

## BRIEF COMMUNICATION

# Rapid identification of multi-strain HBV infection in patient by high-throughput DNA sequencing

Yongchang Zheng<sup>1,†</sup>, Masood ur Rehman Kayani<sup>2,†</sup> and Ting F. Zhu<sup>2,\*</sup>

<sup>1</sup> Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

<sup>2</sup> School of Life Sciences, Center for Synthetic and Systems Biology, MOE Key Laboratory of Bioinformatics, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Tsinghua University, Beijing 100084, China

\* Correspondence: tzhu@biomed.tsinghua.edu.cn

Received March 27, 2015; Revised May 24, 2015; Accepted May 25, 2015

**Hepatitis B is a well-known risk factor for the development of liver cancer and is closely associated with patient morbidity and mortality. Viral mutants and variants have the potential to evade immune response and prolong infection, and thus it is crucial to develop a methodology for the rapid identification of multi-strain hepatitis infections in patients. Here we describe a method based on selective region amplification of viral genome and deep sequencing, which may be used for rapid identification of multi-strain hepatitis B virus (HBV) infection in patients. The method works even with significantly low amounts of patients' serum samples, where the wet-lab procedures take about 1.5 days, followed by a quick bioinformatic analysis to reveal the final results. Our method can potentially be applied to the rapid and reliable identification of multi-strain HBV infection and help improve treatment regimens.**

**Keywords:** hepatitis B; liver cancer; high-throughput sequencing; single nucleotide polymorphism; bioinformatics

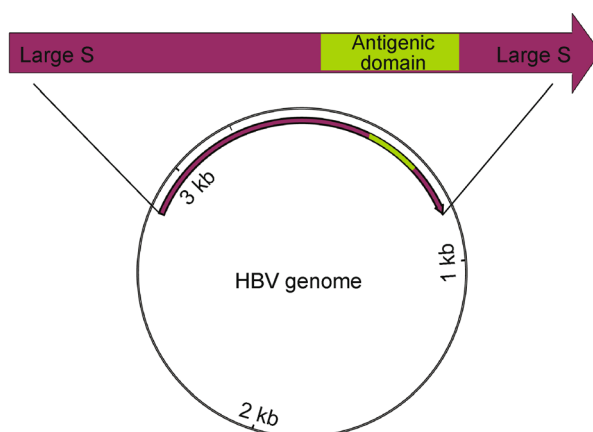
Hepatitis B virus (HBV) results in chronic and acute infections of liver and has affected hundreds of millions of people globally, including in China [1,2]. Approximately one million deaths per annum are caused by chronic hepatitis, hepatocellular carcinoma (HCC) and cirrhosis [3]. HCC is among the most prevalent tumors worldwide and HBV infection is its well-known risk factor [4]. The mechanism by which HBV causes liver cancer has not been well understood, but it may involve the viral biology itself and the subsequent cellular responses (e.g., inflammation) occurring after viral infection. Although current vaccination initiatives have lowered the rate of HBV infections in China, chronic infections still remain a major concern [5].

HBV is a small, non-cytopathic virus with a 3.2-kb DNA genome, which contains 4 overlapping genes: (i) the *PreS/S* gene that encodes the Hepatitis B surface antigen (HBsAg), (ii) the *PreC/C* gene that encodes for the nucleocapsid and e-antigen (HBeAg), (iii) the *P* gene that

encodes for polymerase, and (iv) the *X* gene [6]. The S domain in *PreS/S* gene plays a key role in morphogenesis, facilitating viral entry into the cells. Studies have shown that antibodies designed against S domain have the potential to neutralize HBV infections [7–9]. An “antigenic domain” (residues 100–170) is considered crucial for the presence of genotypic variants in HBV-positive patients [10] (Figure 1). The HBV DNA is replicated through an intermediate reverse transcription step, which generates genetic variations due to lack of proof-reading ability in the viral polymerase [11–13]. This ultimately leads to the generation of quasispecies pools in the host, which further aggravates to the emergence of mutants when selection pressure is exerted, especially by treatment procedures such as vaccination and administration of nucleoside analogs for inhibition of viral polymerase [12].

The current methodology used to confirm the presence of mutant strains in patients include quantitative

<sup>†</sup> These authors contributed equally to this work



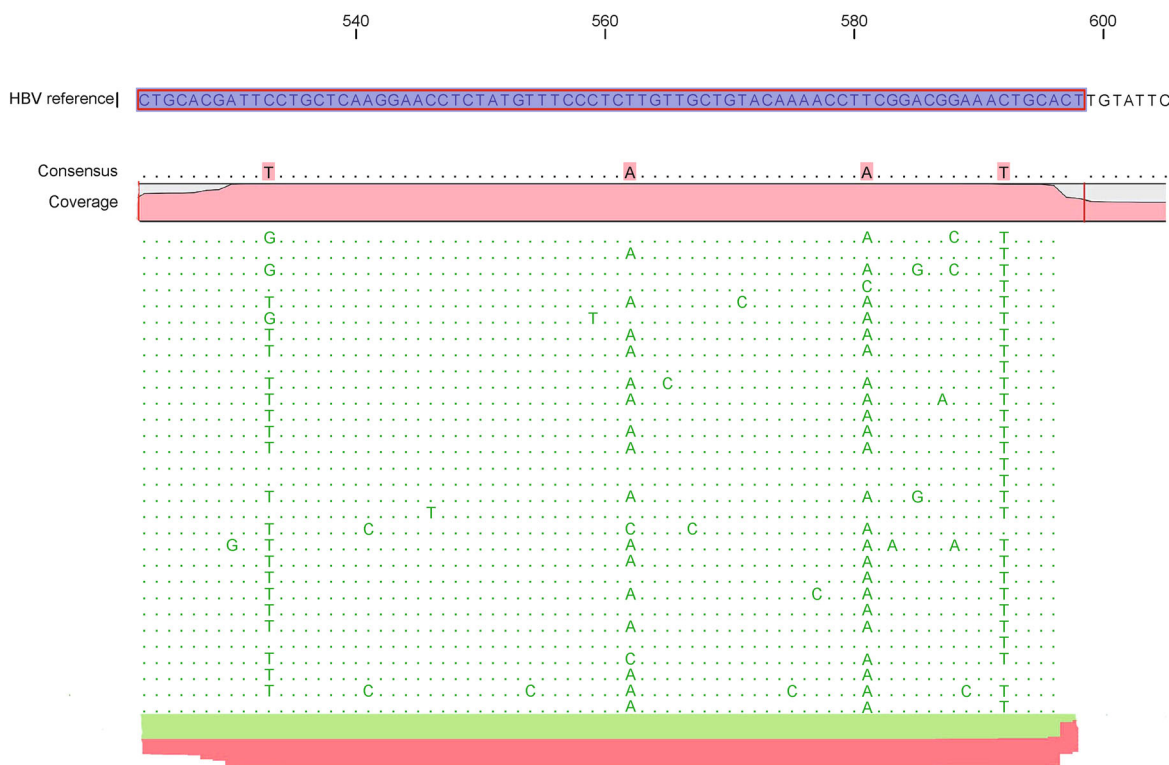
**Figure 1.** HBV genome, with the *PreS/S* gene and the antigenic domain.

polymerase chain reaction (qPCR) and restriction fragment length polymorphism analysis [14]. However, these immunoassays have a limited sensitivity and specificity, e.g., a particular assay can only be used to detect a specific mutant [15,16]. Therefore, it is instrumental to develop a reliable methodology for the identification of multiple HBV mutants and eliminate the need to carry out multiple immunoassays [17]. Here, based on selective region PCR and deep sequencing, we have developed a method for the rapid identification of multi-strain HBV infection in patients. The procedure requires approximately 1.5 days to complete and can potentially be used for the clinical identification of multi-strain HBV infection and help improve treatment regimens.

To demonstrate our method of rapid identification of multi-strain HBV infection in patients, we have obtained serum samples from a late-stage liver cancer patient with clinically suspected multiple re-infections of HBV, enrolled at the Peking Union Medical College Hospital, Beijing, China. A volume of ~5 mL whole blood was collected and the serum was separated by centrifugation at 1000  $\times$ g for 10 min. The serum was collected in aliquots of 500  $\mu$ L and was immediately stored at  $-80^{\circ}\text{C}$  until further use. The patient was informed about the purpose of study and a written consent was obtained before sample collection. DNA was extracted from the serum using the EZNA Viral DNA Kit (Omega Biotek, USA). In brief, 250  $\mu$ L serum was lysed by protease and acrylamide, treated with RNase-A, ethanol washed and eluted. DNA was quantified using Nanodrop 2000c (Thermo Scientific, USA). Serum separation and DNA extraction were performed in a BSL-2 laboratory. For PCR, the primers were designed to amplify the 220 bp antigenic domain region. Forward (5'-CTGGACTACCAAGGTATGT-3') and reverse (5'-GTAAACTGAGCCAGGAGA-3') primers were synthesized by Genewiz, China. PCR cycling

conditions included an initial denaturation at  $94^{\circ}\text{C}$  for 2 min, 30 cycles of denaturation at  $94^{\circ}\text{C}$  for 30 s, annealing at  $55^{\circ}\text{C}$  for 30 s and extension at  $72^{\circ}\text{C}$  for 1 min, and final extension at  $72^{\circ}\text{C}$  for 10 min. The reaction was performed in Veriti thermal cycler (Life technologies, USA) with a total volume of 50  $\mu$ L containing 5 ng of template DNA, 0.3  $\mu$ M of each primer, 200  $\mu$ M of each deoxynucleoside triphosphate (dNTP), 1U of Taq polymerase and 2 $\times$  PCR buffer (supplied with the polymerase). The amplified DNA was electrophoresed on a 1.5% agarose gel for about 30 min at 90V. After electrophoresis, the desired DNA fragment was excised and purified using EZNA gel extraction kit (Omega Biotek, USA) according to manufacturer's instructions.

The purified PCR product was further used for library preparation and high-throughput sequencing. Briefly, the sample was adapter ligated, amplified, size selected, and quantified. The library was sequenced using an Illumina MiSeq system. The barcoded sequencing data consisted of 1,298 paired-end reads with a read length of 150 bp. Data analysis was performed using CLC Genomics Workbench 7, involving the removal of adapter contamination from sequenced reads and trimming the low quality ends from the reads. The threshold for trimming was Q20. We next performed the alignment of the Illumina reads against the HBV reference genome (GenBank accession number NC\_003977) using a mapping quality score of 30. Duplicate sequences were excluded from further analysis. SNPs were also called with the alignment data using CLC Genomics Workbench 7. Per base quality score indicated that a high percentage of the sequences passed the PHRED Q20 and Q30 threshold, indicative of low sequencing error, while any low quality sequences were trimmed from the reads to improve the overall quality of alignment. The average sequence coverage was  $\sim 200\times$ , which provided a



**Figure 2. Alignment of the sequencing data with the HBV reference genome.** The resulting SNPs, the consensus sequence, and depth of coverage are also shown.

sufficient depth and hence increased reliability of the single-nucleotide polymorphism (SNP) call of the 220 bp antigenic domain (Figure 1).

The analysis revealed 4 SNPs, including 2 C to T mutations at position 533 and 592, and 2 T to A mutations at position 562 and 581 (Figure 2). These C to T and T to A mutations were among the high coverage polymorphisms, i.e., they were represented by at least 50% of the reads. One of these variations (position 592) is covered by more than 99% of the reads. Multiple other variations were also observed in these regions, which were relatively under-covered to be considered reliable (Table 1).

It is also noteworthy that mutation C533T substituted amino acid proline with a serine residue, a non-polar to polar change, which may result in structural changes in the antigenic domain. Furthermore, the T581A variation replaced a serine residue with threonine: a polar residue replaced by another polar residue. Two of the other SNPs, i.e., the T562A and C592T mutations did not induce any change in the serine and asparagine residues they code for. Taking into account the SNP data and possible mutations in the antigenic domain, it is reasonable to consider that the patient would have had up to 16 different strains of the virus, either through multiple infections or *de novo* mutation of the virus. Such multi-strain infection would

**Table 1. Four SNPs in the antigenic domain with total and SNP coverage.**

Position	Reference base	SNP	Total coverage	SNP coverage
533	C	T	309	192
562	T	A	312	162
581	T	A	312	233
592	C	T	306	305

be harder for the patient’s immune system and treatment regiments to tackle, making it possible for the virus to persist and cause other diseases.

Here we have developed a method for the rapid identification of multi-strain HBV infections and applied the method to analyze a potential multi-strain HBV infection case in a late-stage liver cancer patient. We amplified and sequenced a small genomic region from the *PreS/S* viral gene, which suggested the existence of genotypic viral variants in the patient. Our data indicated the presence of 4 SNPs, which suggested the possibility of multi-strain infection. These SNPs could arise due to various reasons, e.g., from re-infection and from the selective pressure exerted by the therapeutic regiments. For persistent infection, the virus would have to evade the

host immune system. Hence, variations are introduced to the genome, and the selection of variants favors successful replication cycles. We have amplified only a small region from the viral genome, which could be one of the reasons why we only observed 4 SNPs from sequencing. Sequencing the other genomic regions from the virus may result in an increased number of SNPs; yet, this would still suggest the presence of multiple strains of viruses in the patient. Thus for improving the feasibility and simplicity of the methodology, we sequenced only a small region from *PreS/S* gene. The method we reported here is fast, cost-effective, and may enable clinicians in the rapid identification of multi-strain HBV infection to develop appropriate treatments for chronic HBV infection in patients.

## ABBREVIATIONS

HBV, hepatitis B virus; SNP, single nucleotide polymorphism; HBsAg, Hepatitis B surface antigen; HCC, hepatocellular carcinoma

## ACKNOWLEDGEMENTS

This work was supported in part by funding from the Ministry of Science and Technology of China (973 grant No. 2015CB553402), National Natural Science Foundation of China (grant No. 31470532), the Tsinghua University-Peking University Center for Life Sciences (CLS), the Tsinghua-Janssen Scholarship and research grant, and the Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases.

## COMPLIANCE WITH ETHICS GUIDELINES

The authors Yongchang Zheng, Masood ur Rehman Kayani, and Ting F. Zhu declare they have no conflict of interest.

The study was approved by the Ethics Committee of the Peking Union Medical College Hospital. All patients and control subjects signed informed consent forms.

## REFERENCES

1. Lavanchy, D. (2005) Worldwide epidemiology of HBV infection, disease burden, and vaccine prevention. *J. Clin. Virol.*, 34, S1–S3
2. Zuckerman, J. N. and Zuckerman, A. J. (2000) Current topics in hepatitis B. *J. Infect.*, 41, 130–136
3. Ding, Y., Sheng, Q., Ma, L. and Dou, X. (2013) Chronic HBV infection among pregnant women and their infants in Shenyang, China. *Virol. J.*, 10, 17
4. Di Bisceglie, A. M. (2009) Hepatitis B and hepatocellular carcinoma. *Hepatology*, 49, S56–S60
5. Cui, Y. and Jia, J. (2013) Update on epidemiology of hepatitis B and C in China. *J. Gastroenterol. Hepatol.*, 28, 7–10
6. Tiollais, P., Pourcel, C. and Dejean, A. (1985) The hepatitis B virus. *Nature*, 317, 489–495
7. Glebe, D., Aliakbari, M., Krass, P., Knoop, E. V., Valerius, K. P. and Gerlich, W. H. (2003) Pre-s1 antigen-dependent infection of Tupaia hepatocyte cultures with human hepatitis B virus. *J. Virol.*, 77, 9511–9521
8. Iwarson, S., Tabor, E., Thomas, H. C., Goodall, A., Waters, J., Snoy, P., Shih, J. W. and Gerety, R. J. (1985) Neutralization of hepatitis B virus infectivity by a murine monoclonal antibody: an experimental study in the chimpanzee. *J. Med. Virol.*, 16, 89–96
9. Shearer, M. H., Sureau, C., Dunbar, B. and Kennedy, R. C. (1998) Structural characterization of viral neutralizing monoclonal antibodies to hepatitis B surface antigen. *Mol. Immunol.*, 35, 1149–1159
10. Carman, W. F., Karayiannis, P., Waters, J., Tanzi, E., Thomas, H. C., Zanetti, A. R., Manzillo, M. D. and Zuckerman, A. J. (1990) Vaccine-induced escape mutant of hepatitis B virus. *Lancet*, 336, 325–329
11. Lee, Y. I., Hur, G. M., Suh, D. J. and Kim, S. H. (1996) Novel pre-C/C gene mutants of hepatitis B virus in chronic active hepatitis: naturally occurring escape mutants. *J. Gen. Virol.*, 77, 1129–1138
12. Purdy, M. A. (2007) Hepatitis B virus S gene escape mutants. *Asian J. Transfus. Sci.*, 1, 62–70
13. Yamamoto, K., Horikita, M., Tsuda, F., Itoh, K., Akahane, Y., Yotsumoto, S., Okamoto, H., Miyakawa, Y. and Mayumi, M. (1994) Naturally occurring escape mutants of hepatitis B virus with various mutations in the S gene in carriers seropositive for antibody to hepatitis B surface antigen. *J. Virol.*, 68, 2671–2676
14. Coleman, P. F. (2006) Detecting hepatitis B surface antigen mutants. *Emerg. Infect. Dis.*, 12, 198–203
15. Kazim, S. N., Sarin, S. K., Sharma, B. C., Khan, L. A. and Hasnain, S. E. (2006) Characterization of naturally occurring and Lamivudine-induced surface gene mutants of hepatitis B virus in patients with chronic hepatitis B in India. *Intervirology*, 49, 152–160
16. Ijaz, S., Torre, F., Tedder, R. S., Williams, R. and Naoumov, N. V. (2001) Novel immunoassay for the detection of hepatitis B surface ‘escape’ mutants and its application in liver transplant recipients. *J. Med. Virol.*, 63, 210–216
17. Sheldon, J. and Soriano, V. (2008) Hepatitis B virus escape mutants induced by antiviral therapy. *J. Antimicrob. Chemother.*, 61, 766–768