Coates edited

**HCC-Associated Single Nucleotide Variants and Deletions Identified Using
Genome-Wide High Throughput Analysis of HBV**

Wen-Chun Liu^{1,2}, I-Chin Wu^{1,2}, Yen-Chien Lee^{3,4}, Chih-Peng Lin⁵, Ji-Hong Cheng⁶,
Yih-Jyh Lin⁷, Chia-Jui Yen^{1,2}, Pin-Nan Cheng^{1,2}, Pei-Fu Li⁸, Yi-Ting Cheng⁸, Pei-Wen
Cheng⁹, Koun-Tem Sun⁹, Shu-Ling Yan¹, Jia-Jhen Lin¹, Jui-Chu Yang¹⁰, Kung-Chao
Chang¹⁰, Cheng-Hsun Ho^{1,2}, Vincent S. Tseng¹¹, Bill Chia-Han Chang⁵, Jaw-Ching
Wu^{12,13}, Ting-Tsung Chang^{1,2}

¹ *Department of Internal Medicine, National Cheng Kung University Hospital,
College of Medicine, National Cheng Kung University, Tainan, Taiwan, ROC*

² *Infectious Disease and Signaling Research Center, National Cheng Kung University,
Tainan, Taiwan, ROC*

³ *Department of Oncology, Tainan Hospital, Ministry of Health and Welfare, Tainan,
Taiwan, ROC*

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/path.4938](https://doi.org/10.1002/path.4938)

⁴ *Institute of Clinical Medicine, College of Medicine, National Cheng Kung University,*

Tainan, Taiwan, ROC

⁵ *Yourgene Bioscience, Taipei, Taiwan, ROC*

⁶ *Department of Computer Science and Information Engineering, National Cheng*

Kung University, Tainan, Taiwan, ROC

⁷ *Department of Surgery, National Cheng Kung University College of Medicine and*

Hospital, Tainan, Taiwan, ROC

⁸ *Institute of Medical Informatics, National Cheng Kung University, Tainan, Taiwan,*

ROC

⁹ *Department of Information and Learning Technology, Science and Engineering*

College, National University of Tainan, Tainan, Taiwan, ROC

¹⁰ *Human Biobank, Research Center of Clinical Medicine, National Cheng Kung*

University Hospital, Tainan, Taiwan, ROC

¹¹ *Department of Computer Science, National Chiao Tung University, Hsinchu,*

Taiwan, ROC

¹² *Institute of Clinical Medicine, School of Medicine, National Yang-Ming University,*

Taipei, Taiwan, ROC

¹³ *Translational Research Division, Medical Research Department, Taipei Veterans*

General Hospital, Taipei, Taiwan, ROC

Correspondence to: *Ting-Tsung Chang, Department of Internal Medicine, National*

Cheng Kung University Hospital, College of Medicine, National Cheng Kung

University, 138 Sheng-Li Road, Tainan 70403, Taiwan, ROC. E-mail:

ttchang@mail.ncku.edu.tw and Jaw-Ching Wu, Translational Research Division,

Medical Research Department, Taipei Veterans General Hospital and Institute of

Clinical Medicine, National Yang-Ming University 201 Shih-Pai Road, Sec. 2, Taipei

112, Taiwan. E-mail: jcwu@vghtpe.gov.tw

Running title: HBV Variants and Deletions Associated with HCC

Keywords: Deletion index; Hepatocarcinogenesis; Meta-analysis; Next-generation

sequencing; U-shaped distribution

Conflict of interest

The authors declare that there are no conflicts of interest.

Author Manuscript

Abstract

This study investigated hepatitis B virus (HBV) single nucleotide variants (SNVs) and deletion mutations linked with hepatocellular carcinoma (HCC). Ninety-three HCC patients and 108 non-HCC patients were enrolled for HBV genome-wide next-generation sequencing (NGS) analysis. Systemic literature review and meta-analysis were performed to validate NGS-defined HCC-associated SNVs and deletions. The experimental results identified 60 NGS-defined HCC-associated SNVs, including 41 novel SNVs, and their pathogenic frequencies. Each SNV was specific for either genotype B ($n=24$) or genotype C ($n=34$), except nt53C for both genotypes. The pathogenic frequencies of these HCC-associated SNVs exhibited a distinct U-shaped distribution pattern. According to meta-analysis and literature review, 167 HBV variants from 109 publications were categorized into 4 levels (A–D) of supporting evidence associated with HCC. The proportion of NGS-defined HCC-associated SNVs among these HBV variants declined significantly from 75% of 12 HCC-associated variants by meta-analysis (Level A) to 0% of 10 HCC-unassociated variants by meta-analysis (Level D) ($P<0.0001$). PreS deletions were significantly associated with HCC, in terms of deletion index, for both

genotypes B ($P=0.030$) and C ($P=0.049$). For genotype C, preS deletions involving a specific fragment (nt2977–3013) were significantly associated with HCC (HCC vs non-HCC, 6/34 vs 0/32, $P=0.025$). Meta-analysis of preS deletions showed significant association with HCC (summary odds ratio: 3.0, 95% confidence interval, 2.3 to 3.9). Through transfecting Huh7 cells, all the five novel NGS-defined HCC-associated SNVs in the small surface region influenced hepatocarcinogenesis pathways, including ER-stress and DNA repair systems, by microarray, real-time PCR and western blot analysis. Their carcinogenic mechanisms are worthy of further research.

Introduction

Patients with chronic hepatitis B virus (HBV) are at risk of developing liver cirrhosis and hepatocellular carcinoma (HCC) [1-3]. The HBV genome presents a highly compact organization with four overlapping open reading frames (ORFs) and regulatory elements [4]. Ten HBV genotypes (A–J) have been identified, with genotypes B and C occurring mainly in Asia, while A and D occur mainly in Africa and Europe [5].

Owing to the lack of proofreading activity of HBV reverse transcriptase, the composition of the viral quasispecies evolves over time depending on selective pressure, including host immune response. The characteristics of viral quasispecies have been implicated in the exacerbation of chronic hepatitis B (CHB) and the development of liver cancer [6]. Although numerous HBV variants have been studied as to whether they are genuine risk factors for HCC development, uncertain or controversial results remain [7,8].

Next-generation sequencing (NGS) is a high-throughput analysis for characterizing genetic diversity in viral strains and can characterize minor strains among viral quasispecies in a way that is not possible by direct sequencing or cloning. This study

seeks HCC-associated HBV single nucleotide variants (SNVs) with their pathogenic frequencies using NGS analysis of whole HBV genomes, and compares the findings with systemic literature review and meta-analysis.

Materials and Methods (see supplementary file for details)

Patient enrollment

Patients were retrospectively selected from databases of 475 chronic HBV infected patients with HCC and 1036 chronic HBV infected patients without HCC in National Cheng Kung University Hospital and Taipei Veterans General Hospital. These patients were enrolled with negativity of anti-hepatitis C virus antibodies, anti-hepatitis D antibodies, and anti-human immunodeficiency virus antibodies. All serum samples analyzed were obtained before any antiviral treatment and had HBV viral loads more than 10,000 IU/mL, successful HBV genome PCR amplification and successful NGS. For HCC patients, the serum samples were collected when HCC was diagnosed for the first time. For genotypes B and C, patients in the HCC and non-HCC groups were matched for age and sex. Finally, 93 HCC patients and 108 non-HCC patients were enrolled in the NGS analysis.

Next-generation sequencing

The HBV full genome was amplified by PCR with nine primer sets (Table S1). NGS was performed using the Illumina HiSeq™ 2500. The sample-specific reference sequence was utilized as a mapping reference sequence for NGS datasets [9-12]. The HBV genotypes were determined by real-time PCR [13] and phylogenetic sequence analysis [14]. PreS deletion mutants were classified into types I-V. PreS deletion mutants involving functional sites, *S*, *T*, *N*, *M* and *P* were analyzed [15].

Systemic literature review and meta-analysis

The MEDLINE (1966 to March 10, 2015) and EMBASE (1950 to December 2014) databases were searched using keywords “hepatitis B virus”, “mutation” and “hepatocellular carcinoma”. A random-effects model was applied to obtain summary odds ratios. Publication bias was measured using Begg’s funnel plots and the Egger’s test. Meta-analysis was only performed for HBV variants studied in at least 4 articles.

Construction of plasmid DNA containing HBV DNA and transfection into Huh7

cells

The plasmid pEGFP-N1-LHBS-genoB/C encoding the wild-type large HBV surface protein (LHBS) was used as a template for mutagenesis to create mutations at small surface region. These plasmids were transfected into Huh7 cells.

Real-time PCR

Total RNA was extracted from Huh7 cells 72 hours after transfection. Real-time PCR was performed using primers specific for *APE1*, *GRP78*, and *GAPDH* genes with ABI 7500 Fast Real-Time PCR System using a SYBR Green method.

Microarray analysis and biological interpretation

Gene expression data from transfected Huh7 cells were analyzed by Agilent SurePrint Microarray. The complete dataset is available on the Gene Expression Omnibus database (Accession Number GSE87804). Normalized Microarray results were imported into Ingenuity Pathway Analysis (IPA) software. The major molecular and cellular functions, physiological system developmental functions, and disease and disorder development by these LHBS variants were investigated [16].

Western Blot Analysis

Proteins were extracted from transfected Huh7 cells and proteins related to ER stress and DNA repair were analyzed by western blotting.

Statistical analysis

Continuous variables were compared by Student's t test or Mann-Whitney test.

Categorical variables were compared by Chi-square test or Fisher's exact test. The

SNV frequency ($f_{m/quasi}$) is the frequency of a specific substitution within one subject

($f_{m/quasi}$ = number of reads with a specific substitution / the number of total reads

covering that site in one sample) [17]. The odds ratio (OR) of a specific HBV SNV

for HCC is defined as: $OR = (HCC_E \times non-HCC_{NE}) / (HCC_{NE} \times non-HCC_E)$. The

subscript E indicates the number of patients with SNV frequencies not less than the

cut-off value, and the subscript NE indicates the number of patients with SNV

frequencies less than the cut-off value. For each specific SNV, the cut-off value of

SNV frequency which corresponds to the highest OR was used for statistical analysis.

The P values of these OR were calculated with Chi-square test. With respect to each

genotype, HCC-associated SNVs discovered by NGS analysis were determined after

adjustment for multiple testing by Benjamini and Hochberg's false discovery rate (FDR) method [18] for those HCC-associated variants evidenced by meta-analysis and those SNVs with P values of $OR < 0.05$ by NGS analysis.

Results

Clinical characteristics of patients

Table S3 lists the clinical characteristics of 93 HCC patients and 108 non-HCC patients enrolled in NGS analysis. The distributions of gender, age, AST, ALT, creatinine and HBeAg status were similar between the HCC and non-HCC groups for both genotypes B and C.

HBV genome-wide characterization of substitutions

The mapping quality of NGS was good, as indicated in Table S4. Figure 1 and Figure S2 display the genome-wide characterization of all genotypes B and C HBV substitutions identified in each patient. Complex quasi-species compositions were observed in both HCC and non-HCC groups. SNV I is defined as the dominant variant of HBV in the non-HCC group. SNV II is defined as the variant other than

SNV I at the same nucleotide position. The SNV II frequency in each patient is shown in a heatmap [19].

HCC-associated SNVs discovered by NGS analysis

For genotype B HBV, SNVs located at 23 nucleotide sites were significantly associated with HCC. These SNVs included pre-core (preC) mutations (1896A and 1899A). Nucleotide sites 273A and 2227T were protective factors for HCC, whereas 273G and 2227G were risk factors for HCC. All the other 21 SNVs were risk factors for HCC; of these, 6 were SNV I, and 15 were SNV II (Table 1A). Among the 25 SNVs, 17 were missense mutations which changed the amino acids in the polymerase, preS2, surface and preC/core regions (Table S5A.). Seven of the 17 missense mutations, and 4 of 8 silent mutations, involved the regulatory elements including CpG island I/II/III, X promoter, Enhancer (Enh) I, epsilon loop (μ loop) and basal core promoter (BCP) (Figure S2B).

For genotype C HBV, 35 SNVs located at 35 nucleotide sites were significantly associated with HCC. These SNVs included BCP mutation (1764A and 1653T). All the 35 SNVs were risk factors for HCC, while 18 of them were SNV II (Table 1B).

Among the 35 SNVs, 28 were missense mutations located at 4 ORFs, particularly at the preS1 region and the spacer domain of polymerase (Table S5B). Twenty-one of the 28 missense mutations and 3 of the 6 silent mutations were at the regulatory elements, namely CpG island I/II/III, negative regulatory element (NRE)/core upstream regulatory sequence (CURS)/BCP, Enh I/II, core promoter and S2 promoters (Figure S2C).

Validation of NGS-defined HCC-associated SNVs by systemic literature review and meta-analysis

To validate these 60 NGS-defined HCC-associated HBV SNVs, a systemic literature review of 2,270 potentially relevant articles was conducted, of which 109 articles [15,20-127] were enrolled as appropriate for final analysis (Figure S3 and Table S6).

The articles discuss 167 HBV variants, which were categorized into 4 levels of supporting evidence associated with HCC. The proportions of NGS-defined HCC-associated SNVs among HBV variants with different levels of evidence associated with HCC declined significantly with decreasing levels of supporting evidence from Level A to Level D (chi-square test for the 2x4 contingency table,

$P < 0.0001$, Figure 2A).

Nine of the 12 HCC-associated variants supported by meta-analysis (Level A) were consistently identified as HCC-associated SNVs by NGS analysis. For the only three variants of Level A group not identified by NGS analysis, 1762T was identified as an HCC-associated SNV by our NGS-based subgroup analysis of HBeAg-positive patients with genotype C HBV infection (OR: 6.3, 95% CI = 1.1 to 34.0), whereas 1766T and 1768A mainly expressed in genotypes A and D, but not in genotypes B and C among our patients (Figure 2B and Figure S4L, S4N, S4O).

Seven (12%) of 60 HCC-associated HBV variants studied by less than 4 reports and supported in at least one report (Level B) were identified by NGS analysis. Only 2 (2%) of 85 HCC-unassociated variants supported by all previous studies with fewer than 4 reports (Level C) were identified by NGS analysis (Table S7.) None of the 10 HCC-unassociated variants by meta-analysis (Level D) were identified by NGS analysis (Figures S4 and S5).

The U-shaped distribution pattern of SNV frequency

A distinctive U-shaped distribution pattern of SNV II frequency was identified, with

most having either high (>80%) or low values (<20%) (Figure S6). The cut-off values of HCC-associated SNV frequency, *i.e.* pathogenic frequency, clearly showed a similar U-shaped distribution pattern (Table 1 and Figure 2C). Among 60 NGS-defined HCC-associated SNVs, 18 of 19 reported previously had cut-off values of SNV frequencies >20% and 16/19 had cut-off values of SNV frequencies >50%. Among the 41 novel HCC-associated SNVs, 25 and 13 had cut-off values of SNV frequencies >80% and <20%, respectively.

Deletion mutations in HBV genome

Deletion hotspots located in the preC/C gene, preS region and BCP region in both genotypes B and C HBV. The preS region had the greatest deletion frequencies and the most complex deletion patterns (Figure 3). Tables S8 and S9 list the deletion mutations of HBV in each subject discovered by NGS analysis. No significant difference was observed between HCC and non-HCC patients in the proportion having deletion mutations, either for genotype B (10/40 *vs* 8/47, $P=0.36$) or genotype C (34/53 *vs* 32/61, $P=0.21$). The proportion of patients with deletion mutations in the preS region (especially preS1) was much higher in the genotype C group than in the

genotype B group, among both HCC patients and non-HCC patients (Table S10). Intriguingly, the HCC group had more patients with deletion mutations involving nt2977–3013 (amino acid (aa.) 43–56) (6/34 vs 0/32, $P=0.025$, Figure 3B and 3D), deletion patterns II or III (18/34 vs 8/32, $P=0.039$), and deletion mutations at functional sites S (17/34 vs 7/32, $P=0.034$) and T (17/34 vs 7/32, $P=0.034$) in the preS region [15] (**Figure S7**), than the non-HCC group among genotype C infected patients with HBV deletions. Deletion index is calculated as: $\text{Deletion index} = \Sigma[(L_{\text{del}} \times R_{\text{del}})] / L_{\text{whole}}$, where L_{del} indicates length of the deletion fragment within a gene or region; R_{del} indicates ratio of the deletion fragment in a quasispecies within one subject; L_{whole} indicates whole length of a gene or region. HCC patients had higher deletion index at preS region than non-HCC patients, for genotype B ($P=0.030$) and genotype C ($P=0.045$) (Figure S8). Meta-analysis of preS deletions also showed a significant association with HCC (summary OR: 3.0, 95% CI, 2.3 to 3.9) (Figure S4W).

NGS-defined HCC-associated SNVs in small S region induce ER-stress, regulate DNA repair and affect hepatocarcinogenesis pathways

The levels of *GRP78* mRNA were 1.27-fold ($P=0.030$) and 1.38-fold ($P=0.033$)

higher in A273G mutant (genotype B) and C293G mutant (genotype C), respectively, than controls. The levels of *APE1* mRNA were 1.57-fold ($P=0.026$) and 1.41-fold ($P=0.018$) higher in C446G mutant (genotype C) and C446G/A456G mutant (genotype C), respectively, than controls (Figure S9). From microarray and Ingenuity Pathway Analysis, all the novel NGS-defined HCC-associated SNVs and their combinations in small S region could be involved in hepatocarcinogenesis pathways, by inducing cell apoptosis, regulating cell cycle, death and survival, and affecting DNA replication, recombination and repair. Intriguingly, these double or triple mutants enhanced those effects regarding hepatocarcinogenic pathways (Figure 4). For example, the triple mutant (A293G/C446G/A456G) in genotype C significantly affected 19805 genes compared to wild type construct. Among those affected genes, 311 belonged to the DNA repair-related genes with 206 down-regulated and 105 up-regulated. The up-regulated genes (Mu/Wt, $\log_2\text{ratio} \geq 1$) included PML, HMGA2, PARP1, XRCC1, PLK1, SLX1A and RMI2. The down-regulated genes (Mu/Wt, $\log_2\text{ratio} \leq -1$) included PCNA, PMS2, XRCC4, ATM, MDC1 and RAD51C. Consistent with the results of real-time PCR analysis, our microarray analysis showed significantly higher expression of the ER-stress markers *GRP78* and *APE1* mRNA in

small S mutants than controls (*GRP78*: 1.53-fold in A273G and 1.93-fold in C293G; *APE1*: 1.58-fold in C446G and 1.56-fold in C446G/A456G).

At the protein level, all NGS-defined HCC-associated SNVs in small S region of genotype B influenced ER stress-related factors such as PERK, IRE1 $\pm$, Ero1-L α , GRP78 and calnexin, except for T216C LHBS variant with only a trend to increase GRP78 expression (Figure S10A). In genotype C, three LHBS single mutated variants (A293G, C446G, and A456G) influenced PERK, IRE1 $\pm$, Ero1-L α and calnexin, except GRP78 up-regulated to some extent. Most of the double and triple LHBS affected the expression of PERK, IRE1 $\pm$, Ero1-L α , and calnexin, but not GRP78. Interestingly, the variants of genotype C resulted in accumulation of LHBS proteins more prominently than those of genotype B (Figure S10B). Additionally, all these genotype B LHBS variants and most of genotype C LHBS variants affected DNA repair systems through inducing *XRCC1* and reducing *APE1* (Figure S10). The C446G/A456G LHBS variants of genotype C induced Ape1 protein, compatible with results at the RNA level (Figures S9 and S10).

Discussion

This is the first large study to investigate HCC-associated HBV SNVs with their pathogenic frequencies using NGS analysis of whole HBV genomes. It is also the first to compare NGS-defined SNVs with HBV variants studied previously using systemic literature review and meta-analysis. This high-throughput sequencing analysis revealed that SNVs exist in 49.9% and 51.3% of nucleotide sites in HBV genomes of genotypes B and C, respectively. The preS region, including S2 promoter, had the highest mutation rate (the proportion of nucleotide sites with SNVs in a gene or region), while DR1, DR2, ϵ loop and poly-A signal had lower ones (Figures 1, S2 and S11.).

Sixty NGS-defined HCC-associated SNVs were found. The proportions of NGS-defined HCC-associated SNVs among HBV variants declined significantly along with the decreasing levels of supporting evidence from Level A to Level D after combining the results from NGS, systemic literature review and meta-analysis (Figure 2A), indicating that these NGS results were valid. All HCC-associated HBV variants supported by meta-analysis (Level A), except 1766T and 1768A (mainly expressed in genotypes A and D), but including the subgroup analysis of 1762T, were identified by NGS analysis [23,45,61]. These 10 HCC-associated HBV variants were located in the

C-terminal transactivation domain of HBx/BCP (1613A, 1653T, 1674C, 1753G, 1762T and 1764A), preC region (1846T, 1896A and 1899A) and preS2 region (53C).

This study provides the first meta-analysis to report association with HCC for 53C, 1613A, 1674C and 1846T. No NGS-defined HCC-associated SNVs were discovered among the 10 HCC-unassociated variants by meta-analysis (Level D).

Notably, this HBV whole-genome high-throughput sequencing analysis revealed 41 novel HCC-associated SNVs. These novel SNVs were discovered probably for the following reasons. First, NGS analysis can define HCC-associated SNVs with their specific sites, nucleotides and pathogenic frequencies, instead of presence or absence of specific major stains identified mainly by direct sequencing as in previous studies. Second, consistent with the unique U-shaped distribution pattern of SNV frequency [17], 32% of 41 novel HCC-associated SNVs had cut-off values of SNV frequencies below 20%, which could not be detected by direct sequencing. The HBV SNVs with low frequencies still have high viral content in CHB patients with high viral loads. Therefore, HCC-associated SNVs with low frequencies could have carcinogenic effects [128]. Coexistence of specific HCC-associated SNVs, even some with low frequencies, could increase the risk of HCC occurrence according to microarray

analysis (Figure 4). Third, 40 of 41 novel HCC-associated SNVs did not locate in previously reported hotspot regions, such as X gene, BCP region and preC region of an HBV genome [88,95,129]. For the genotype B group, 17 of 18 novel HCC-associated SNVs were located in the core region, small S region and P gene, while only one located in the preC region. For the genotype C group, all 23 novel HCC-associated SNVs were located in the preS region, small S region, core region and P gene (Figure S2). Point mutations of the preS region has rarely been explored, although the association of preS region deletions and HCC is extensively studied [130-132]. These novel HCC-associated SNVs might play a significant role in HCC development. The SNVs with a characteristic U-shape frequency distributions probably resulted from evolution processes and are consistent with a previous report [17,133]. Our results indicate that HCC-associated SNVs that lead to hepatocarcinoma are not related to the presence or absence of the specific SNVs, but with certain pathogenic frequencies.

Intriguingly, the 60 NGS-defined HCC-associated SNVs were distinct for both genotype B (n=24) and genotype C (n=34), except 53C for both genotypes. Genotype C HBV showed HCC-associated SNVs in the X gene (n=6), BCP region (n=6) and

preS1 region (n=15), while genotype B HBV did not. Genotype B HBV bear more HCC-associated SNVs in preC/core gene than genotype C HBV (n=10 vs n=2) (Figure S2). Previous studies also indicated that baseline BCP 1762T/1764A mutations increase the risk of HCC development exclusively in genotype C and not in genotype B [128]. Moreover, only two of the 12 HCC-associated variants supported by meta-analysis could receive further subgroup analysis of genotypes. All variants showed significant association with HCC in genotype C but not in genotype B (1653T, summary OR: 2.9, 95% CI, 2.1 to 3.9; 1753V, summary OR: 2.0, 95% CI, 1.1 to 3.6), which is consistent with our NGS results. Distinct genotype-specific HCC-associated SNVs possibly cause distinct clinical characteristics in HCC patients infected with different HBV genotypes. For instance, BCP 1762T/1764A mutations were found to be distinct HCC-associated SNVs for genotype C in this study, and independent predictors for postoperative recurrence in HCC in a previous study [134]. This might be the reason that genotype C HBV is associated with increased risk of HCC and an increased tumor recurrence rate [128,135,136]. In addition, the dN/dS ratios of the codons affected by these SNVs were calculated by DiversiTool (<http://josephhughes.github.io/btctools/>) [137]. Purifying selection (dN/dS<1)

involved about 60% of these codons (Figure S12 and Table S11). These findings are worthy of further studies.

From real-time PCR and microarray analysis, the NGS-defined HCC-associated SNVs and their combinations in small S region may induce ER-stress and influence hepatocarcinogenesis through affecting cell cycle, cell apoptosis, cell death, cell survival, DNA replication, DNA recombination, DNA repair, etc. The correlation of SNV combinations with HCC were supported by the results of SKAT analysis [66,138] (Table S12). For example, 311 DNA repair-related genes were affected by the triple mutant (A293G/C446G/A456G) in genotype C, involving DNA double-strand break repair by homologous recombination (HDR), non-homologous end joining (NHEJ), mismatch repair (MMR) and nucleotide excision repair (NER). Most genes involved in HDR, MMR and NER were down-regulated (Table S13). Among the NHEJ genes, those participating in the canonical NHEJ (C-NHEJ) were reduced but those participating in the alternative NHEJ (A-NHEJ), a major source of genomic instability, were induced. Western blot analysis demonstrated that LHBS variants with HCC-associated HBV SNVs in small S region induced ER stress with LHBS protein accumulation and altered DNA repair systems through XRCC1 and Ape1 (Figure

S10).

Our meta-analysis, which included 32 studies and evaluated a total of 5563 HBV-infected patients, demonstrated that infection with preS deletions was associated with a 3.0-fold increased risk of HCC (Figure S4W). For both genotypes B and C HBV, only preS deletion was significantly associated with HCC in terms of deletion index. The deletion index is composed of the deletion length and the deletion frequency by NGS analysis. Our sequencing analysis found that deletions were more often located in the 3'-terminus of the pre-S1 region and the 5'-terminus of the pre-S2 region (Figure 3), as reported previously [15]. Both preS1 and preS2 deletions can cause unbalanced production of envelope proteins, with a consequent accumulation of mutated large surface protein (LHBs) in the ER of hepatocytes, activating ER stress and ultimately HCC development [132,139,140]. Furthermore, among genotype C HBV infected patients with any deletion mutation, the HCC group had more patients with preS deletion mutations involving nt2977–3013 (aa.43–56) (Figure 3), deletion patterns II or III, and functional sites S and T (Figure S7), than the non-HCC group. The deletion involved in the function of envelope protein synthesis via S-promoter and CAD/Hsc70 binding activity. HCC-associated preS deletion

mutations involved a specific fragment (nt2977–3013, aa.43–56), which is essential for efficient hepadnaviral infectivity [141,142]. However, the hepatocarcinogenic mechanism of this fragment requires further elucidation.

According to the preceding systemic literature review and *in vitro* studies, the NGS-defined HCC-associated SNVs were unlikely to be discovered by chance. However, there are some limitations in this study. Because of the limited sample size, the novel HCC-associated SNVs discovered by our NGS analysis need to be confirmed by further cellular and animal model experiments. These NGS-defined HCC-associated HBV SNVs might not be applicable to patients with low viral load because of possible selection bias in this study. Furthermore, HBV variants in liver might not fully contribute to those in serum of CHB patients. Some differences of viral quasispecies at preC region have been found between those in serum and those in liver of CHB patients [143]. Several preS deletion mutants could result in less ability to replicate or produce infectious particles in the liver, contribute to HBV integration and ER stress via the accumulation of S protein, and probably play an important role in hepatocarcinogenesis [144-149]. It will be valuable to investigate HBV quasispecies in liver in the future.

Studies on the properties of HBV quasispecies indicate that distinct viral variants with greatest fitness evolve through immune selection during the natural course of chronic HBV infection [150] and play crucial roles in HCC development [151,152]. The development of HCC is a highly heterogeneous process involving mutations in HBV proteins and regulatory elements. Unique HBV oncoproteins (HBx isoforms and preS mutants) and mutated preC/core proteins can induce hepatocarcinogenesis through activation of ER-independent pathways [152], induction of ER and oxidative stress [153-155], regulation of microRNA expression [156], lipid metabolism disturbance [157] or epigenetic modification through modified genomic methylation status [158]. Mutations of HBV regulatory elements probably induce hepatocarcinogenesis through oncoprotein expression modulation [151,152], HBV DNA integration leading to chromosomal instability (DR1/DR2) [159] or HBV DNA methylation (CpG islands) [160] (Figure 5). We confirmed that all our novel NGS-defined HCC-associated SNVs in the small S region, particularly those with double and triple mutants, could influence hepatocarcinogenesis pathways (Figure 4) and some of them could induce ER-stress (Figures S9 and S10). In conclusion, NGS analysis revealed 41 novel HCC-associated SNVs and preS deletion, which involved HBV ORFs and regulatory

elements and are worthy of further investigation in hepatocarcinogenesis.

Acknowledgements

This work was supported by grants from the Ministry of Science and Technology, R.O.C (MOST 102-2628-B-006-008-MY3, NSC 99-2314-B-006-028-MY3, and NSC 95-2321-B-010-005), the Ministry of Health and Welfare, Executive Yuan, R.O.C (MOHW104-TDU-B-211-124-001 and MOHW106-TDU-B-211-144-003 for the Center of Excellence), and the Infectious Disease and Signaling Research Center of National Cheng Kung University. We thank the Human Biobank of the Research Center of Clinical Medicine and Cancer Data Bank of the Cancer Center of National Cheng Kung University Hospital Cancer Center for providing HCC samples. We are grateful to Dr. Sheng-Hsiang Lin for providing statistical consulting services from the Biostatistics Consulting Center, National Cheng Kung University Hospital. We thank Chun-Pei Cheng, Yi-Cheng Tsai and Chia-Hua Li for assisting NGS analysis. We are grateful to Jui-Chu Yang for providing the bioinformatics consulting services of IPA from the Human Biobank, Research Center of Clinical Medicine, National Cheng Kung University Hospital. We are indebted to Professor Wenya Huang and Professor

Kung-Chia Young for invaluable suggestions, and Ted Kony and I-Chen Li for English editing.

Author Contributions: W-CL and I-CW contributed equally to this article. W-CL handled experimental design, experiment performance, clinical and NGS data analyses, biometric consulting and manuscript writing; I-CW was responsible for statistical analysis and manuscript writing; Y-CL, S-LY and J-JL helped in meta-analysis; C-PL and BC-HC executed NGS and analyzed high-throughput data using bio-informatic software; J-HC, P-FL, Y-TC, P-WC, VST and K-TS assisted in analyzing the classification-based association rules; Y-JL, C-JY and P-NC provided clinical samples from HCC patients; J-CY and K-CC provided the bioinformatics consulting services of IPA; C-HH helped in sample preparation, clinical data arrangement and statistical analysis; T-TC and J-CW coordinated the study, gathered clinical data and serum samples from patients with chronic hepatitis B, and composed the manuscript.

Abbreviations: amino acid (aa.), basal core promoter (BCP), chronic hepatitis B

(CHB), core upstream regulatory sequence (CURS), epsilon loop (μ loop), Enhancer (Enh) I, hepatitis B virus (HBV), hepatocellular carcinoma (HCC), homologous recombination (HDR), Ingenuity Pathway Analysis (IPA), mismatch repair (MMR), negative regulatory element (NRE), next-generation sequencing (NGS), non-homologous end joining (NHEJ), nucleotide excision repair (NER), odds ratio (OR), open reading frames (ORFs), pre-core (preC), single nucleotide variants (SNVs), small surface (S).

SUPPLEMENTARY MATERIAL ONLINE

Supplementary material file including Material and Methods, Legends for Figures S1-S12, and Tables S1- S13

Figure S1. Begg's funnel plot of 1753V, 1896A and 1899A

Figure S2. HBV genome-wide characterization of substitutions with distinct

NGS-defined HCC-associated SNVs in genotype B and genotype C

Figure S3. Flowchart of article selection

Figure S4. Summary odds ratios of hepatocellular carcinoma

Figure S5. Odds ratios of NGS analysis in 10 HBV variants with Level D evidence

Figure S6. The U-shaped distribution pattern of SNV II frequency

Figure S7. The percentage of patients with distinct preS deletion patterns among patients with any deletion mutation of HBV full genome comparing HCC and non-HCC groups

Figure S8. Distribution of deletion index for different HBV gene fragments in HCC and non-HCC patients

Figure S9. HBV small S mutants induces the ER stress compared with HBV wildtype small S proteins

Figure S10. LHBS variants with HCC-associated HBV SNVs in small S region

induced ER stress and involved DNA repair system

Figure S11. Mutation rates of HBV open reading frames and regulatory elements

Figure S12. The dN/dS ratios of the codons affected by NGS-defined

HCC-associated SNVs in each HBV gene

Table S1. Primer sets for amplifying full HBV genome in next-generation sequencing (NGS)

Table S2. Oligonucleotides used for PCR mutagenesis

Table S3. Clinical characteristics of chronic hepatitis B patients with and without HCC

Table S4. Mapping statistics of NGS reads

Table S5A. Seventeen significant HCC-associated SAVs among genotype B HBV infected patients

Table S5B. Twenty-eight significant HCC-associated SAVs among genotype C HBV infected patients

Table S6. Characteristics of 109 studies included in the meta-analysis

Table S7. NGS-defined HCC-associated HBV SNVs with Level B evidence or Level C evidence

Table S8. Characteristics of genotype B HBV deletion patterns among patients with

and without HCC

Table S9. Characteristics of genotype C HBV deletion patterns among patients with
and without HCC

Table S10. Proportions of patients with deletion mutations among HBV genome
comparing genotype B and genotype C groups

Table S11. The substitution patterns of NGS-defined HCC-associated SNVs

Table S12. The association of NGS-defined SNVs and their combinations at small S
gene with HCC by SKAT

Table S13. Expression profiles of essential genes of the five major DNA repair
mechanisms in Huh7 cell with small S triple mutants of genotype C HBV

References

1. Ganem D, Prince AM. Hepatitis B virus infection--natural history and clinical consequences. *New Engl J Med* 2004; **350**: 1118-1129.
2. Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Viral Hepatitis* 2004; **11**: 97-107.
3. Lok AS, McMahon BJ. Chronic hepatitis B. *Hepatology* 2007; **45**: 507-539.
4. Seeger C, Mason WS. Molecular biology of hepatitis B virus infection. *Virology* 2015; **479-480**: 672-686.
5. Cassidy A, Mossman S, Olivieri A, *et al.* Hepatitis B vaccine effectiveness in the face of global HBV genotype diversity. *Expert Rev Vaccines* 2011; **10**: 1709-1715.
6. Rodriguez-Frias F, Buti M, Tabernero D, *et al.* Quasispecies structure, cornerstone of hepatitis B virus infection: mass sequencing approach. *World J Gastroenterol* 2013; **19**: 6995-7023.
7. Liao Y, Hu X, Chen J, *et al.* Precore mutation of hepatitis B virus may contribute to hepatocellular carcinoma risk: evidence from an updated meta-analysis. *PloS One* 2012; **7**: e38394.
8. Liu S, Zhang H, Gu C, *et al.* Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. *J Natl Cancer Inst* 2009; **101**: 1066-1082.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-1760.
10. Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078-2079.
11. Liu WC, Lin CP, Cheng CP, *et al.* Aligning to the sample-specific reference sequence to optimize the accuracy of next-generation sequencing analysis for hepatitis B virus. *Hepatol Int* 2016; **10**: 147-157.
12. Ye K, Schulz MH, Long Q, *et al.* Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009; **25**: 2865-2871.
13. Liu WC, Mizokami M, Buti M, *et al.* Simultaneous quantification and genotyping of hepatitis B virus for genotypes A to G by real-time PCR and

- two-step melting curve analysis. *J Clin Microbiol* 2006; **44**: 4491-4497.
14. Tamura K, Stecher G, Peterson D, *et al.* MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 2013; **30**: 2725-2729.
 15. Chen BF, Liu CJ, Jow GM, *et al.* High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases. *Gastroenterology* 2006; **130**: 1153-1168.
 16. Yu G, Wang LG, Han Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; **16**: 284-287.
 17. Li F, Zhang D, Li Y, *et al.* Whole genome characterization of hepatitis B virus quasispecies with massively parallel pyrosequencing. *Clin Microbiol Infect* 2015; **21**: 280-287.
 18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B-Methodol* 1995; **57**: 289-300.
 19. Khomtchouk BB, Van Booven DJ, Wahlestedt C. HeatmapGenerator: high performance RNAseq and microarray visualization software suite to examine differential gene expression levels using an R and C++ hybrid computational pipeline. *Source Code Biol Med* 2014; **9**: 30.
 20. Abe K, Thung SN, Wu HC, *et al.* Pre-S2 deletion mutants of hepatitis B virus could have an important role in hepatocarcinogenesis in Asian children. *Cancer Sci* 2009; **100**: 2249-2254.
 21. Ahn SH, Yuen L, Han KH, *et al.* Molecular and clinical characteristics of hepatitis B virus in Korea. *Journal of medical virology* 2010; **82**: 1126-1134.
 22. Asim M, Malik A, Sarma MP, *et al.* Hepatitis B virus BCP, Precore/core, X gene mutations/genotypes and the risk of hepatocellular carcinoma in India. *J Med Virol* 2010; **82**: 1115-1125.
 23. Bahramali G, Sadeghizadeh M, Amini-Bavil-Olyaei S, *et al.* Clinical, virologic and phylogenetic features of hepatitis B infection in Iranian patients. *World J Gastroenterol* 2008; **14**: 5448-5453.
 24. Bai X, Zhu Y, Jin Y, *et al.* Temporal acquisition of sequential mutations in the enhancer II and basal core promoter of HBV in individuals at high risk for hepatocellular carcinoma. *Carcinogenesis* 2011; **32**: 63-68.
 25. Baptista M, Kramvis A, Kew MC. High prevalence of 1762(T) 1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black

- Africans with hepatocellular carcinoma compared with asymptomatic carriers. *Hepatology* 1999; **29**: 946-953.
26. 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.
 27. Cao Z, Bai X, Guo X, *et al.* High prevalence of hepatitis B virus pre-S mutation and its association with hepatocellular carcinoma in Qidong, China. *Arch Virol* 2008; **153**: 1807-1812.
 28. Chan HL, Hui AY, Wong ML, *et al.* Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. *Gut* 2004; **53**: 1494-1498.
 29. Chen CH, Changchien CS, Lee CM, *et al.* Combined mutations in pre-s/surface and core promoter/precore regions of hepatitis B virus increase the risk of hepatocellular carcinoma: a case-control study. *J Infect Dis* 2008; **198**: 1634-1642.
 30. Chen CH, Changchien CS, Lee CM, *et al.* A study on sequence variations in pre-S/surface, X and enhancer II/core promoter/precore regions of occult hepatitis B virus in non-B, non-C hepatocellular carcinoma patients in Taiwan. *Int J Cancer* 2009; **125**: 621-629.
 31. Chen QY, Harrison TJ, Sabin CA, *et al.* The Effect of HBV Genotype C on the Development of HCC Differs Between Wild-Type Viruses and Those With BCP Double Mutations (T(1762)A(1764)). *Hepat Mon* 2014; **14**: e16214.
 32. Chen YM, Wu SH, Qiu CN, *et al.* Hepatitis B virus subgenotype C2- and B2-associated mutation patterns may be responsible for liver cirrhosis and hepatocellular carcinoma, respectively. *Braz J Med Biol Res* 2013; **46**: 614-622.
 33. Cho EY, Choi CS, Cho JH, *et al.* Association between Hepatitis B Virus X Gene Mutations and Clinical Status in Patients with Chronic Hepatitis B Infection. *Gut Liver* 2011; **5**: 70-76.
 34. Cho EY, Kim HC, Choi CS, *et al.* Nucleotide changes related to hepatocellular carcinoma in the enhancer 1/x-promoter of hepatitis B virus subgenotype C2 in cirrhotic patients. *Cancer Sci* 2010; **101**: 1905-1912.
 35. Choi CS, Cho EY, Park R, *et al.* X gene mutations in hepatitis B patients with cirrhosis, with and without hepatocellular carcinoma. *J Med Virol* 2009; **81**:

- 1721-1725.
36. 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**: 161-168.
 37. Chu CM, Lin CC, Chen YC, *et al.* Basal core promoter mutation is associated with progression to cirrhosis rather than hepatocellular carcinoma in chronic hepatitis B virus infection. *Br J Cancer* 2012; **107**: 2010-2015.
 38. Chu CM, Lin CC, Lin SM, *et al.* Viral load, genotypes, and mutants in hepatitis B virus-related hepatocellular carcinoma: special emphasis on patients with early hepatocellular carcinoma. *Dig Dis Sci* 2012; **57**: 232-238.
 39. Datta S, Ghosh A, Dasgupta D, *et al.* Novel point and combo-mutations in the genome of hepatitis B virus-genotype D: characterization and impact on liver disease progression to hepatocellular carcinoma. *PloS One* 2014; **9**: e110012.
 40. Fan W, Shi B, Wei H, *et al.* Comparison of hepatitis B X gene mutation between patients with hepatocellular carcinoma and patients with chronic hepatitis B. *Virus Genes* 2011; **42**: 162-170.
 41. Fang ZL, Sabin CA, Dong BQ, *et al.* HBV A1762T, G1764A mutations are a valuable biomarker for identifying a subset of male HBsAg carriers at extremely high risk of hepatocellular carcinoma: a prospective study. *Am J Gastroenterol* 2008; **103**: 2254-2262.
 42. Fang ZL, Sabin CA, Dong BQ, *et al.* Hepatitis B virus pre-S deletion mutations are a risk factor for hepatocellular carcinoma: a matched nested case-control study. *J Gen Virol* 2008; **89**: 2882-2890.
 43. Fang ZL, Yang J, Ge X, *et al.* Core promoter mutations (A(1762)T and G(1764)A) and viral genotype in chronic hepatitis B and hepatocellular carcinoma in Guangxi, China. *J Med Virol* 2002; **68**: 33-40.
 44. Gao ZY, Li T, Wang J, *et al.* Mutations in preS genes of genotype C hepatitis B virus in patients with chronic hepatitis B and hepatocellular carcinoma. *J Gastroenterol* 2007; **42**: 761-768.
 45. Gopalakrishnan D, Keyter M, Shenoy KT, *et al.* Hepatitis B virus subgenotype A1 predominates in liver disease patients from Kerala, India. *World J Gastroenterol* 2013; **19**: 9294-9306.
 46. Guo X, Jin Y, Qian G, *et al.* 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**: 718-725.
47. Huang CJ, Wu CF, Lan CY, *et al.* Impact of genetic heterogeneity in polymerase of hepatitis B virus on dynamics of viral load and hepatitis B progression. *PloS One* 2013; **8**: e70169.
 48. Huy TT, Ushijima H, Quang VX, *et al.* Characteristics of core promoter and precore stop codon mutants of hepatitis B virus in Vietnam. *J Med Virol* 2004; **74**: 228-236.
 49. Huy TT, Ushijima H, Win KM, *et al.* High prevalence of hepatitis B virus pre-s mutant in countries where it is endemic and its relationship with genotype and chronicity. *J Clin Microbiol* 2003; **41**: 5449-5455.
 50. Ito K, Tanaka Y, Kato M, *et al.* Comparison of complete sequences of hepatitis B virus genotype C between inactive carriers and hepatocellular carcinoma patients before and after seroconversion. *J Gastroenterol* 2007; **42**: 837-844.
 51. Ito K, Tanaka Y, Orito E, *et al.* T1653 mutation in the box alpha increases the risk of hepatocellular carcinoma in patients with chronic hepatitis B virus genotype C infection. *Clin Infect Dis* 2006; **42**: 1-7.
 52. Jang JS, Kim HS, Kim HJ, *et al.* Association of concurrent hepatitis B surface antigen and antibody to hepatitis B surface antigen with hepatocellular carcinoma in chronic hepatitis B virus infection. *J Med Virol* 2009; **81**: 1531-1538.
 53. Jang JW, Chun JY, Park YM, *et al.* Mutational complex genotype of the hepatitis B virus X /precore regions as a novel predictive marker for hepatocellular carcinoma. *Cancer Sci* 2012; **103**: 296-304.
 54. Jang JW, Lee YC, Kim MS, *et al.* A 13-year longitudinal study of the impact of double mutations in the core promoter region of hepatitis B virus on HBeAg seroconversion and disease progression in patients with genotype C chronic active hepatitis. *J Viral Hepat* 2007; **14**: 169-175.
 55. Jiang S, Yang Z, Li W, *et al.* Re-evaluation of the carcinogenic significance of hepatitis B virus integration in hepatocarcinogenesis. *PloS One* 2012; **7**: e40363.
 56. Jiang SZ, Gao ZY, Li T, *et al.* T3098C and T53C mutations of HBV genotype C is associated with HBV infection progress. *Biomed Environ Sci* 2009; **22**: 511-517.
 57. Juniastuti, Utsumi T, Aksono EB, *et al.* Predominance of precore mutations

- and clinical significance of basal core promoter mutations in chronic hepatitis B virus infection in Indonesia. *Biomed Rep* 2013; **1**: 522-528.
58. Kao JH, Chen PJ, Lai MY, *et al.* Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. *Gastroenterology* 2003; **124**: 327-334.
 59. Kao JH, Liu CJ, Jow GM, *et al.* Fine mapping of hepatitis B virus pre-S deletion and its association with hepatocellular carcinoma. *Liver Int* 2012; **32**: 1373-1381.
 60. Karino Y, Toyota J, Sato T, *et al.* Early mutation of precore (A1896) region prior to core promoter region mutation leads to decrease of HBV replication and remission of hepatic inflammation. *Dig Dis Sci* 2000; **45**: 2207-2213.
 61. Khan A, Al Balwi MA, Tanaka Y, *et al.* Novel point mutations and mutational complexes in the enhancer II, core promoter and precore regions of hepatitis B virus genotype D1 associated with hepatocellular carcinoma in Saudi Arabia. *Int J Cancer* 2013; **133**: 2864-2871.
 62. 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**: 1002-1008.
 63. 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**: 3575-3580.
 64. Kusakabe A, Tanaka Y, Inoue M, *et al.* A population-based cohort study for the risk factors of HCC among hepatitis B virus mono-infected subjects in Japan. *J Gastroenterol* 2011; **46**: 117-124.
 65. Laskus T, Radkowski M, Nowicki M, *et al.* Association between hepatitis B virus core promoter rearrangements and hepatocellular carcinoma. *Biochem Biophys Res Commun* 1998; **244**: 812-814.
 66. Lee S, Emond MJ, Bamshad MJ, *et al.* Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. *Am J Hum Genet* 2012; **91**: 224-237.
 67. Lee SA, Mun HS, Kim H, *et al.* Naturally occurring hepatitis B virus X deletions and insertions among Korean chronic patients. *J Med Virol* 2011; **83**: 65-70.

68. Li MS, Lau TC, Chan SK, *et al.* The G1613A mutation in the HBV genome affects HBeAg expression and viral replication through altered core promoter activity. *PloS One* 2011; **6**: e21856.
69. Li W, Chen G, Yu X, *et al.* Accumulation of the mutations in basal core promoter of hepatitis B virus subgenotype C1 increase the risk of hepatocellular carcinoma in Southern China. *Int J Clin Exp Pathol* 2013; **6**: 1076-1085.
70. Lin CL, Liu CH, Chen W, *et al.* Association of pre-S deletion mutant of hepatitis B virus with risk of hepatocellular carcinoma. *J Gastroenterol Hepatol* 2007; **22**: 1098-1103.
71. Liu CJ, Chen BF, Chen PJ, *et al.* Role of hepatitis B viral load and basal core promoter mutation in hepatocellular carcinoma in hepatitis B carriers. *J Infect Dis* 2006; **193**: 1258-1265.
72. Liu CJ, Chen BF, Chen PJ, *et al.* Role of hepatitis B virus precore/core promoter mutations and serum viral load on noncirrhotic hepatocellular carcinoma: a case-control study. *J Infect Dis* 2006; **194**: 594-599.
73. Liu S, Xie J, Yin J, *et al.* A matched case-control study of hepatitis B virus mutations in the preS and core promoter regions associated independently with hepatocellular carcinoma. *J Med Virol* 2011; **83**: 45-53.
74. Lyu H, Lee D, Chung YH, *et al.* Synergistic effects of A1896, T1653 and T1762/A1764 mutations in genotype c2 hepatitis B virus on development of hepatocellular carcinoma. *J Viral Hepat* 2013; **20**: 219-224.
75. Malik A, Singhal DK, Albanyan A, *et al.* Hepatitis B virus gene mutations in liver diseases: a report from New Delhi. *PloS One* 2012; **7**: e39028.
76. Mendy ME, Kaye S, Le Roux E, *et al.* Application of a novel, rapid, and sensitive oligonucleotide ligation assay for detection of cancer-predicting mutations in the precore and basal core promoter of hepatitis B virus. *J Clin Microbiol* 2008; **46**: 2723-2730.
77. Mun HS, Lee SA, Jee Y, *et al.* The prevalence of hepatitis B virus preS deletions occurring naturally in Korean patients infected chronically with genotype C. *J Med Virol* 2008; **80**: 1189-1194.
78. Munoz A, Chen JG, Egner PA, *et al.* Predictive power of hepatitis B 1762T/1764A mutations in plasma for hepatocellular carcinoma risk in Qidong, China. *Carcinogenesis* 2011; **32**: 860-865.

79. Muroyama R, Kato N, Yoshida H, *et al.* Nucleotide change of codon 38 in the X gene of hepatitis B virus genotype C is associated with an increased risk of hepatocellular carcinoma. *J Hepatol* 2006; **45**: 805-812.
80. Nagasaka A, Hige S, Marutani M, *et al.* Prevalence of mutations in core promoter/precore region in Japanese patients with chronic hepatitis B virus infection. *Dig Dis Sci* 1998; **43**: 2473-2478.
81. Panigrahi R, Biswas A, Banerjee A, *et al.* Subgenotype D5, BCP and MHR mutations in hepatic complications among hepatitis B virus infected patients from Orissa, India. *Infect Genet Evol* 2012; **12**: 1622-1629.
82. Park YM, Jang JW, Yoo SH, *et al.* Combinations of eight key mutations in the X/preC region and genomic activity of hepatitis B virus are associated with hepatocellular carcinoma. *J Viral Hepat* 2014; **21**: 171-177.
83. Pollicino T, Campo S, Raimondo G. PreS and core gene heterogeneity in hepatitis B virus (HBV) genomes isolated from patients with long-lasting HBV chronic infection. *Virology* 1995; **208**: 672-677.
84. Poustchi H, Mohamadkhani A, Bowden S, *et al.* Clinical significance of precore and core promoter mutations in genotype D hepatitis B-related chronic liver disease. *J Viral Hepat* 2008; **15**: 753-760.
85. Qu L, Kuai X, Liu T, *et al.* Pre-S deletion and complex mutations of hepatitis B virus related to young age hepatocellular carcinoma in Qidong, China. *PloS One* 2013; **8**: e59583.
86. Qu LS, Liu JX, Liu TT, *et al.* Association of hepatitis B virus pre-S deletions with the development of hepatocellular carcinoma in Qidong, China. *PloS One* 2014; **9**: e98257.
87. Qu LS, Liu TT, Jin F, *et al.* Combined pre-S deletion and core promoter mutations related to hepatocellular carcinoma: A nested case-control study in China. *Hepatol Res* 2011; **41**: 54-63.
88. Qu LS, Zhu J, Liu TT, *et al.* Effect of combined mutations in the enhancer II and basal core promoter of hepatitis B virus on development of hepatocellular carcinoma in Qidong, China. *Hepatol Res* 2014; **44**: 1186-1195.
89. Sakamoto T, Tanaka Y, Orito E, *et al.* Novel subtypes (subgenotypes) of hepatitis B virus genotypes B and C among chronic liver disease patients in the Philippines. *J Gen Virol* 2006; **87**: 1873-1882.
90. 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**: 3191-3197.
91. Song BC, Kim H, Kim SH, *et al.* Comparison of full length sequences of hepatitis B virus isolates in hepatocellular carcinoma patients and asymptomatic carriers of Korea. *J Med Virol* 2005; **75**: 13-19.
 92. Song LH, Duy DN, Binh VQ, *et al.* Low frequency of mutations in the X gene, core promoter and precore region of hepatitis B virus infected Vietnamese. *J Viral Hepat* 2005; **12**: 160-167.
 93. Sugauchi F, Ohno T, Orito E, *et al.* Influence of hepatitis B virus genotypes on the development of preS deletions and advanced liver disease. *J Med Virol* 2003; **70**: 537-544.
 94. 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**: 1687-1697.
 95. 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**: 3604-3611.
 96. Tanaka Y, Mukaide M, Orito E, *et al.* Specific mutations in enhancer II/core promoter of hepatitis B virus subgenotypes C1/C2 increase the risk of hepatocellular carcinoma. *J Hepatol* 2006; **45**: 646-653.
 97. Tangkijvanich P, Sa-Nguanmoo P, Mahachai V, *et al.* A case-control study on sequence variations in the enhancer II/core promoter/precore and X genes of hepatitis B virus in patients with hepatocellular carcinoma. *Hepatol Int* 2010; **4**: 577-584.
 98. Tatsukawa M, Takaki A, Shiraha H, *et al.* Hepatitis B virus core promoter mutations G1613A and C1653T are significantly associated with hepatocellular carcinoma in genotype C HBV-infected patients. *BMC Cancer* 2011; **11**: 458.
 99. Thongbai C, Sa-nguanmoo P, Kranokpiruk P, *et al.* Hepatitis B virus genetic variation and TP53 R249S mutation in patients with hepatocellular carcinoma in Thailand. *Asian Pac J Cancer Prevent* 2013; **14**: 3555-3559.
 100. Tong MJ, Blatt LM, Kao JH, *et al.* Precore/basal core promoter mutants and hepatitis B viral DNA levels as predictors for liver deaths and hepatocellular carcinoma. *World J Gastroenterol* 2006; **12**: 6620-6626.

101. Tong MJ, Blatt LM, Kao JH, *et al.* Basal core promoter T1762/A1764 and precore A1896 gene mutations in hepatitis B surface antigen-positive hepatocellular carcinoma: a comparison with chronic carriers. *Liver Int* 2007; **27**: 1356-1363.
102. Tong MJ, Huynh TT, Siripongsakun S. Familial clustering of hepatocellular carcinoma in HBsAg-positive patients in the United States. *Hepatol Int* 2013; **7**: 1019-1029.
103. Truong BX, Seo Y, Yano Y, *et al.* Genotype and variations in core promoter and pre-core regions are related to progression of disease in HBV-infected patients from Northern Vietnam. *Int J Mol Med* 2007; **19**: 293-299.
104. Utama A, Siburian MD, Fanany I, *et al.* Low prevalence of hepatitis B virus pre-S deletion mutation in Indonesia. *J Med Virol* 2011; **83**: 1717-1726.
105. Venard V, Corsaro D, Kajzer C, *et al.* Hepatitis B virus X gene variability in French-born patients with chronic hepatitis and hepatocellular carcinoma. *J Med Virol* 2000; **62**: 177-184.
106. Wang YM, Ng WC, Lo SK. Detection of pre-S/S gene mutants in chronic hepatitis B carriers with concurrent hepatitis B surface antibody and hepatitis B surface antigen. *Journal of gastroenterology* 1999; **34**: 600-606.
107. Wang Z, Tanaka Y, Huang Y, *et al.* Clinical and virological characteristics of hepatitis B virus subgenotypes Ba, C1, and C2 in China. *Journal of clinical microbiology* 2007; **45**: 1491-1496.
108. Welschinger R, Kew MC, Viana R, *et al.* T1653 mutation in the enhancer II region of the hepatitis B virus genome in southern African Blacks with hepatocellular carcinoma. *European journal of gastroenterology & hepatology* 2010; **22**: 541-545.
109. Wu Y, Gan Y, Gao F, *et al.* Novel natural mutations in the hepatitis B virus reverse transcriptase domain associated with hepatocellular carcinoma. *PloS One* 2014; **9**: e94864.
110. Xiao L, Zhou B, Gao H, *et al.* Hepatitis B virus genotype B with G1896A and A1762T/G1764A mutations is associated with hepatitis B related acute-on-chronic liver failure. *J Med Virol* 2011; **83**: 1544-1550.
111. Xu L, Qian G, Tang L, *et al.* Genetic variations of hepatitis B virus and serum aflatoxin-lysine adduct on high risk of hepatocellular carcinoma in Southern Guangxi, China. *J Hepatol* 2010; **53**: 671-676.

112. Yan T, Li K, Li F, *et al.* T1846 and A/G1913 are associated with acute on chronic liver failure in patients infected with hepatitis B virus genotypes B and C. *J Med Virol* 2011; **83**: 996-1004.
113. Yang HI, Yeh SH, Chen PJ, *et al.* Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma. *J Nat Cancer Inst* 2008; **100**: 1134-1143.
114. Yeung P, Wong DK, Lai CL, *et al.* Association of hepatitis B virus pre-S deletions with the development of hepatocellular carcinoma in chronic hepatitis B. *J Infect Dis* 2011; **203**: 646-654.
115. Yin J, Xie J, Liu S, *et al.* Association between the various mutations in viral core promoter region to different stages of hepatitis B, ranging of asymptomatic carrier state to hepatocellular carcinoma. *Am J Gastroenterol* 2011; **106**: 81-92.
116. Young Min Park ET, Boo Sung Kim. Mutations of precore and proximal core regions of hepatitis B virus genome in serum of hepatocellular carcinoma patients. *Hepatol Res* 1999; **16**.
117. Yousif M, Mudawi H, Bakhiet S, *et al.* Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. *BMC Infect Dis* 2013; **13**: 328.
118. Yuan J, Zhou B, Tanaka Y, *et al.* Hepatitis B virus (HBV) genotypes/subgenotypes in China: mutations in core promoter and precore/core and their clinical implications. *J Clin Virol* 2007; **39**: 87-93.
119. Yuan JM, Ambinder A, Fan Y, *et al.* Prospective evaluation of hepatitis B 1762(T)/1764(A) mutations on hepatocellular carcinoma development in Shanghai, China. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 590-594.
120. 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.
121. Zhang D, Ma S, Zhang X, *et al.* Prevalent HBV point mutations and mutation combinations at BCP/preC region and their association with liver disease progression. *BMC Infect Dis* 2010; **10**: 271.
122. Zhang Q, Yin J, Zhang Y, *et al.* HLA-DP polymorphisms affect the outcomes of chronic hepatitis B virus infections, possibly through interacting with viral mutations. *J Virol* 2013; **87**: 12176-12186.

123. Zheng JX, Zeng Z, Zheng YY, *et al.* Role of hepatitis B virus base core and precore/core promoter mutations on hepatocellular carcinoma in untreated older genotype C Chinese patients. *J Viral Hepat* 2011; **18**: e423-431.
124. Zhong S, Chan JY, Yeo W, *et al.* Hepatitis B envelope protein mutants in human hepatocellular carcinoma tissues. *J Viral Hepat* 1999; **6**: 195-202.
125. Zhong S, Chan JY, Yeo W, *et al.* Frequent integration of precore/core mutants of hepatitis B virus in human hepatocellular carcinoma tissues. *J Viral Hepat* 2000; **7**: 115-123.
126. Zhu R, Zhang HP, Yu H, *et al.* Hepatitis B virus mutations associated with in situ expression of hepatitis B core antigen, viral load and prognosis in chronic hepatitis B patients. *Pathol Res Pract* 2008; **204**: 731-742.
127. Zhu Y, Jin Y, Guo X, *et al.* Comparison study on the complete sequence of hepatitis B virus identifies new mutations in core gene associated with hepatocellular carcinoma. *Cancer Epidemiol Biomarkers Prev* 2010; **19**: 2623-2630.
128. 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**: 1134-1143.
129. Zhang XD, Wang Y, Ye LH. Hepatitis B virus X protein accelerates the development of hepatoma. *Cancer biology & medicine* 2014; **11**: 182-190.
130. Gunther S. Genetic variation in HBV infection: genotypes and mutants. *J Clin Virol* 2006; **36 Suppl 1**: S3-S11.
131. 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**: 466-477.
132. Wang HC, Huang W, Lai MD, *et al.* Hepatitis B virus pre-S mutants, endoplasmic reticulum stress and hepatocarcinogenesis. *Cancer Sci* 2006; **97**: 683-688.
133. Haegeman B, Weitz JS. A neutral theory of genome evolution and the frequency distribution of genes. *BMC Genomics* 2012; **13**: 196.
134. Yeh CT, So M, Ng J, *et al.* Hepatitis B virus-DNA level and basal core promoter A1762T/G1764A mutation in liver tissue independently predict postoperative survival in hepatocellular carcinoma. *Hepatology* 2010; **52**:

- 1922-1933.
135. Croagh CM, Desmond PV, Bell SJ. Genotypes and viral variants in chronic hepatitis B: A review of epidemiology and clinical relevance. *World J Hepatol* 2015; **7**: 289-303.
 136. Chen JD, Liu CJ, Lee PH, *et al.* Hepatitis B genotypes correlate with tumor recurrence after curative resection of hepatocellular carcinoma. *Clin Gastroenterol Hepatol* 2004; **2**: 64-71.
 137. Xia H, Beck AS, Gargili A, *et al.* Transstadial Transmission and Long-term Association of Crimean-Congo Hemorrhagic Fever Virus in Ticks Shapes Genome Plasticity. *Sci Rep* 2016; **6**: 35819.
 138. Ionita-Laza I, Lee S, Makarov V, *et al.* Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. *Am J Hum Genet* 2013; **92**: 841-853.
 139. Hsieh YH, Su IJ, Wang HC, *et al.* Hepatitis B virus pre-S2 mutant surface antigen induces degradation of cyclin-dependent kinase inhibitor p27Kip1 through c-Jun activation domain-binding protein 1. *Mol Cancer Res* 2007; **5**: 1063-1072.
 140. Hung JH, Su IJ, Lei HY, *et al.* Endoplasmic reticulum stress stimulates the expression of cyclooxygenase-2 through activation of NF-kappaB and pp38 mitogen-activated protein kinase. *J Biol Chem* 2004; **279**: 46384-46392.
 141. He Y, Ye X, Tiollais P, *et al.* Selection of HBV preS1-binding penta-peptides by phage display. *Acta Biochim Biophys Sin* 2014; **46**: 691-698.
 142. Glebe D, Urban S. Viral and cellular determinants involved in hepadnaviral entry. *World J Gastroenterol* 2007; **13**: 22-38.
 143. Dienes HP, Gerken G, Goergen B, *et al.* Analysis of the precore DNA sequence and detection of precore antigen in liver specimens from patients with anti-hepatitis B e-positive chronic hepatitis. *Hepatology* 1995; **21**: 1-7.
 144. Caselmann WH, Meyer M, Kekule AS, *et al.* A trans-activator function is generated by integration of hepatitis B virus preS/S sequences in human hepatocellular carcinoma DNA. *Proc Natl Acad Sci U S A* 1990; **87**: 2970-2974.
 145. Hildt E, Saher G, Bruss V, *et al.* The hepatitis B virus large surface protein (LHBs) is a transcriptional activator. *Virology* 1996; **225**: 235-239.
 146. Hildt E, Munz B, Saher G, *et al.* The PreS2 activator MHBs(t) of hepatitis B

- virus activates c-raf-1/Erk2 signaling in transgenic mice. *EMBO J* 2002; **21**: 525-535.
147. Bonilla Guerrero R, Roberts LR. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. *J Hepatol* 2005; **42**: 760-777.
 148. Luan F, Liu H, Gao L, *et al.* Hepatitis B virus protein preS2 potentially promotes HCC development via its transcriptional activation of hTERT. *Gut* 2009; **58**: 1528-1537.
 149. Pollicino T, Cacciola I, Saffioti F, *et al.* Hepatitis B virus PreS/S gene variants: pathobiology and clinical implications. *J Hepatol* 2014; **61**: 408-417.
 150. Homs M, Buti M, Tabernero D, *et al.* Quasispecies dynamics in main core epitopes of hepatitis B virus by ultra-deep-pyrosequencing. *World J Gastroenterol* 2012; **18**: 6096-6105.
 151. Wang Q, Xu Y, Zhou W, *et al.* The viral oncoprotein HBx of Hepatitis B virus promotes the growth of hepatocellular carcinoma through cooperating with the cellular oncoprotein RMP. *Int J Biol Sci* 2014; **10**: 1181-1192.
 152. Su IJ, Wang LH, Hsieh WC, *et al.* The emerging role of hepatitis B virus pre-S2 deletion mutant proteins in HBV tumorigenesis. *J Biomed Sci* 2014; **21**: 98.
 153. Yang JC, Teng CF, Wu HC, *et al.* Enhanced expression of vascular endothelial growth factor-A in ground glass hepatocytes and its implication in hepatitis B virus hepatocarcinogenesis. *Hepatology* 2009; **49**: 1962-1971.
 154. Lee H, Kim H, Lee SA, *et al.* Upregulation of endoplasmic reticulum stress and reactive oxygen species by naturally occurring mutations in hepatitis B virus core antigen. *J Gen Virol* 2015; **96**: 1850-1854.
 155. Hsieh YH, Su IJ, Wang HC, *et al.* Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage. *Carcinogenesis* 2004; **25**: 2023-2032.
 156. Sidhu K, Kapoor NR, Pandey V, *et al.* The "Macro" World of microRNAs in Hepatocellular Carcinoma. *Front Oncol* 2015; **5**: 68.
 157. Oehler N, Volz T, Bhadra OD, *et al.* Binding of hepatitis B virus to its cellular receptor alters the expression profile of genes of bile acid metabolism. *Hepatology* 2014; **60**: 1483-1493.
 158. Watanabe Y, Yamamoto H, Oikawa R, *et al.* DNA methylation at hepatitis B

- viral integrants is associated with methylation at flanking human genomic sequences. *Genome Res* 2015; **25**: 328-337.
159. Sung WK, Zheng H, Li S, *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012; **44**: 765-769.
160. Jain S, Chang TT, Chen S, *et al.* Comprehensive DNA methylation analysis of hepatitis B virus genome in infected liver tissues. *Sci Rep* 2015; **5**: 10478.

Table 1A. Twenty-five significant HCC-associated SNVs with their cut-off values of SNV frequency ($f_{m/quasi}$) among genotype B HBV infected patients through NGS analysis.

Site	NT	Median SNV frequency (f _{m/quasi} , %)		Odds ratio ^b	Cut-off	Proportion of patients with SNV		P	False discovery rate (FDR)			
		(10 th , 25 th , 75 th , and 90 th percentiles)				frequency ³ ≥ cut-off value (%)						
		HCC group	Non-HCC group			(95%CI)	value (%)		HCC group	Non-HCC group	value	Corrected
		(N=40)	(N=47)						(N=40)	(N=47)	threshold	Decision
53	C	38 (0.0, 0.0, 100,100)	0.0 (0.0, 0.0, 23, 100)	4.5 (1.8, 11.5)	32	55	21.3	0.0025	0.008	Significant		
216	C	0.0 (0.0, 0.0, 80, 100)	8.8 (0.0, 0.0, 55, 82)	5.6 (1.2, 28.3)	100	20	4.3	0.0391	0.046	Significant		
273	G	21 (0.0, 0.0, 100, 100)	0.0 (0.0, 0.0, 5.6, 100)	5.0 (2.0, 12.8)	12.2	57.5	21.3	0.0012	0.004	Significant		

273	A	79 (0.0, 0.0, 100, 100)	100 (0.0, 94, 100, 100)	0.3 (0.1, 0.8)	55	42.5	19.1	0.0326	0.044	Significant
529	G	20 (0.0, 0.0, 100, 100)	3.7 (0.0, 0.0, 18, 100)	4.1 (1.4, 12.0)	100	37.5	12.8	0.0149	0.022	Significant
530	A	100 (80, 100, 100, 100)	100 (32, 89, 100, 100)	4.2 (1.3, 14.0)	100	90	68.1	0.0186	0.028	Significant
724	C	100 (87, 95, 100, 100)	100 (81, 87, 100, 100)	4.0 (1.3, 12.0)	91.7	87.5	63.8	0.0138	0.02	Significant
1173	G	100 (97, 100, 100, 100)	100 (85, 100, 100, 100)	9.2 (1.1, 76.5)	94.1	97.5	80.9	0.0183	0.026	Significant
1221	C	0.0 (0.0, 0.0, 4.6, 45)	0.0 (0.0, 0.0, 0.0, 16)	4.1 (1.1, 14.1)	4.3	27.5	8.5	0.0244	0.038	Significant
1242	G	0.0 (0.0, 0.0, 62, 100)	0.0 (0.0, 0.0, 0.0, 71)	5.1 (1.8, 14.6)	5.5	42.5	12.8	0.0038	0.012	Significant
1359	A	0.0 (0.0, 0.0, 3.7, 75)	0.0 (0.0, 0.0, 0.0, 4.2)	3.6 (1.0, 12.5)	2.4	25	8.5	0.0447	0.048	Significant
1896	A	100 (0.0, 1.1, 100, 100)	65 (0.0, 0.0, 100, 100)	3.5 (1.1, 10.7)	96.5	57.5	23.4	0.0025	0.01	Significant

1899	A	0.0 (0.0, 0.0, 99, 100)	0.0 (0.0, 0.0, 0.0, 94)	5.8 (1.1, 31.0)	96.1	32.5	6.4	0.0021	0.006	Significant
1913	C	100 (58, 100, 100, 100)	100 (31, 86, 100, 100)	5.5 (1.1, 28.2)	94.3	62.5	97.9	0.0001	0.002	Significant
2095	G	100 (0.0, 0.0, 100, 100)	0.0 (0.0, 0.0, 100, 100)	3.2 (1.3, 8.0)	10.1	72.5	44.7	0.0165	0.024	Significant
2120	G	0.0 (0.0, 0.0, 65, 100)	0.0 (0.0, 0.0, 0.0, 67)	3.6 (1.0, 12.5)	67.2	25	8.5	0.0447	0.05	Significant
2213	G	0.0 (0.0, 0.0, 0.0, 85)	0.0 (0.0, 0.0, 0.0, 0.0)	9.8 (1.2, 83.1)	1.7	17.5	2.1	0.0217	0.032	Significant
2226	T	0.0 (0.0, 0.0, 0.0, 91)	0.0 (0.0, 0.0, 0.0, 0.0)	9.8 (1.2, 83.1)	1.7	17.5	2.1	0.0217	0.034	Significant
2227	G	0.0 (0.0, 0.0, 0.0, 91)	0.0 (0.0, 0.0, 0.0, 0.0)	9.8 (1.2, 83.1)	1.8	17.5	2.1	0.0217	0.036	Significant
2227	T	100 (0.0, 93, 100, 100)	100 (100, 100, 100, 100)	0.2 (0.1, 0.8)	95.5	25	6.4	0.0316	0.04	Significant
2441	C	0.0 (0.0, 0.0, 20, 100)	0.0 (0.0, 0.0, 0.0, 75)	11.5 (1.4, 96.5)	100	20	2.1	0.0102	0.016	Significant

2444	C	100 (68, 100, 100, 100)	100 (42, 80, 100, 100)	4.8 (1.6, 14.3)	92.6	85	57.4	0.0104	0.018	Significant
2525	T	0.0 (0.0, 0.0, 95, 100)	0.0 (0.0, 0.0, 14, 71)	7.5 (1.5, 36.7)	95.5	25	4.3	0.01	0.014	Significant
2583	G	0.0 (0.0, 0.0, 91, 100)	0.0 (0.0, 0.0, 14, 85)	6.5 (1.2, 32.3)	100	22.5	4.3	0.0201	0.03	Significant
2690	A	0.0 (0.0, 0.0, 12, 100)	0.0 (0.0, 0.0, 0.0, 8.6)	4.9(1.2, 19.3)	12	25	6.4	0.0316	0.042	Significant

Note: ^a SNV frequency ($f_{m/quasi}$) of a specific HBV SNV in one subject was counted as the frequency of a specific substitution within one subject, i.e. the number of reads with a specific substitution/the number of total reads covering that site in one sample. Data are expressed as median, 10th, 25th, 75th, and 90th percentiles. ^b Odds ratio (OR) of a specific HBV SNV for HCC was defined as: $OR = (HCC_E \times non-HCC_{NE}) / (HCC_{NE} \times non-HCC_E)$. The subscript _E means the number of patients with SNV frequency no less than the cut-off value; the subscript _{NE} means the number of patients with SNV frequency less than the cut-off value. ^c For each specific SNV, the cut-off value of SNV frequency which corresponds to the highest odds ratio was used for statistical analysis. *P* values of odds

ratios were calculated by Chi-square test. With respect to genotype B, HCC-associated SNVs discovered by NGS were determined after adjustment for multiple testing by Benjamini and Hochberg's false discovery rate (FDR) method for those HCC-associated variants evidenced by meta-analysis and those SNVs with P values of $OR < 0.05$ by NGS analysis.

Table 1B. Thirty-five significant HCC-associated SNVs with their cut-off values of SNV frequency ($f_{m/quasi}$) among genotype C HBV infected patients through NGS analysis.

Site	NT	Median SNV frequency (f _{m/quasi} , %)		Odds ratio ^b	Cut-off	Proportion of patients with		<i>P</i>	False discovery rate			
		(10 th , 25 th , 75 th , and 90 th percentiles)				SNV frequency ^a cut-off value			(FDR)			
						(95%CI)	value (%)				value	
		HCC group	Non-HCC group						HCC group	Non-HCC group		Corrected
									Decision			
		(N=53)	(N=61)			(N=53)	(N=61)		threshold			
53	C	0.0 (0.0, 0.0, 57, 91)	0.0 (0.0, 0.0, 6.2, 29)	5.1 (1.6, 16.7)	52.4	26.4	6.6	0.0045	0.0157	Significant		
293	G	3.6 (0.0, 0.0, 51, 100)	0.0 (0.0, 0.0, 6.1, 87)	2.9 (1.3, 6.4)	1.8	50.9	26.2	0.0117	0.0214	Significant		
446	G	0.0 (0.0, 0.0, 17, 87)	0.0 (0.0, 0.0, 0.0, 46)	5.6 (2.0, 15.2)	6.4	37.7	9.8	0.0009	0.0029	Significant		
456	G	0.0 (0.0, 0.0, 0.0, 21)	0.0 (0.0, 0.0, 0.0, 0.0)	6.0 (1.2, 29.3)	10.2	17	3.3	0.0226	0.0371	Significant		

633	A	0.0 (0.0, 0.0, 3.3, 50)	0.0 (0.0, 0.0, 0.0, 0.0)	5.6 (1.1, 27.5)	3.6	20.8	6.6	0.0295	0.0414	Significant
834	G	0.0 (0.0, 0.0, 95, 100)	0.0 (0.0, 0.0, 17, 52)	7.6 (2.1, 28.2)	76.9	28.3	4.9	0.0007	0.0014	Significant
1092	C	100 (95, 100, 100, 100)	100 (0.0, 100, 100, 100)	5.6 (1.2, 26.6)	50.1	96.2	82	0.0192	0.0329	Significant
1155	C	100 (98, 100, 100, 100)	100 (0.0, 100, 100, 100)	5.6 (1.2, 26.6)	87.1	96.2	82	0.0192	0.0343	Significant
1386	A	3.6 (0.0, 0.0, 87, 100)	0.0 (0.0, 0.0, 13, 83)	5.2 (1.1, 25.9)	96.7	15.1	3.3	0.0482	0.0486	Significant
1613	A	7.0 (0.0, 0.0, 75, 100)	0.0 (0.0, 0.0, 39, 81)	10.1 (1.2, 83.6)	86.7	18.9	3.3	0.0117	0.0229	Significant
1653	T	11 (0.0, 0.0, 90, 100)	0.0 (0.0, 0.0, 6.5, 100)	4.2 (1.3, 14.2)	60.8	28.3	11.5	0.0421	0.0471	Significant
1674	C	0.0 (0.0, 0.0, 38, 100)	0.0 (0.0, 0.0, 4.2, 54)	11.5 (1.4, 94.8)	55.8	18.9	4.9	0.0354	0.0443	Significant
1753	G	0.0 (0.0, 0.0, 0.0, 23)	0.0 (0.0, 0.0, 0.0, 12)	3.1 (1.1, 9.4)	100	32.1	9.8	0.0066	0.0186	Significant

1764	A	100 (0.0, 100, 100, 100)	97 (0.0, 7.8, 100, 100)	4.6 (1.7, 12.4)	100	69.8	41	0.0038	0.0143	Significant
1846	T	8.7 (0.0, 0.0, 52, 100)	0.0 (0.0, 0.0, 21, 94)	3.1 (1.2, 8.2)	21.1	35.8	13.1	0.0086	0.02	Significant
2201	T	100 (96, 100, 100, 100)	100 (0.0, 100, 100, 100)	3.7 (1.1, 11.9)	93.3	92.5	77	0.0376	0.0457	Significant
2573	C	0.0 (0.0, 0.0, 100, 100)	0.0 (0.0, 0.0, 0.0, 100)	4.2 (1.3, 14.2)	100	32.1	13.1	0.0269	0.0386	Significant
2594	A	0.0 (0.0, 0.0, 100, 100)	0.0 (0.0, 0.0, 3.9, 100)	4.0 (1.3, 12.4)	100	34	13.1	0.0154	0.0243	Significant
2708	G	0.0 (0.0, 0.0, 100, 100)	0.0 (0.0, 0.0, 0.0, 97)	5.8 (1.5, 22.1)	100	34	9.8	0.0035	0.0129	Significant
2840	T	14 (6.4, 10, 17, 22)	9.6 (0.0, 4.4, 15, 23)	8.3 (1.8, 38.3)	4.4	96.2	75.4	0.0028	0.01	Significant
2875	A	4.8 (0.0, 0.0, 95, 100)	5.7 (0.0, 0, 50, 93)	5.6 (1.7, 18.3)	94.8	28.3	6.6	0.0023	0.0086	Significant
2889	G	100 (87, 100, 100, 100)	100 (7.8, 79, 100, 100)	5.5 (1.8, 17.6)	85.3	92.5	68.9	0.002	0.0071	Significant

2901	T	100 (90, 100, 100, 100)	100 (73, 90, 100, 100)	4.1 (1.1, 15.4)	85.1	94.3	80.3	0.049	0.05	Significant
2931	C	100 (92, 100, 100, 100)	100 (76, 86, 100, 100)	9.1 (2.0, 41.6)	87.1	96.2	73.8	0.0014	0.0043	Significant
2988	C	100 (100, 100, 100, 100)	100 (88, 100, 100, 100)	9.0 (1.1, 73.6)	92.1	98.1	85.2	0.0191	0.0257	Significant
2989	A	100 (100, 100, 100, 100)	100 (88, 100, 100, 100)	9.0 (1.1, 73.6)	91.7	98.1	85.2	0.0191	0.0271	Significant
2997	T	100 (98, 100, 100, 100)	100 (89, 100, 100, 100)	9.0 (1.1, 73.6)	90.8	98.1	85.2	0.0191	0.0286	Significant
2998	C	100 (100, 100, 100, 100)	100 (89, 100, 100, 100)	9.0 (1.1, 73.6)	92.2	98.1	85.2	0.0191	0.03	Significant
3006	A	100 (100, 100, 100, 100)	100 (84, 98, 100, 100)	14.1 (1.8, 111.8)	94.2	98.1	78.7	0.0014	0.0057	Significant
3009	G	100 (97, 100, 100, 100)	100 (88, 100, 100, 100)	5.0 (1.0, 24.0)	92.9	96.2	83.6	0.0342	0.0429	Significant
3016	C	100 (100, 100, 100, 100)	100 (90, 100, 100, 100)	9.0 (1.1, 73.6)	92.9	98.1	85.2	0.0191	0.0314	Significant

3021	A	100 (100, 100, 100, 100)	100 (84, 100, 100, 100)	11.4 (1.4, 91.9)	92.6	98.1	82	0.0052	0.0171	Significant
3066	T	100 (94, 100, 100, 100)	100 (72, 94, 100, 100)	5.6 (1.2, 26.6)	86.8	96.2	82	0.0192	0.0357	Significant
3097	A	100 (95, 100, 100, 100)	100 (84, 100, 100, 100)	4.5 (1.2, 16.8)	93.5	94.3	78.7	0.0282	0.04	Significant
3120	G	0.0 (0.0, 0.0, 100, 100)	0.0 (0.0, 0.0, 19, 95)	4.8 (1.6, 14.4)	95.9	30.2	8.2	0.0033	0.0114	Significant

Note: ^a SNV frequency ($f_{m/quasi}$) of a specific HBV SNV in one subject was counted as the frequency of a specific substitution within one subject, i.e. the number of reads with a specific substitution/the number of total reads covering that site in one sample. Data are expressed as median, 10th, 25th, 75th, and 90th percentiles. ^b Odds ratio (OR) of a specific HBV SNV for HCC was defined as follows: $OR = (HCC_E \times non-HCC_{NE}) / (HCC_{NE} \times non-HCC_E)$. The subscript _E means the number of patients with SNV frequency no less than the cut-off value; the subscript _{NE} means the number of patients with SNV frequency less than the cut-off value. ^c For each specific SNV, the cut-off value of SNV frequency which corresponds to the highest odds ratio was used for statistical analysis. *P* values of

odds ratios were calculated by Chi-square test. With respect to genotype C, HCC-associated SNVs discovered by NGS were determined after adjustment for multiple testing by Benjamini and Hochberg's false discovery rate (FDR) method for those HCC-associated variants evidenced by meta-analysis and those SNVs with P values of $OR < 0.05$ by NGS analysis.

Figure Legends

Figure 1. HBV genome-wide characterization of substitutions in genotype B

(HCC, n=40; non-HCC, n=47). Comparing HCC and non-HCC groups, the distribution of total substitutions along the hepatitis B virus (HBV) genome and their frequencies within quasispecies were demonstrated. HBV genome: four HBV genes and their nucleotide positions; Sequencing coverage: the average sequencing coverage at each site; Mutation rate per 9-bp: proportions of nucleotides with SNVs in each 9-bp window shown as blue-red color scale at one base sliding step; Average SNV II frequency: the average values of SNV II frequency (SNV II $f_{m/quasi}$).; SNV II frequency in each patient: SNV II and their SNV frequency ($f_{m/quasi}$) at each site shown as blue-red color scale in each patient. SNV I was defined as the dominant variant of HBV in non-HCC group. SNV II was the variant other than SNV I at the same nucleotide position. Almost all SNV I have SNV frequency higher than 80% either in HCC or non-HCC group. H: HCC group. N: Non-HCC group.

Figure 2. Validation of NGS-defined HCC-associated SNVs by a systemic

literature review and meta-analysis. (A) The proportions of NGS-defined

HCC-associated SNVs among HBV variants with different levels of supporting evidence declined significantly with decreasing levels from Level A to Level D. One hundred and sixty-seven HBV variants, which had been studied previously, were categorized into 4 groups according to supporting evidence associated with HCC. The proportion of NGS-defined HCC-associated SNVs among each group declined significantly along with decreasing levels of supporting evidences from A to D (chi-square test for the 2x4 *contingency table*, $P < 0.0001$). A: HCC-associated HBV variants favored by meta-analysis with at least 4 studies; B: HCC-associated HBV variants favored by at least one study with fewer than 4 relevant studies; C: HBV variants not associated with HCC provided by all studies with fewer than 4 relevant studies; D: HBV variants not associated with HCC provided by meta-analysis with at least 4 studies. (B) **Odds ratios of NGS analysis in 12 HCC-associated HBV variants with Level A evidence.** Level A included HCC-associated HBV variants evidenced by meta-analysis with at least 4 studies. Nine of them were identified by NGS analysis. * 1766T and 1768A mainly expressed in genotype A and D HBV and were not identified as HCC-associated SNVs by this NGS analysis. † 1762T could be identified as HCC-associated SNVs in a subgroup analysis of HBeAg-positive

patients with genotype C HBV infection by NGS analysis (odds ratio: 6.3, 95% confidence interval [CI] = 1.1 to 34.0). (C) **The distribution of cut-off values of HCC-associated SNV frequency ($f_{m/quasi}$) among the genotype B and genotype C groups.** Red circles represent SNVs reported to be associated with HCC. Yellow circles represent SNVs investigated in only one study, but with a negative result. Empty circles represent SNVs not investigated previously. The SNV, nt53C, was the risk variant for HCC in both genotypes B and C groups.

Figure 3. Genome-wide deletion distribution of HBV comparing HCC and non-HCC patients. Y axis represents the proportion of patients with deletion for each nucleotide or amino acid among patients with any deletion mutation of HBV full genome. The number of patients with any deletion was 10 in HCC group and 8 in non-HCC group for genotype B, while 34 in HCC group and 32 in non-HCC group for genotype C, respectively. Inverted Y axis represents the average deletion frequency ($f_{del/quasi}$) for each nucleotide or amino acid among patients with any deletion mutation of HBV full genome. (A) Full genome of genotype B HBV; (B) Full genome of genotype C HBV. The upper portions of (A) and (B) show ORFs for

all genes and 4 domains of the P gene. (C) PreS region of genotype B HBV; (D) PreS region of genotype C HBV. The upper portions of (C) and (D) show B and T cell epitopes and 5 functional sites (*S*, *T*, *N*, *M*, and *P*) among preS region. $f_{\text{del/quasi}}$: the deletion frequency within one subject, i.e. the number of reads with deletion/ the number of total reads covering that site in one sample. TP: terminal protein of polymerase; SP: spacer domain of polymerase; RT: reverse transcriptase of polymerase; RH: RNase H of polymerase. S: S-promoter; T: topology (CAD or Hsc70 binding site); N: nucleocapsid binding site; M: the start codon of M protein; P: pHSA site.

Figure 4. Ingenuity Pathway Analysis of microarray data and the proposed model for transformation of the Huh7 cell line with different HBV small S mutant expression. (A) Biological functions in transfected Huh7 cell with different three genotype B HBV small S mutants (T216C, A273G and T216C/A273G double mutants); (B) Biological functions in transfected Huh7 cell with different seven genotype C HBV small S mutants (A293G, C446G, A456G, A293G/C446G, C446G/A456G, A293G/A456G, and A293G/C446G/A456G triple mutants). Upper:

Key functional differences in molecular and cellular functions; Middle: Key functional differences in physiological system development and function; Lower: Key functional differences in disease and disorders. Bars indicate significance. The length of the bar indicates only that the differentially expressed proteins are related to this pathway but does not indicate upregulation or downregulation of the pathway.

Figure 5. Overview of HBV variants in hepatocarcinogenesis.

Figure 1.tif

Figure 2.tif

Figure 3.tif

Figure 4.tif

Figure 5.tif

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Liu, W-C;Wu, I-C;Lee, Y-C;Lin, C-P;Cheng, J-H;Lin, Y-J;Yen, C-J;Cheng, P-N;Li, P-F;Cheng, Y-T;Cheng, P-W;Sun, K-T;Yan, S-L;Lin, J-J;Yang, J-C;Chang, K-C;Ho, C-H;Tseng, VS;Chang, BC-H;Wu, J-C;Chang, T-T

Title:

Hepatocellular carcinoma-associated single-nucleotide variants and deletions identified by the use of genome-wide high-throughput analysis of hepatitis B virus.

Date:

2017-10

Citation:

Liu, W. -C., Wu, I. -C., Lee, Y. -C., Lin, C. -P., Cheng, J. -H., Lin, Y. -J., Yen, C. -J., Cheng, P. -N., Li, P. -F., Cheng, Y. -T., Cheng, P. -W., Sun, K. -T., Yan, S. -L., Lin, J. -J., Yang, J. -C., Chang, K. -C., Ho, C. -H., Tseng, V. S., Chang, B. C. -H. ,... Chang, T. -T. (2017). Hepatocellular carcinoma-associated single-nucleotide variants and deletions identified by the use of genome-wide high-throughput analysis of hepatitis B virus.. J Pathol, 243 (2), pp.176-192. <https://doi.org/10.1002/path.4938>.

Persistent Link:

<http://hdl.handle.net/11343/293330>