No. & %	CHB (55 isolates) No. & %	<i>P</i>
Pre-S1 (aa 1-119)	5(11.63)	14(25.45)	0.086
Pre-S2 (aa 1-55)	11(25.58)	11(20.00)	0.511
Epitope Mapping			
pS1-B1 (aa 12-53 of pre-S1)	1(2.33)	0(0)	0.439*
pS1-B2 (aa 72-78 of pre-S1)	2(4.65)	2(3.64)	1.000*
pS1-BT (aa 94-117 of pre-S1)	3(6.98)	5(9.09)	1.000*
pS2-B1 (aa 1-6 of pre-S2)	3(6.98)	7(12.73)	0.505*
pS2-B2 (aa 3-15 of pre-S2)	11(25.58)	9(16.36)	0.261
pS2-B3 (aa 13-24 of pre-S2)	9(20.93)	7(12.73)	0.276
pS2-T1 (aa 21-48 of pre-S2)	9(20.93)	7(12.73)	0.276
Functional mapping			
L start condon	0(0)	5(9.09)	0.065
HBS (aa 2-48)	0(0)	6(10.91)	0.033*
S promoter (nt 3045-3189)	4(9.30)	5(9.09)	1.000*
CBF (nt 3137-3147)	2(4.65)	2(3.64)	1.000*
HSC70 ( aa 74-118)	4(9.30)	7(12.73)	0.751*
CAD (aa 81-105)	3(6.98)	5(9.09)	1.000*
NBS (aa 103-127)	7(16.28)	8(14.55)	0.813
M start condon	1(2.33)	7(12.73)	0.075*
VS (aa 1-5 of pre-S2)	3(6.98)	7(12.73)	0.505*
pHSA (aa 3-16 of pre-S2)	11(25.58)	10(18.18)	0.376

\* means Fisher's exact test.

Abbreviations: CAD, Cytosolic anchorage determinant; CBF, CCAAT binding factor; CHB, Chronic hepatitis B; HBS, Hepatocyte binding site; HCC, Hepatocellular carcinoma; HSC 70, Heat shock protein 70; NBS, Nucleocapsid binding site; pHSA, Polymerized human serum albumin; VS, Viral secretion.

## Figures



**Figure 1**

Distribution of HBV genotypes in six HCC study regions of China. The proportion of HBV recombinant CD and genotype B, C and I in each region is shown. The HBV prevalence rate is shown by the color of region in the map.

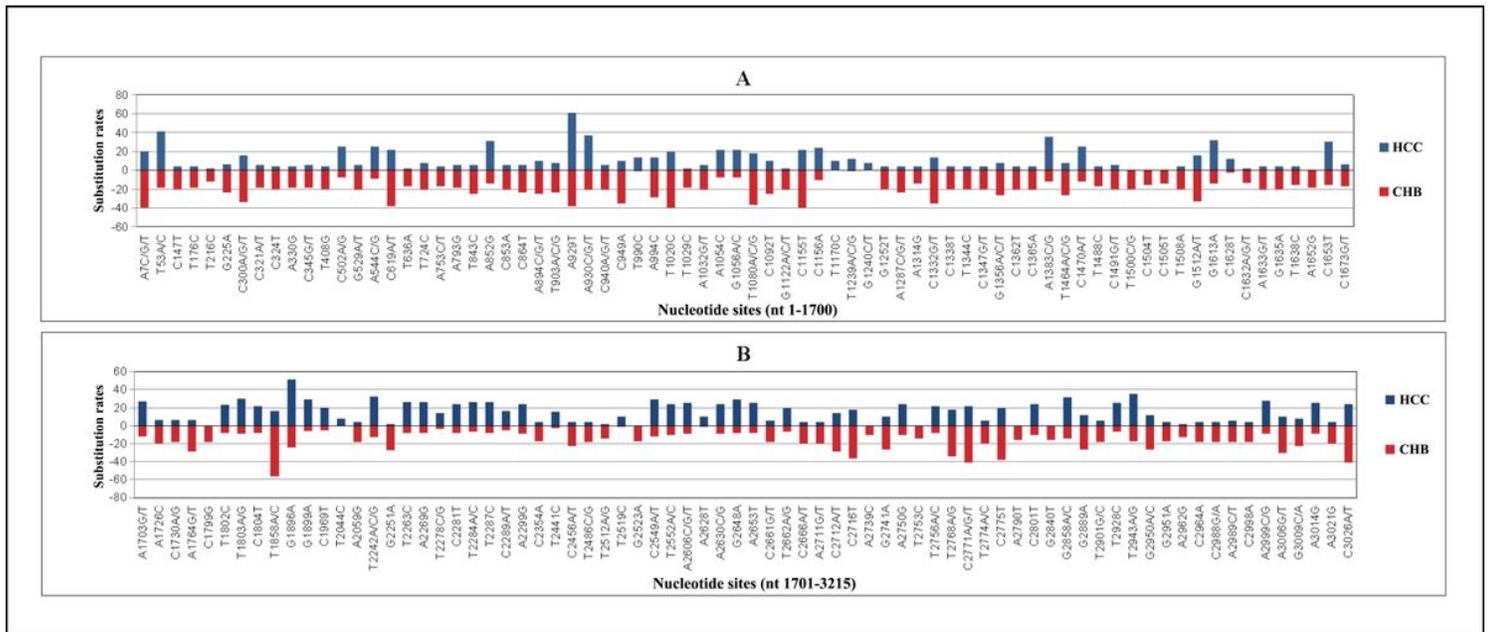


Figure 2

Distribution of nucleotide substitutions in HBV genome. Each bar represents the substitution rates in the nucleotide position.

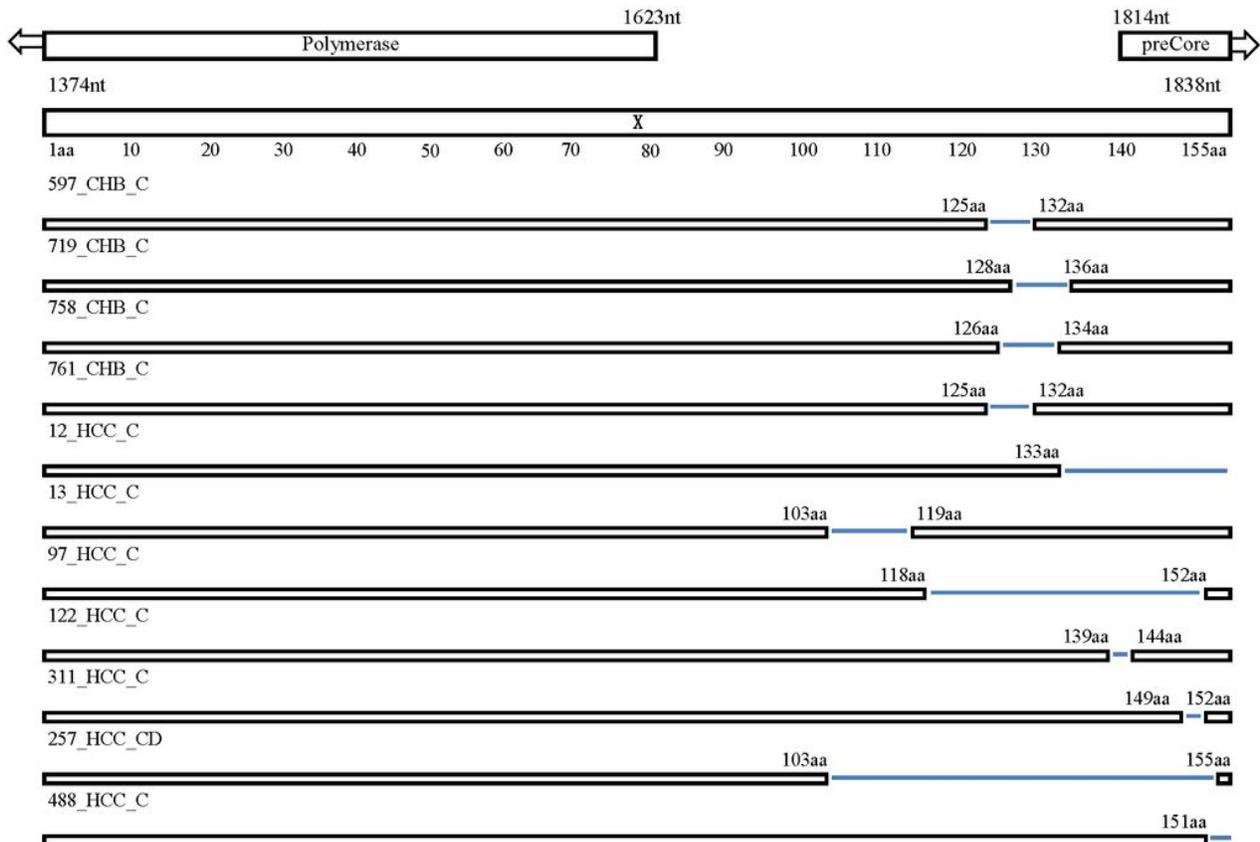


Figure 3

Visualization of deletions within the HBx protein. The deletion between nt1374-nt1838 and correspondent amino acid positions in four CHB and seven HCC genomes are labeled with blue dash.