Research Article

JOURNAL OF HEPATOLOGY

Novel pre-mRNA splicing of intronically integrated HBV generates oncogenic chimera in hepatocellular carcinoma

Yung-Tuen Chiu[1,2,†], John K.L. Wong[3,†], Shing-Wan Choi[3], Karen M.F. Sze[1,2], Daniel W.H. Ho[1,2], Lo-Kong Chan[1,2], Joyce M.F. Lee[1,2], Kwan Man[1,4], Stacey Cherny[3], Wan-Ling Yang[5], Chun-Ming Wong[1,2,*], Pak-Chung Sham[3,*], Irene O.L. Ng[1,2,*]

[1]*State Key Laboratory for Liver Research, The University of Hong Kong, Hong Kong;* [2]*Department of Pathology, The University of Hong Kong, Hong Kong;* [3]*Department of Psychiatry, The University of Hong Kong, Hong Kong;* [4]*Department of Surgery, The University of Hong Kong, Hong Kong;* [5]*Department of Pediatrics and Adolescent Medicine, The University of Hong Kong, Hong Kong*

Background & Aims: Hepatitis B virus (HBV) integration is common in HBV-associated hepatocellular carcinoma (HCC) and may play an important pathogenic role through the production of chimeric HBV-human transcripts. We aimed to screen the transcriptome for HBV integrations in HCCs.

Methods: Transcriptome sequencing was performed on paired HBV-associated HCCs and corresponding non-tumorous liver tissues to identify viral-human chimeric sites. Validation was further performed in an expanded cohort of human HCCs.

Results: Here we report the discovery of a novel pre-mRNA splicing mechanism in generating HBV-human chimeric protein. This mechanism was exemplified by the formation of a recurrent HBV-cyclin A2 (CCNA2) chimeric transcript (A2S), as detected in 12.5% (6 of 48) of HCC patients, but in none of the 22 non-HCC HBV-associated cirrhotic liver samples examined. Upon the integration of HBV into the intron of the *CCNA2* gene, the mammalian splicing machinery utilized the foreign splice sites at 282 nt. and 458 nt. of the HBV genome to generate a pseudo-exon, forming an in-frame chimeric fusion with *CCNA2*. The A2S chimeric protein gained a non-degradable property and promoted cell cycle progression, demonstrating its potential oncogenic functions.

Conclusions: A pre-mRNA splicing mechanism is involved in the formation of HBV-human chimeric proteins. This represents a novel and possibly common mechanism underlying the formation of HBV-human chimeric transcripts from intronically integrated HBV genome with functional impact.

Lay summary: HBV is involved in the mammalian pre-mRNA splicing machinery in the generation of potential tumorigenic HBV-human chimeras. This study also provided insight on the impact of intronic HBV integration with the gain of splice sites in the development of HBV-associated HCC.

Keywords: Splicing; HBV integration; Cyclin A2; Hepatocellular carcinoma.
Received 9 June 2015; received in revised form 28 January 2016; accepted 3 February 2016; available online 8 February 2016

* Corresponding authors. Addresses: Department of Pathology, The University of Hong Kong, Hong Kong (C.-M. Wong), or Department of Psychiatry, The University of Hong Kong, Hong Kong (P.-C. Sham), or Department of Pathology, The University of Hong Kong, Hong Kong (I.O.L. Ng).
E-mail addresses: jackwong@pathology.hku.hk (C.-M. Wong), pcsham@hku.hk (P.-C. Sham), iolng@hku.hk (I.O.L. Ng).
† The authors contributed equally as joint first authors.

Abbreviations: HBV, hepatitis B virus; HCC, hepatocellular carcinoma; NT-L, non-tumorous liver; A2S, HBV-CCNA2 chimeric transcript; D-box, destruction box; DR1, Direct Repeat 1; CDKs, cyclin-dependent kinases; CHX, cycloheximide; HBsAg, hepatitis B surface antigen; GSNAP, Genomic Short-read Nucleotide Alignment Program.

Introduction

Hepatocellular carcinoma (HCC) is a leading cancer worldwide and specifically prevalent in China and Southeast Asia [1,2], due to prevalent hepatitis B virus (HBV) infection. In most of the HCCs developed in chronic HBV carriers, the viral DNA sequence is clonally integrated into the host genome [3] and HBV integration into human genome is a major causative factor for HBV-associated HCC [4]. In some cases, insertion of the HBV promoter sequences into the host chromosomal DNA leads to abnormal or increased transcription of host sequences [5]. HBV integration was also found to disrupt chromosomal stability with subsequent chromosomal re-arrangements. It is also generally reported that truncation or alteration of host genes by HBV insertions generates novel fusion proteins with altered oncogenic properties [6,7]. Although HBV integration is shown to be involved in hepatocarcinogenesis, the role of HBV integration in HCC development has not been fully elucidated.

Recently, with the advances in next generation sequencing technologies [8], the preferential integration sites and functional impact of viral integration into the host genome can be studied in a more global and extensive manner. For instance, Sung *et al.*, using whole-genome sequencing, identified recurrent HBV integrations at several cancer-related genes (such as *TERT*, *MLL4* and *ROCK1*) in human HCC [9]. However, that study focused on viral genomic integration sites only and the functional consequences of genome disruption by viral insertion were minimally addressed. As the functional impact of HBV integration can also be mediated through transcriptional modification, a comprehensive analysis of the HBV insertion sites focusing on the transcriptome would allow identification of chimeric HBV-human transcripts in HCC and their functional consequences.

In our present study, we aimed to identify novel HBV-human chimeras and delineate the underlying mechanisms in generation of chimeras with an unbiased, transcriptional level and single-base resolution tool in human HCCs. With whole-transcriptome sequencing performed on 16 pairs of HCCs and non-tumorous livers (NT-Ls), we identified a novel mechanism for the formation of HBV-human chimeric transcripts from intronically integrated HBV genome, which was further validated in an expanded cohort of 32 pairs of HCCs and NT-Ls. This mechanism resulted in the generation of a novel and recurrent HBV-cyclin A2 (CCNA2) fusion transcript, A2S. CCNA2 is a cell cycle regulatory protein and acts as a regulatory subunit of cyclin-dependent kinase (CDKs). Here we uniquely demonstrated that A2S was produced by a novel pre-mRNA splicing of the HBV genome, after HBV integration at the intron 2 of *CCNA2* gene, to form a 177 bp in-frame pseudo-exon. Upon the HBV pseudo-exon fusion, the destruction box (D-box) of A2S protein was disrupted, resulting in a gain of non-degradable property. Moreover, A2S retained an intact function in enhancing the cell cycle progression, hence demonstrating its potential role in tumor development. Altogether, we identified a novel pre-mRNA splicing mechanism resulting in a HBV-human chimeric transcript from intronically integrated HBV genome with functional impact.

Materials and methods

Patient samples

Forty-eight pairs of human HCCs and their corresponding NT-L tissues from patients with liver resection for HCC between 1991 and 2001 at Queen Mary Hospital, Hong Kong were selected. All these 48 patients had positive serum hepatitis B surface antigen (HBsAg) status. The age of patients ranged from 24 to 71 years; 36 were male and 12 female. All specimens were snap-frozen in liquid nitrogen and kept at −80 °C. Frozen sections were cut from HCC and NT-L blocks separately and stained for histological examination to ensure a homogenous cell population of tissues. Use of human samples was approved by the institutional review board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW 09-185).

Transcriptome sequencing

Transcriptome sequencing of both HCC and corresponding NT-L tissue samples was performed by Axeq (Seoul, Korea), using 100 bp pair-end sequencing on the HiSeq 2000 platform. On average, 10 million reads per sample were generated. The number of reads aligned to HBV genome (DQ089769.1) is 26935.5 in median. Aligned reads were further processed for HBV-human junction identification. Sequencing data is deposited in Sequence Read Archive (SRA) with ID SRP062885. Refer to the Supplementary materials and methods for details.

Detection and experimental validation of A2S transcript

Multiple sequencing reads were found to align perfectly to a hypothesized HBV-CCNA2 fusion transcript. The hypothesized transcript contains exon 2 of *CCNA2*, a *HBs* gene fragment (282 nt. −458 nt.) and exon 3 of *CCNA2*. Since the partial *HBs* gene insert of the hypothesized fusion transcript was only 177 bp long, library design with 300 bp long insert size was able to span both chimeric junctions to indicate they coexist on a single transcript. To validate the hypothesized A2S transcript by PCR, primers were designed to flank exon 2 and exon 3 of *CCNA2*. Sanger sequencing was performed on the PCR product to confirm its identity.

Protein degradation assay

Determination of protein degradation was assayed as described previously. [32] In brief, cells were transfected with flag- c-Myc-tagged WT CCNA2 and A2S constructs for 48 h and collected after treatment with cycloheximide (50 μg/ml) for the indicated time points. Cells were then lysed for Western blotting with anti-flag or -c-Myc antibody. The band intensity of the Western blotting result was measured by ImageJ. The protein abundance was then plotted against time. The protein level of β-actin was used as an internal control.

Results

Detection of HBV-human junction sites by transcriptome sequencing

Transcriptome sequencing was performed on 16 pairs of HBV-associated HCCs and their corresponding NT-L tissues. HBV sequences were detected in all patients, 13 of whom were infected with HBV genotype C strain, while the others were with genotype B strain (Supplementary Fig. 1). To detect transcripts containing viral-human chimeric junctions, sequence reads were aligned to human and HBV reference sequences and by GSNAP through two alignment stages to maximize alignment sensitivity and specificity. A total of 506 distinct HBV-human junctions were detected; of these, 413 were found in the NT-L tissues, while 94 were found in the HCCs (Supplementary Table 2). However, the average sequencing read depth covering the HBV-human junctions was higher in HCCs (49.6) than in NT-Ls (25.4). This reduced diversity of highly expressed HBV-human junctions in HCCs may reflect the clonal origin of tumor cells, or positive selection of tumor cells with growth-promoting HBV integration events.

To identify hotspots for HBV-human junctions, coordinates of junctions on the viral arm were clustered in intervals of 50 nt. for enrichment analysis. Of all the HBV-human junctions detected, 57% located on the *HBx* gene. Furthermore, the enrichment of chimeric junctions on *HBx* indicates preferential breaks of HBV at the Direct Repeat 1 (DR1) site (1800 nt. −1850 nt.) upon integration (Fig. 1A).

Significant enrichment of HBV-human junctions at the human splice sites

Our transcriptome sequencing results indicate additional peaks (Fig. 1A) over HBV-human junctions, which were not observed in previous whole-genome sequencing studies. To understand the observation, we tested whether there was a significant difference in the proportion of junctions formed with the gene features present on human transcribed region: exonic and splicing (Supplementary Fig. 2). The results showed that HBV-human junctions tended to form on human splice sites than on other exonic regions remarkably. ($p < 2.2 \times 10^{-16}$, Z-test). Surprisingly, we found that these junctions were recurrently formed in multiple genes (*CCNA2*, *TERT*, *NR3C2* and *MLL4*; Fig. 1C) and were most frequently detected at 458 nt., followed by 282 nt. and 2449 nt. on the HBV genome (in *HBs* and *HBc* genes) (Fig. 1B), suggesting that these sites might be responsible for multiple splicing disruption events over different genes.

Identification of a novel HBV-CCNA2 chimeric transcript (A2S) in HCC

Among the junctions found to be enriched on splice sites, two HBV-human junctions (HBV 458 nt. and HBV 282 nt.) were most common. These two junctions resulted in a novel chimeric transcript where human *CCNA2* fused with a portion of the *HBs* gene (282 to 458 nt.), as identified in sample 149T (Fig. 2) (Alignment data shown in Supplementary Fig. 3). To validate this chimeric