Introduction

Hepatitis B virus (HBV) infection leads to a wide spectrum of liver injuries, ranging from acute self-limiting infection and fulminant hepatitis to chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (HCC) [1]. The viral genome contains four partially overlapping open reading frames (ORFs) (pre-S/S, pre-C/C, P and X) [2]. HBV infection is usually diagnosed when the circulating HBV surface antigen (HBsAg) is detected. Spontaneous loss of HBsAg is a rare event in chronic hepatitis B (CHB) infection [3–6]. Loss of serum HBsAg is also observed in patients receiving treatment for CHB with interferon and nucleoside/nucleotide analogues [7,8]. Although the continued presence of detectable HBsAg in serum is a prominent feature of HBsAg-positive CHB infection or overt CHB infection, HBsAg seroclearance may not signify eradication of HBV. The introduction of sensitive HBV DNA detection tests has revealed the existence of HBsAg-negative HBV infection. This new entity of HBV infection is defined as “the presence of viral DNA in the liver (with detectable or undetectable HBV DNA in the serum) of individuals testing negative for the HBsAg” [9,10]. Some of these subjects may be people with or without prior medical history of HBV infection can develop to the occult phase [11]. All these scenarios are collectively termed occult hepatitis B infection (OBI) [10].

Clinical evidence reveals that there is generally no reduction in the risk for HCC in CHB patients with HBsAg seroclearance.
Materials and Methods

Patients and Samples
A total of 185 serum samples (156 from CHB patients with detectable HBsAg at the first presentation and 60 healthy controls) were collected from HBsAg-positive treatment-naive CHB patients and healthy subjects, respectively. The serum samples were used for the study after informed consent was obtained from all the patients.

Rolling Circle Amplification and Sequencing of Full-Length HBV Genome
RCA, a powerful PCR-based technique for the amplification of low viral load circular DNA, has been previously described [24]. Briefly, HBV plus-strand DNA inside the virions in the serum was used for the endogenous HBV polymerase before extraction using the QIAamp MinElute Virus Vacuum Kit (Qiagen). Nicks in the HBV DNA strands were then completed by ligase (Epicentre, Madison, WI). For the liver biopsy samples, total DNA was extracted using the QIAamp AllPrep Kit (Qiagen). The HBV DNA samples were subjected to RCA using eight primers spanning the whole HBV genome (Table S1) [24]. Single-genome-length HBV DNA was either retrieved directly by SpeI digestion (New England Biolabs) or further amplified using the Expand High Fidelity PCR System (Roche, Mannheim, Germany) [25]. This amplified HBV DNA was then purified and sequenced using sequencing primers that cover the whole viral genome (Table S1). Sensitivity and conditions for RCA reaction were tested against samples with known copies of viral DNA (quantified by the Artus end-time PCR kit; Qiagen).

Nucleotide Sequence and Amino Acid Analysis
Full-length HBV sequences were assembled using the CLC Main Workbench 6.6.2 (CLC Bio, Aarhus, Denmark). HBV genotype was determined using the NCBI genotyping tool and phylogenetic analysis. HBV sequences from CHB subjects were compared with each other from the control subjects or genotype-matched HBV sequences of Chinese CHB patients retrieved from the GenBank [26]. Analysis of nucleotide and amino acid diversity (d), as well as phylogenetic analysis, was performed using the MEGA version 5 software [27]. The program SimPlot [28] was used to search for evidence of recombination. Amino acid mutation frequency was defined as the number of amino acid variations per total number of amino acid residues within a particular genomic region. Prediction of RNA secondary structure was performed using the MFOLD software [29].

In-vitro Analysis on the Effects of Genetic Mutations on Specific Regulatory Regions on HBV Promoter Activities
HBV promoter regions from selected OBI cases and wide-type controls were amplified using primers containing KpnI and NheI restriction sites. The KpnI- and NheI-digested PCR products, containing different mutated HBV promoter regions, were cloned into the reporter vector pGL3-Basic. Positive constructs were confirmed and co-transfected with pRL-TK Renilla reporter plasmid into Huh-7 cells using Lipofectamine 2000 (Invitrogen). After 24 hours of incubation, the cells were assayed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI). Promoter activities were expressed as a ratio of firefly luciferase to Renilla luciferase luminescence. Results were expressed in arbitrary units (AU).

Statistical Analyses
Differences between categorical variables were analyzed using the Fisher’s exact test or Chi-square test. For continuous variables, the Student’s t-test was used. All statistical analysis was done using GraphPad Prism 5.0 (GraphPad Software, Inc. San Diego, CA). Data are expressed as percentage or mean ± SD. A p-value of < 0.05 was considered to be statistically significant.

Results
Rolling Circle Amplification for HBV Genomes from OBI
By using known copies of HBV DNA isolated from CHB patients, we demonstrated RCA could amplify down to 15 copies/reaction (Figure 1). Full-length HBV genomes were successfully amplified from 18/60 (30%) liver biopsies and 4/185 (2.2%) serum samples from OBI subjects using this RCA method. Complete HBV genomes were also amplified from all 11 controls. Seventeen out of these 18 OBI subjects had detectable intraha-
patic HBV DNA levels (median: 4.07 copies/cell; range: 0.07–15.34 copies/cell). HBV viral load in the 4 serum samples were all under the detection limit of COBAS assay of 20 IU/ml. Full-length HBV genomes were also amplified from 11 HBsAg-positive controls (median: 3.51 × 10^5 IU/ml; range: 100–7.34 × 10^7 IU/ml).

Demographic Data and Phylogenetic Analysis
Demographic data of the study subjects are summarized in Table 1. Genotype C was dominant in both OBI group (17/22; 77.3%; 11 with subtype C1, 4 with subtype C2 and 2 outliers) and controls (9/11; 81.8%; all subtype C1). Genotype B was detected in the remaining OBI subjects (5/22; 22.7%; all subtype B2) and controls (2/11; 18.2%; 1 with subtype B2 and 1 outlier) (Figure S1). Evidence of intergenotypic recombination between genotypes B/C was detected in these 3 outliers but only constituted less than 20% of the major parental genotype. The full-length HBV genomes and multiple sub-genomic regions reproduce the same genotyping results, these 3 outliers (2 occult samples and 1 control) were clustered into genotype C and B respectively. Of note, phylogenetic analysis based on the full-length HBV sequences or the 4 individual ORFs cannot distinctly separate OBI from CHB controls (data not shown).

Comparison of HBV Genomic Diversity between OBI and Control Subjects
The nucleotide and amino acid diversity were evaluated with respective to their genotypes. The nucleotide diversity over the entire HBV genome with genotype C was significantly higher in the 17 subjects with OBI when compared with the 9 control subjects (diversity, d = 0.040 ± 0.002 vs. 0.026 ± 0.002, p = 0.008). The nucleotide diversity in these 17 OBI subjects was also significantly higher than that in 49 genotype C reference sequences retrieved from Chinese CHB patients [26] (d = 0.04 ± 0.002 vs. d = 0.030 ± 0.001, p = 0.025). There was no significant difference in the nucleotide diversity between OBI subjects and 5 genotype B reference sequences retrieved from Chinese CHB patients showed no significant difference (d = 0.016 ± 0.002 vs. d = 0.015 ± 0.002, p = 0.792) [30] (Table 2). In terms of amino acid diversity, there

Figure 1. Rolling circle amplification (RCA) of full-length HBV genome. The number of initial HBV DNA template present in each RCA reaction is shown at the top. M, molecular weight marker; H2O, negative control. (A) The high-molecular-weight raw RCA products containing multiple copies of the initial HBV template. (B) The full-length HBV genomes recovered by restriction enzyme, SpeI digestion. (C) Full-length HBV genome amplified by using the RCA products as PCR template.
doi:10.1371/journal.pone.0099028.g001
was no significant difference between OBI and control sequences for both genotypes (data not shown).

Further analysis on the nucleotide and amino acid diversity on the 4 individual HBV ORFs indicated OBI cases presented slightly higher nucleotide and amino acid diversity than control cases (Table 2). For genotype C cases, the nucleotide diversity in the pre-S1, pre-C and P ORFs was significantly higher in OBI than control cases \( (p = 0.048, p = 0.047 \text{ and } p = 0.032, \text{ respectively}) \). For genotype B cases, OBI subjects had a significantly higher nucleotide diversity in the pre-C/C and C regions \( (p = 0.045 \text{ and } p = 0.031, \text{ respectively}) \) (Table 2). There was no significant difference on the amino acid diversity between OBI and control sequences on the 4 ORFs in both genotypes (data not shown). In summary, OBI

| HBV regions       | Genotype B           | Genotype C           |
|-------------------|----------------------|----------------------|
|                   | OBI                  | Control              | P-value | OBI                  | Control              | P-value |
| Pre-S/S           | 0.008±0.002          | 0.006±0.002          | 0.490    | 0.030±0.003          | 0.018±0.003          | 0.069  |
| Pre-S1            | 0.009±0.003          | 0.007±0.003          | 0.833    | 0.053±0.007          | 0.022±0.005          | 0.048* |
| Pre-S2            | 0.011±0.005          | 0.002±0.002          | 0.320    | 0.032±0.008          | 0.024±0.009          | 0.603  |
| S                 | 0.007±0.002          | 0.006±0.002          | 0.706    | 0.020±0.003          | 0.015±0.003          | 0.461  |
| Pre-C/Core        | 0.028±0.005          | 0.009±0.003          | 0.045*   | 0.041±0.004          | 0.024±0.004          | 0.059  |
| Pre-C             | 0.017±0.010          | 0.018±0.011          | 0.924    | 0.043±0.013          | 0.014±0.008          | 0.047* |
| Core              | 0.030±0.005          | 0.008±0.003          | 0.031*   | 0.041±0.004          | 0.025±0.004          | 0.099  |
| P                 | 0.013±0.002          | 0.014±0.002          | 0.911    | 0.039±0.003          | 0.026±0.002          | 0.032* |
| X                 | 0.016±0.004          | 0.036±0.006          | 0.990    | 0.027±0.004          | 0.018±0.004          | 0.267  |
| Pre-S1 promoter   | 0.027±0.005          | 0.011±0.003          | 0.110    | 0.055±0.006          | 0.034±0.005          | 0.097  |
| Pre-S2/S promoter | 0.010±0.004          | 0.008±0.003          | 0.684    | 0.052±0.007          | 0.020±0.005          | 0.037* |
| Full-length HBV   | 0.021±0.003          | 0.016±0.002          | 0.279    | 0.040±0.002          | 0.026±0.002          | 0.008* |

\(^{a}\)Full-length HBV: OBI sequence compared with 5 genotype B and 49 genotype C reference sequence obtained from Chinese patients with chronic HBV infection. \(^{*}\)A statistical difference with \( p<0.05 \).

doi:10.1371/journal.pone.0099028.t002
cases had a higher nucleotide diversity than control cases, and these changes were scattered over the entire viral genome and may not lead to coding changes. This was evidenced by the comparable amino acid diversity between the two groups.

### Mutational Analysis on the pre-S/S Coding Region

Details of the mutations observed in subjects with OBI and CHB infection were illustrated in Table S2. The total number of amino acid variations over the entire PreS/S region was significantly higher in OBI than control cases [89/400 (22.2%) vs. 33/400 (8.25%), \( p = 0.0001 \)] (Figure 2). The number of amino acid substitutions in individual pre-S1, pre-S2 and S regions were also significantly higher in OBI than control cases [pre-S1: 21/119 (17.6%) vs. 3/119 (2.5%), \( p = 0.0001 \); pre-S2: 20/55 (36.4%) vs. 7/55 (12.7%), \( p = 0.001 \); S: 48/226 (21.2%) vs. 23/226 (10.2%), \( p = 0.001 \)], respectively (Figure 2B). Some mutation patterns were uniquely observed in OBI cases in pre-S/S ORF, including sequence deletions (3/22, 13.6%) and abolishment of the pre-S2 start codon (ATG) by a point mutation (3/22, 13.6%). The analysis of S ORF showed that the clinically important amino acid substitutions were mainly located in the major hydrophilic region (MHR) (residues 103–173). These included I126S, Q129N, and M133T. I126S, Q129N, and M133T are associated with diagnostic problems [15,31,32], whereas G145A is known as vaccine escape mutant and associated with OBI [33].

### Mutational Analysis on the pre-C/C Coding Regions

The number of amino acid substitutions in OBI cases was significantly higher than control cases in both pre-C and core ORFs (pre-C: 18/29 (62.1%) vs. 3/29 (10.3%); core: 99/183 (54.1%) vs. 37/183 (20.2%), both with \( p = 0.0001 \)) (Figure 3A, 3B and Table S2). Within the pre-C region, the G1896A mutation introducing a stop signal at codon 28 (W28*) which abolishes HBeAg production was found in 7/22 occult cases and 1/11 control cases (31.8% vs. 9.1%, \( p = 0.2176 \)) (Figure 3B). Mutations in core region were mainly presented within different antigen immunogenic epitopes. As shown in Figure 3B, OBI cases had a higher number of amino acid substitutions than control cases within the CD8+ CTL-epitope (residues 18–27) [6/10 (60%) vs. 1/10 (10%), \( p = 0.0573 \)] and CD4+ T-cell epitope (residues 48–69) [9/22 (40.9%) vs. 3/22 (13.6%), \( p = 0.0883 \)]. Similarly, multiple amino acid substitutions were also detected within the HbcAg domain (residues 80–120). In particular, 9 amino acid changes (S74G, E77D, E77Q, P79Q, A80T, E83D, L84A, S87G and S87R) within the Hbc/e1 epitope (residues 75–90) were solely detected in OBI cases (56.3% vs. 0%, \( p = 0.015 \)), and that was known to reduce both HBe and Hbc antigenicity [34]. In addition, amino acid substitutions within the Hbc/e2 epitope (residues 120–

---

**Figure 2. Schematic representation showing types and positions of mutations affecting the HBV pre-S/S region.** (A) Binding sites in the pre-S1 and pre-S2/S promoters for the liver-specific transcription factors (HNF1 and HNF3), Sp1, TATA binding protein (TBP) and the CCAAT box. Distinct mutations in TF binding sites (broken arrows), premature termination of surface protein synthesis (*), nucleotide duplication (broken arrows), deletion (a flash) and abolishment of pre-S2/S start codon (X) were indicated. (B) Amino acid changes in the pre-S1, pre-S2 and S genes. (Position with more than one amino acid changes was bracketed. Amino acid changes in “a” determinant region were underlined).

doi:10.1371/journal.pone.0099028.g002
in OBI cases were also higher than controls [5/21 (23.8%) vs. 2/21 (9.52%), p = 0.041].

Mutational Analysis on the X Coding Region
The total number of amino acid substitutions in the X ORF was significantly higher in subjects with OBI than controls [83/154 (53.9%) vs. 56/154 (36.4%), p < 0.005]. A number of amino acid mutations were solely detected in subjects with OBI, including G22S (13.6%), P38S (13.6%), A102T (13.6%) and K118T (13.6%). There was a trend of a higher frequency of H86R mutation detected in OBI than in control cases [12/22 (54.5%) vs. 2/11 (9.1%), p = 0.067].

Mutational Analysis on the P Coding Region
The total number of amino acid substitutions over the P ORF was significantly higher in OBI than control cases [197/843 (30.1%) vs. 63/843 (7.6%), p < 0.0001] (Table S2). Deletions in the pre-S/S region which lead to deletions and truncated proteins in the overlapping P coding region were found in 3/22 (13.6%) OBI cases. 3 out of 6 (50%) mutations detected in “a” determinant region of HBsAg also caused mutation in overlapping rt region, these included sI126S to rtD134E, sQ129N to rtS137Q and sT131N to rtN139K (Figure 2B). Nucleotide T128A mutation in the finger domain of the rt was uniquely detected in 3/22 (13.6%) OBI cases. It is speculated that rtT128A mutation might result in defective replication activity [35].

Distinct Mutations in the Regulatory Regions
The mutation frequency within the HBV key regulatory regions was comparable between the OBI group and control group (Table 1). However, a significant difference in the nucleotide diversity was noticed over the pre-S2/S promoter when comparing the OBI with the control groups with genotype C (p = 0.037; Table 2). Point mutations were the most noticeable changes that

Figure 3. Schematic diagrams showing types and positions of the mutations affecting the overlapping HBV key regulatory regions. (A) The overlapping regions of HBV genome contain core promoter (CP), basal core promoter (BCP), enhancer II (Enh II), DR1 (nt 1824–1834) and DR2 (nt 1590–1600) and coding regions of preC/C and X proteins. Clinically important mutations in BCP that results in the alteration of HNF1 binding site and the corresponding amino acid changes in the overlapping X gene. (aa: amino acid, nt: nucleotide, DR: direct repeat). Distinct mutations in TF binding sites, premature termination of protein synthesis, nucleotide duplication and deletion were indicated. Percentage of X mutants in overlapping regions in occult and control cases were listed. (B) Amino acid changes in the pre-C and core genes. (Position with more than one aa changes was bracketed). doi:10.1371/journal.pone.0099028.g003
scattered over these key regulatory elements. Most of them result in interference with liver specific transcription factor (TF) binding sites and were more frequently detected in OBI cases than controls (Figure 2A, Figure 3A and Table S3). Large deletions at nt position 3127–55, 728–1255 and 1754–1771 were found in 3/22 (13.6%) OBI cases. These deletions led to loss of pre-S2/S and X promoters and disruption of BCP promoters, respectively. It is noteworthy that unique nucleotide duplications at the promoter regions were observed in 2/22 (9.1%) OBI cases. These unique nucleotide duplications included a 21 nt duplication (nt 3107–3127) in the Sp1 binding sites within the pre-S2/S promoter and a 31 nt duplication (nt 1644–1674) that interrupted the essential element box β in enhancer II (Table S3) [36].

Recent studies suggested that pre-S2/S mRNA splicing is essential for HBsAg expression [20,22]. This RNA splicing event is controlled by a 5’9 splice donor site (nt 426–464) and the post-transcriptional element (PRE) that contains the 3’9 splice acceptor site. Nucleotide G458A mutation was found to influence HBV RNA splicing and HBsAg production [20]. No G458A mutation was detected in this study. One OBI case with both T429C and T441G mutations in the 5’ splice donor site, which were predicted to affect mRNA secondary structure by MFOLD (Figure 4) [20]. The influence of RNA secondary structure on the activity of a 5’ splice site of the S mRNA and optimal production of HBsAg needs further functional analysis.

Functional Analysis of the Distinct Nucleotide Alterations in Regulatory Elements

We next explored whether the distinct surface promoter and core promoter (CP) mutations identified in OBI cases caused a change in their level of transcription activity. Based on the findings of the mutations in the regulatory regions in OBI cases, 5 mutants and two wild-type controls were constructed. These included T2768A mutation located in Sp1 binding sites of pre-S1 (MUT1), 21 nt duplication between 3128–3148 on pre-S2/S (MUT2), C3015T mutation on preS2/S (MUT3) and two mutations G1677A (MUT4) and C1706T (MUT5) in CP. Luciferase activities of the plasmids containing these mutant promoters (MUT1–5) were compared to that of their corresponding wild-type promoter. The absolute luciferase activity of the control wild-type promoter was set to 100% and all the other mutant constructs were compared accordingly. As shown in Figure 5, mutation significantly affected the promoter activity in MUT2 (85%, p < 0.05), but not in MUT1 (92%, p = 0.226) and MUT3 (87%, p = 0.156). The promoter activity was significantly decreased in MUT4 (72%, p < 0.05) and MUT5 (40%, p < 0.005) with respective to the control CP.

Discussion

We performed mutational analysis of the RCA-amplified full-length HBV genomes from OBI and control cases. In order to exclude natural polymorphism and/or differences related to the geographical origin of the patients, it is imperative to have a robust comparison of sequences obtained from cases with the same genotype and origin. In this sense, we compared OBI sequences with both genotype-matched controls and reference sequences obtained from 49 genotype C (39 with sub-genotype C1 and 10 with sub-genotype C2) [26] and 5 genotype B2 CHB patients from Hong Kong [30].

Our full-length HBV sequence analysis revealed that HBV nucleotide diversity in OBI cases was significantly higher than control cases, indicating that more sequence variations were

Figure 4. Prediction of RNA secondary structure at the 5’ splice site of occult hepatitis B virus infection (OBI) and control HBV strain. (A) RNA structure around the 5’ splice site at position 458 in control wild-type HBV strain (B) Predicted RNA structure of the OBI case with nucleotide T429C and T441G mutations. (Mutations are encircled and an arrow indicates the splice site).

doi:10.1371/journal.pone.0099028.g004
observed in OBI. Despite the strict criteria, a variety of mutations have been detected which are scattered throughout the viral genome in both occult and overt cases. Many of the mutations identified were in the surface proteins (Figure 2), which affect HBsAg detection assays, immune response recognition, HBV infectivity and virion morphogenesis [15,20,23]. However, we did not identify any single prevailing mutation or genetic signature which was associated with OBI. In this study, a mixture of mutation patterns like point mutations, deletions and nucleotide duplications were identified in OBI cases. This finding of multiple mutations, rather than prevailing mutations or genetic signatures, is responsible for the OBI status.

It is likely that multiple mechanisms are involved, and a combination of mutations, rather than any prevailing genetic signatures, is responsible for the OBI status.

Supporting Information

Figure S1 Phylogenetic analysis using the entire HBV genomes amplified from the 22 occult and 11 control HBV infected subjects. The filled circles represent the occult subjects, and the empty circles represent control subjects. Phylogenetic comparison was done by neighbor-joining algorithm based on Kimura two-parameter distance estimation. Bootstrap
values more than 75% are indicated on the major nodes. The scale of the evolutionary distances is shown at the bottom (scale bar). References sequences retrieved from GenBank are indicated by their accession numbers.

### Table S1 HBV-specific primers for rolling circle amplification (RCA), PCR amplification of full-length HBV genomes, sequencing and regulatory elements analysis.

### Table S2 Amino acid changes in HBV coding genes from subjects with occult and overt HBV infections.

#### References

1. Modla N, Kew M, Arbuthnot P (2002) Regulatory elements of hepatitis B virus transcription. J Viral Hepat 9: 323–331.

2. Tiofali P, Pourcel G, Dejean A (1985) The hepatitis B virus. Nature 317: 489–495.

3. Yuan MF, Wong DK, Fung J, Ip P, But D, et al. (2008) HBsAg Seroconversion in chronic hepatitis B Asian patients: replicative level and risk of hepaticocellular carcinoma. Gastroenterology 135: 1192–1199.

4. Da Silva LC, Madruga CL, Carillo FJ, Pinho JR, Sasa-Alequeuzar A, et al. (1996) Spontaneous hepatitis B surface antigen clearance in a long-term follow-up study of patients with chronic type B hepatitis. Lack of correlation with hepatitis C and D virus superinfection. J Gastroenterol 31: 696–701.

5. Liaw YF, Sheen IS, Chen TJ, Chu CM, Pao CC (1991) Incidence, determinants and significance of delayed clearance of serum HBsAg in chronic hepatitis B virus infection: a prospective study. Hepatology 15: 627–631.

6. Abadie WL, McMahan BJ, Hall DB, Heyward WL, Francisco DP, et al. (1985) The long-term serological course of asymptomatic hepatitis B virus carriers and the development of primary hepato cellular carcinoma. J Infect Dis 151: 604–611.

7. Lai CL, Shouval D, Lok AS, Chang TT, Cheinquer H, et al. (2006) Entecavir versus lamivudine for patients with HBsAg-negative chronic hepatitis B. N Engl J Med 354: 1011–1020.

8. Fattovich G, Giustina G, Sanchez-Tapias J, Quero C, Mas A, et al. (1998) Analysis of mutant surface promoter. Gastroenterology 127: 1356–1371.

9. Allain JP (2004) Occult hepatitis B virus infection: implications in transfusion. Vox Sang 86: 83–91.

10. Raimondo G, Allain JP, Brunetto MR, Buendia MA, Chen DS, et al. (2008) Hepatitis B virus: new evidence of occult infection. J Viral Hepat 15: 251–266.

11. Raimondo G, Caccamo G, Filonico R, Pollicino T (2013) Occult HBV infection. Semin Immunopathol 35: 39–52.

12. Huang WT, Wu JF, Lee CH, Yeh YH, et al. (2008) Serological evidence of occult hepatitis B virus infection in Chinese children with chronic liver disease: full-length genome and analysis of mutant surface promoter. Gastroenterology 132: 1156–1167.

13. Liu CL, Shouval D, Lok AS, Chang TT, Cheinquer H, et al. (2006) Entecavir versus lamivudine for patients with HBsAg-negative chronic hepatitis B. N Engl J Med 354: 1011–1020.

14. De Mitri MS, Cassini R, Bernardi M (2010) Hepatitis B virus-related hepatocarcinogenesis: molecular oncogenic potential of clear or occult infections. Eur J Cancer 46: 2178–2186.

15. Hou J, Wang Z, Cheng J, Lin Y, Luo G, et al. (2001) Prevalence of naturally occurring surface gene variants of hepatitis B virus in immunized surface antigen-negative Chinese carriers. Hepatology 34: 1027–1034.

16. Fang Y, Teng X, Xu WZ, Li D, Zhao HW, et al. (2009) Molecular characterization and functional analysis of occult hepatitis B virus infection in Chinese patients infected with genotype C. J Med Virol 81: 826–833.

17. Pollicino T, Amaducci G, Rostuccia A, Raffa G, Alibrandi A, et al. (2012) Impact of hepatitis B virus (HBV) preS/NS genomic variability on HBV surface antigen and HBV DNA serum levels. Hepatology 56: 434–443.

18. Chandnani V, Tayan K, Nayak B, Acharya SK, Panda SK (2004) Occult hepatitis B virus infection in chronic liver disease: full-length genome and analysis of mutant surface promoter. Gastroenterology 127: 1156–1167.

19. Rock CT, Tillmann HL, Manns MP, Trautwein C (1999) The pre-S region determines the intracellular localization and appearance of hepatitis B virus. Hepatology 30: 850–855.

20. Candoni D, Lin CK, Belkhir D, Sakuldamrongpanich T, Bissas S, et al. (2012) Occult hepatitis B infection in blood donors from South East Asia: molecular characterisation and potential mechanisms of occurrence. Gut 61: 1744–1753.

21. Allain JP, Belkhir D, Vermeulen M, Crookes R, Cable R, et al. (2008) Characterization of occult hepatitis B virus strains in South African blood donors. Hepatology 49: 1680–1686.

### Table S3 Nucleotide changes in key regulatory regions of the hepatitis B virus (HBV) genome.

#### Author Contributions

Conceived and designed the experiments: FYH DKHW JF CLL MFY. Performed the experiments: FYH WKS. Analysed the data: FYH AZY. Contributed reagents/materials/analysis tools: CK. Lee MFY. CK, Lin. Wrote the paper: FYH DKHW. Recruitment of study subjects: CLL MFY.