Research Article

CrossMark

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.

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

Research Article

A HBV-human junctions (Global)

B HBV-human junctions (Splice site)

C Genes detected with HBV-human junction at human splice sites

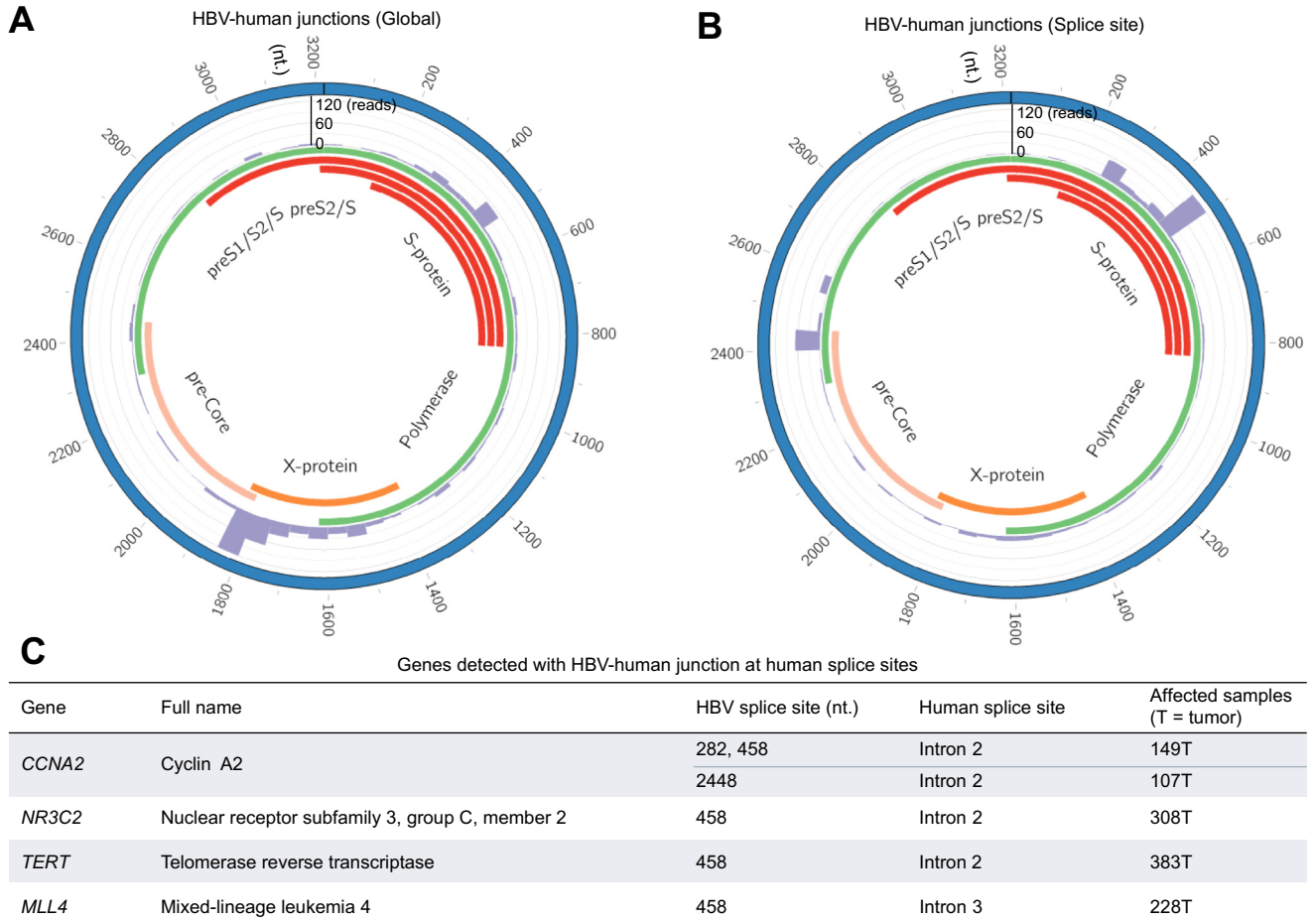| Gene	Full name	HBV splice site (nt.)	Human splice site	Affected samples (T = tumor)
CCNA2	Cyclin A2	282, 458	Intron 2	149T
		2448	Intron 2	107T
NR3C2	Nuclear receptor subfamily 3, group C, member 2	458	Intron 2	308T
TERT	Telomerase reverse transcriptase	458	Intron 2	383T
MLL4	Mixed-lineage leukemia 4	458	Intron 3	228T

Fig. 1. HBV-human chimeric junctions at splice sites. (A) Abundance of chimeric junctions on HBV genome. Note that the chimeric junctions were found more frequently at site of ~1800 nt. and 460 nt. of the HBV genome. (B) Chimeric junctions limited to human splice sites. Note that multiple peaks were found at 282 nt., 458 nt. and 2448 nt. The locations of HBV genes were annotated as S gene, polymerase gene, X gene, and pre-core gene, respectively. The blue bars indicate the number of reads detected at the position. (C) List of genes detected with HBV-human junction at human splice sites. Only tumor-specific events are listed. (This figure appears in colour on the web.)
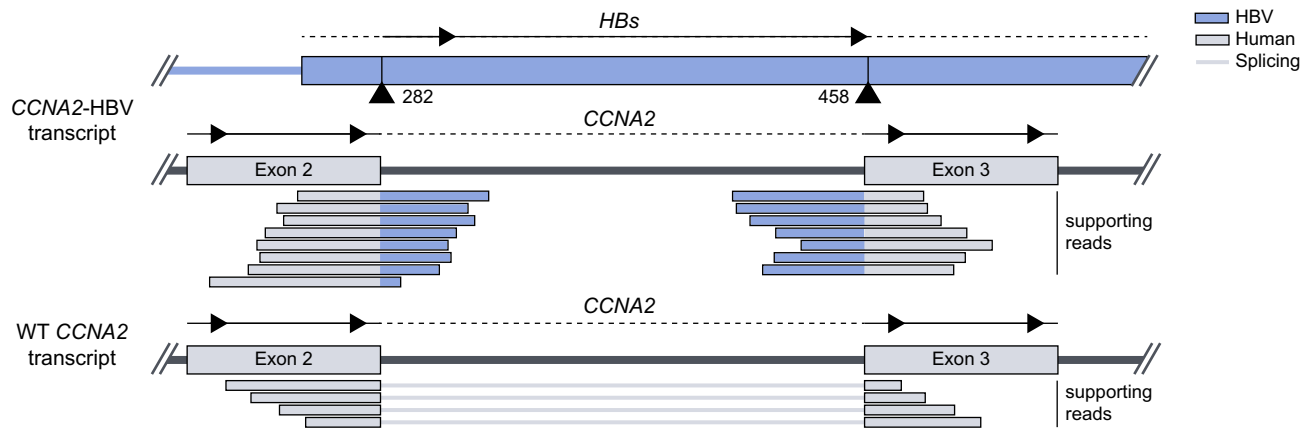

Fig. 2. Graphical illustration of CCNA2-HBs chimeric fusion transcript (A2S) sequence alignment. Reads alignment to the human *CCNA2* and HBV are shown. Note that *HBs* gene (282 nt. to 458 nt.) was integrated into intron 2 of *CCNA2*. Original data are shown in Supplementary Fig. 2.

transcript (RNA), PCR amplification was performed on cDNA using primers flanking *HBs* and *CCNA2* (Supplementary Fig. 4A and B), and two PCR products were detected. Sanger sequencing performed on both PCR products confirmed that the *HBs* gene

from 282 nt. to 458 nt. was fused between the exon 2 and exon 3 of *CCNA2* in the larger-sized product (Supplementary Fig. 4C). This fusion of the *HBs* gene resulted in a novel HBV-CCNA2 chimeric transcript, which we named A2S.

Pre-mRNA splicing mechanism as a common mechanism in generation of HBV-human chimera

To further examine whether A2S was generated at the genomic or transcriptional level, genomic DNA of 149T was PCR-amplified and Sanger-sequenced on *CCNA2*, with primers spanning exons 2 and 3. Two PCR products, approximately of 1500 bp (CCNA2 E2/E3 S) and 3000 bp (CCNA2 E2/E3 L), were found (Supplementary Fig. 5A). The sequence of CCNA2E2/E3 S was identical to the wild-type *CCNA2*, while that of CCNA2E2/E3 L showed integration of a 1450 bp contiguous segment of the HBV genome (3179–3215 nt. and 1–1412 nt.) into its intron 2 (Supplementary Fig. 5B). This confirmed the presence of HBV integration into the *CCNA2* gene.

Intriguingly, there was a large discrepancy between the integrated HBV sequence (3179–3215 nt. and 1–1412 nt.) in the genomic *CCNA2* gene and that (282–458 nt.) in the *CCNA2* transcript, with the insert in the transcript being only a small portion of the insert in the gene (Fig. 3A and Supplementary Fig. 5B). This difference between the genomic and transcriptional levels strongly suggested that the A2S chimeric transcript was generated via a post-transcriptional pre-mRNA splicing mechanism upon genomic integration.

Moreover, we observed that all splice donor and acceptor sites of *CCNA2* remained intact after the HBV integration. Under normal transcriptional splicing, the integrated HBV segment within *CCNA2* intron should have been removed. Therefore, the retention of HBV 282–458 nt. in the A2S transcript required further explanation. Inspection of the HBV sequence at 282–458 nt. revealed the presence of a possible splice acceptor (AG) at 282 nt. and a possible splice donor site (GT) at 458 nt. (Fig. 3B). These splice acceptor and donor sites likely might have participated in the alternative splicing event of *CCNA2*, leading to the retention of the integrated HBV sequence from 282 nt. −458 nt. upon post-transcriptional excision.

To further test the hypothesis on the formation of A2S through pre-mRNA splicing of HBV and human genome, we ectopically expressed a c-Myc-tagged construct of *CCNA2* containing the 1450 bp contiguous segment of the HBV genome (3178–3215 nt. and 1–1412 nt.) in intron 2 (HBVIn) in BEL-7402 HCC cells (Supplementary Fig. 5C and D). Upon HBVIn ectopic expression, the protein and transcript produced had the same size with those of A2S, indicating an occurrence of splicing event on the pre-mRNA of HBVIn (Fig. 3C and D). This indicates that the A2S chimeric protein was likely produced through post-transcriptional pre-mRNA splicing between HBV and human genome.

Prevalence of A2S chimeric transcript in human HCCs, the corresponding non-tumorous livers, and non-HCC HBV-associated cirrhotic livers

In the transcriptome sequencing, the A2S chimeric transcript was found in the tumors in one of the 16 pairs of human HCCs. To further study the prevalence of A2S chimeric transcript in HBV-associated HCCs, the presence of A2S was examined in the cDNA of a separate cohort of 32 pairs of HCCs and their corresponding NT-Ls by PCR and Sanger sequencing. Two pairs of primers, 1) HBVsF/R8 and 2) F2/HBVsR flanking different regions between the *HBs* gene and *CCNA2*, were used for the PCR. As shown in Fig. 4A, the A2S fusion transcript was found mainly in the HCC

tumors. Sanger sequencing result of the 32 cases from the expanded cohort further confirmed the presence of the A2S chimeric transcript in 4 more tumors (247T, 346T, 284T and 340T). In another case (343NT), A2S was found in the NT-L tissue only and not in the HCC tumor (Fig. 4B). Altogether, A2S was found in 5 (10.4%) of the 48 HCC tumors (16 from transcriptome sequencing and 32 from the expanded cohort) (Fig. 4C). In addition, the presence of A2S chimeric transcripts was also examined in 22 cases of non-HCC HBV-associated cirrhotic livers (Supplementary Fig. 6) and the results showed no A2S transcript was found in all of these cases. These results indicate that A2S was preferentially found in the HCC tumors and raised the question of its potential tumorigenicity. Interestingly, all 6 *HBV-CCNA2* fusion transcripts have identical *HBs* sequence fragment inserted between exon 2 and exon 3 of *CCNA2* transcript, supporting our hypothesis of pre-mRNA splicing mechanism in the generation of HBV-human chimeric transcript. Other than the transcript, the encoded A2S protein was also found to be expressed in 149T tumor, 247T tumor and 346T tumor liver (Supplementary Fig. 7). This agreed with the occurrence of A2S at the transcript level. This confirmed the plausibility in the generation of the A2S chimeric protein.

Induction of cell cycle progression by A2S chimeric protein, with same effect as wild-type CCNA2 overexpression

CCNA2 is a member of the cyclin family, which regulates cell cycle progression in somatic cells. It is important in the initiation and progression of S phase and necessary for G2-M exit and transition [10,11]. Taken into account the function of *CCNA2* acting as a cell cycle regulator, we postulated that alteration of the *CCNA2* activity in proper cell cycle upon HBV integration might result in neoplastic transformation.

To assess this hypothesis, the cell cycle patterns were analyzed by flow cytometry in BEL-7402 and SMMC-7721 HCC cells. We ectopically expressed IRES-EGFP-tagged A2S and wild-type *CCNA2* in these cells and only the EGFP-positive cells, i.e. CCNA2- or A2S-overexpressing cells, were analyzed. In Fig. 5A and B, both A2S and wild-type CCNA2 transfectants demonstrated enhanced S phase progression with increased percentages of cells in S and G2-M phases, as compared with the empty vector control. Both CCNA2 and A2S BEL7402 stable transfectants (Supplementary Fig. 8A) also showed a significantly stronger Bromodeoxyuridine (BrdU) staining indicating higher DNA synthesis rate in BrdU cell proliferation assay, increased numbers of colonies formed in colony formation and soft agar colony formation assays, when compared to the empty vector control (Supplementary Fig. 8B, C and D). In addition, the enhanced cell growth ability by CCNA2 and A2S was further evidenced in an immortalized normal liver cell line LO2. CCNA2 and A2S LO2 transfectants showed a higher proliferation rate in cell proliferation assay (Supplementary Fig. 9A), a stronger BrdU staining in BrdU proliferation assay (Supplementary Fig. 9B), an induction of S phase progression in flow cytometry (Supplementary Fig. 9C), and an increased number of colonies formed in colony formation assay (Supplementary Fig. 9D). These results showed that the A2S fusion protein retained the ability of wild-type CCNA2 in enhancing cell proliferation and, notably, this ability was comparable to that of the overexpression of wild-type CCNA2.
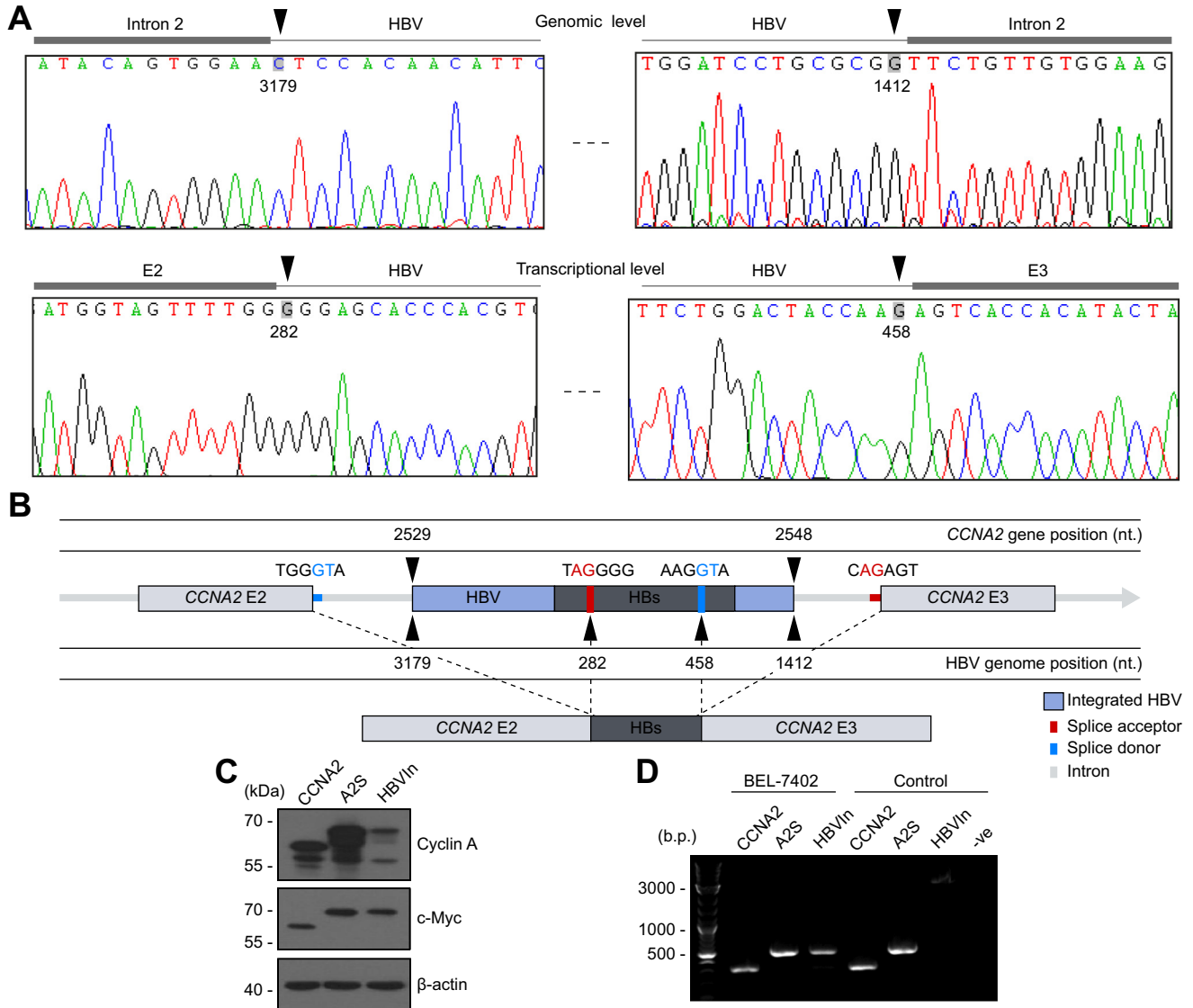
Research Article

Fig. 3. Generation of novel CCNA2-HBs chimera (A2S) by alternative splicing. (A) Discrepancy of *CCNA2*-HBV fusion sites between genomic and transcriptional level in 149T. PCRs were performed on DNA (genomic level) and cDNA (transcriptional level) of 149T as described in Supplementary Figs. 2 and 3. At the genomic level, HBV gene sequence from 3178 nt. to 1412 nt. was integrated into intron 2 of *CCNA2* gene, whereas at the transcriptional level, HBV gene sequence from 282 nt. to 458 nt. was located between exon 2 and exon 3 of *CCNA2*, suggesting a splicing event accounting for this dissimilarity. (B) The HBV gene was integrated between 2529 nt. and 2548 nt. of intron 2 of *CCNA2* gene at the genomic level. New potential splice acceptor (AG) at 282 nt. and acceptor (GT) at 458 nt. were observed upon HBV integration at the genomic level. (C) Generation of protein with the same size as A2S upon expression of HBVIn construct. BEL-7402 cells expressed with c-Myc tagged CCNA2, A2S and HBVIn constructs were subjected to Western blotting. β-actin was used as a loading control. (D) Expression of A2S chimeric transcript upon ectopic expression of HBVIn in BEL-7402. The region of exon 2 and exon 3 was PCR-amplified. PCR amplifications on CCNA2, A2S and HBVIn constructs were used as a control. H$_2$O was used as negative control which is denoted as −ve. (This figure appears in colour on the web.)

Disruption of proper degradation upon partial HBs fusion in A2S chimeric protein

A proper CCNA2 degradation is required for a rigid cell cycle progression. CCNA2 starts to accumulate during the S phase and is abruptly destroyed before the metaphase by ubiquitin-mediated proteolysis [12]. Therefore, we first investigated the influence of *HBs* gene fusion on the change of structural domains by protein structure prediction. Although the insertion of partial HBs protein did not directly disrupt the functional motifs e.g. cyclin box or D-box motif in CCNA2 (Supplementary Fig. 10), the analysis illustrated more frequent formation of α-helix

subunits upon partial S protein fusion in the A2S chimeric protein (Fig. 6A). Moreover, the new α-helix subunits mainly located near the N-terminus (1 aa to 235 aa) of A2S and disrupted the original conformation of destruction box (D-Box) motif (47 aa to 57 aa) of CCNA2. An intact D-box motif is responsible for the destruction of cyclins by programmed proteolysis at the end of mitosis to exert cell cycle control [13,14]. Therefore, we hypothesized that the A2S protein was rendered a non-degradable property upon the HBs fusion event.

Consistent with this hypothesis, the degradation rate of A2S protein was slower than that of the wild-type CCNA2 protein, as shown in BEL7402 expressing flag-tagged A2S and wild-type
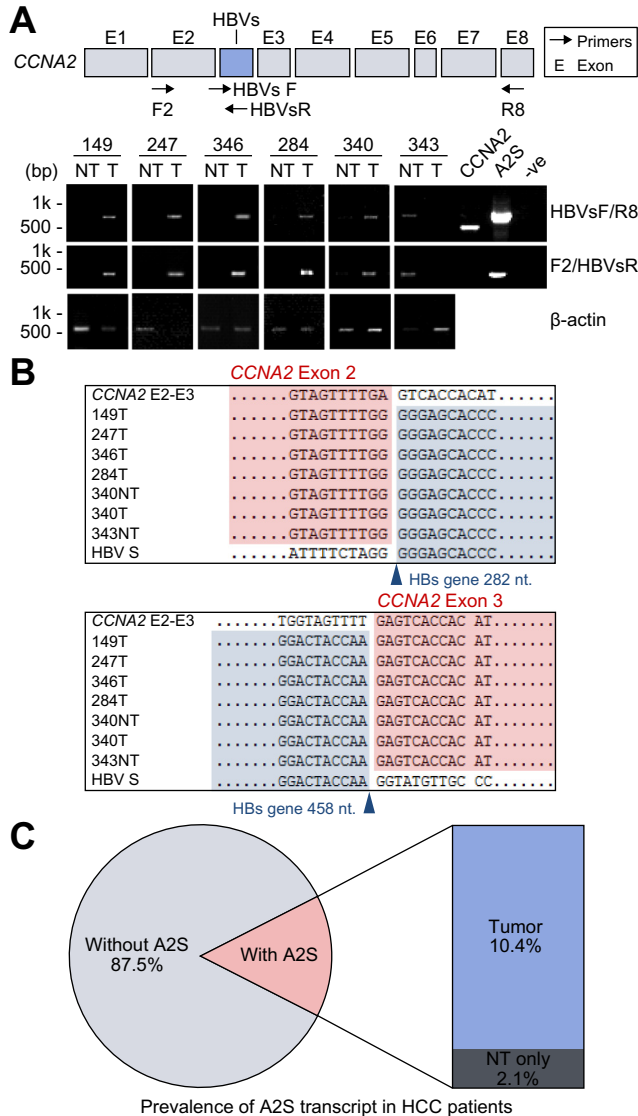
Fig. 4. The A2S chimeric transcript was more frequently detectable in tumors than NT-Ls. (A) Primers flanking different regions of A2S transcript were annotated. Presence of the A2S fusion transcript was examined in cDNA extracted from 48 HCCs and their corresponding NT-Ls by PCR. A2S plasmid were used as positive controls and CCNA2 plasmid and H$_2$O as negative control which is denoted at −ve. β-actin acted as a loading control. (B) Nucleotide sequences of the PCR products of the samples were confirmed by Sanger sequencing. Noted that all 7 samples showed the same *HBV-CCNA2* fusion sites (282 nt. and 458 nt.) (C) Chart showing the prevalence of A2S transcript in HCC patients. Six (12.5%) of the 48 patients were found to carry the A2S transcript. A2S transcript was detectable in the tumors in 5 (10.6%) of the six A2S-positive patients. (This figure appears in colour on the web.)

CCNA2 (Fig. 6B). In BEL7402 cells, the levels of A2S protein at 1, 2 and 3 h were similar to that at 0 h after supplementation of cycloheximide (CHX), while more than 50% of wild-type CCNA2 protein was degraded at the first hour. These results were further confirmed in the immortalized liver cell line LO2 (Fig. 6C). The result provided us with the first hint on how partial *HBs* gene (282 nt. to 458 nt.) fusion affected the degradation profile of A2S chimeric protein.
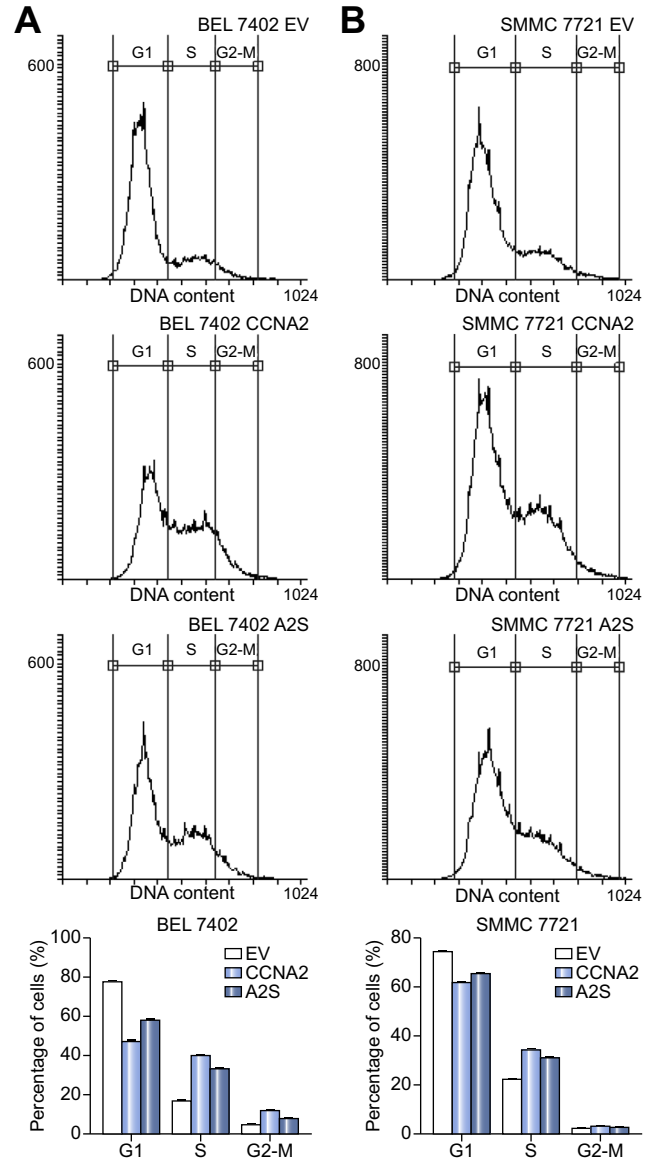

Fig. 5. A2S exerted a similar effect on cell cycle progression as wild-type CCNA2 overexpression and induced S phase progression in HCC cells. (A) BEL-7402 and (B) SMMC-7721 cells ectopically expressed with IRES-EGFP-tagged CCNA2 and A2S constructs were subjected to cell cycle analysis by flow cytometry. Both CCNA2 and A2S transfectants showed increased percentages of cells in the S and G2-M phases.

Discussion

In this study, we identified a novel mechanism underlying the formation of HBV-human chimeric transcripts from intronic-integrated HBV genome, exemplified by the recurrent HBV-CCNA2 fusion transcript A2S. We uniquely demonstrated that the A2S was produced by a novel pre-mRNA splicing of the host genome after HBV integration at the intron 2 of *CCNA2* gene, with subsequent formation of a 177 bp in-frame pseudo-exon and the fusion transcript A2S acquired potential oncogenic properties.

HBV integration into human genome is a major causative factor for hepatocarcinogenesis. A number of recent next generation

Research Article

A   Comparison of predicted structure of *CCNA2* and A2S

Destruction box (47aa to 57 aa)
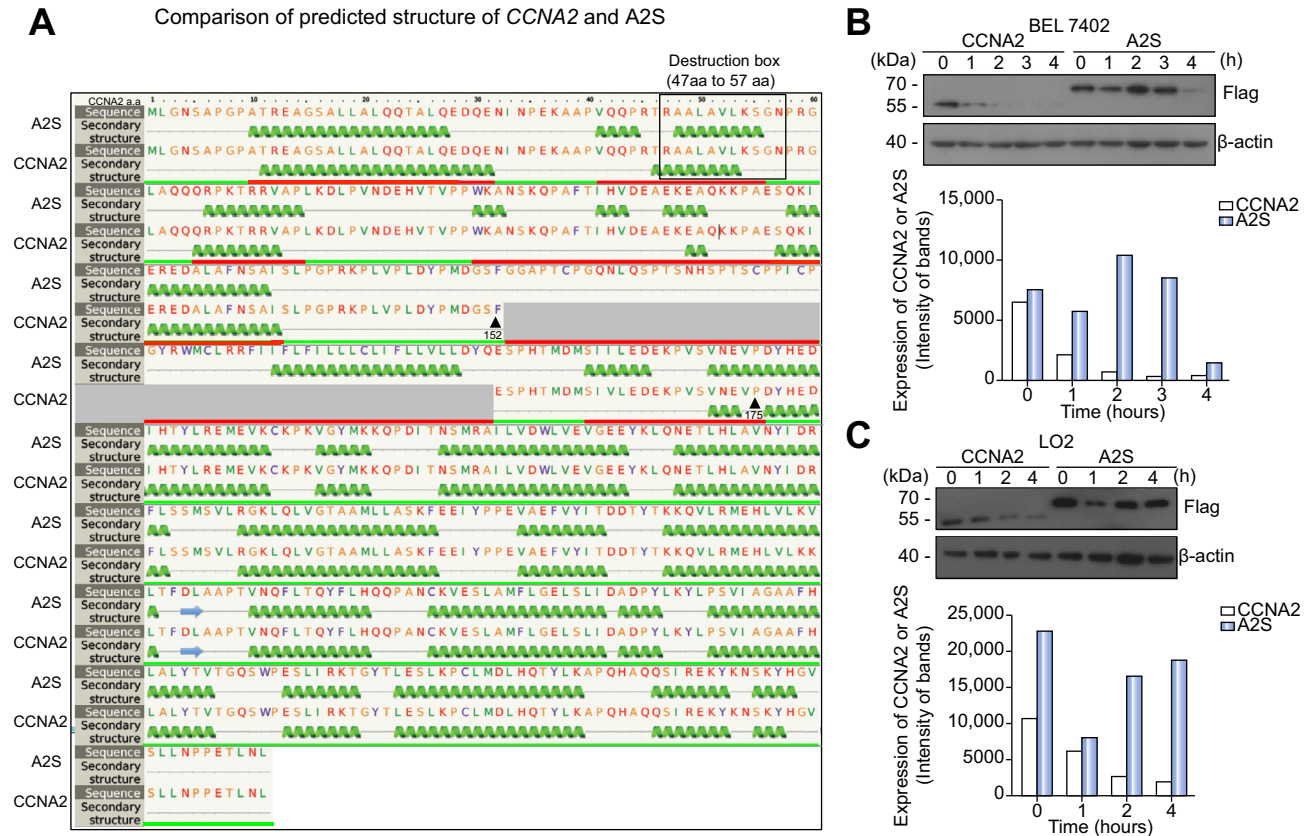
B   BEL 7402

C   LO2

Fig. 6. Structure prediction and degradation profile of CCNA2 and A2S. (A) Protein secondary structure prediction of CCNA2 and A2S. The structural change was only from 1 aa to 235 aa of A2S and included the CCNA2 degradation box (47 aa to 57 aa), while the downstream structures were predicted largely to be undisrupted. The red bar indicates difference on α-helix structure between the CCNA2 and A2S. (B & C) The BEL7402 and LO2 cells were transiently transfected with flag-tagged CCNA2 or A2S constructs. 48 h after transfection, 50 μg/ml cyclohexamide was added and the cells were harvested at indicated time points. Western blot analysis was performed using flag antibody and β-actin was used as internal control. The protein level of CCNA2 was decreased while that of A2S was unchanged in BEL7402 and LO2 cells. The degradation chart of CCNA2 and A2S. The band intensity of the Western blotting was measured by ImageJ and the optical intensity plotted against time. Note that the protein level of AS2 in the transfected cells was unchanged in the degradation chart. (This figure appears in colour on the web.)

sequencing studies have highlighted recurrent regions of HBV integration [9,15]. In a very recent whole-transcriptome sequencing study characterizing transcribed HBV-human fusions, although HBV-human chimeric transcripts were widely studied, only HBV-positive HCC cell lines were investigated for viral integrants, thus with restriction at *in vitro* level [16]. In our present study, we employed transcriptome sequencing and functional assays in an attempt to examine the impact of post-transcriptional modification on HBV integration *in vivo* and its chimera derivatives in paired human HCCs and their NT-Ls. In accord with the previous studies [17–19], recurrent HBV integration was also detected in *CCNA2*, *TERT* and *MLL4* genes in our findings (Fig. 1C). Beyond these findings, this present study has provided the first evidence detailing the significant enrichment of HBV-human junction hotspots confined to splice sites (exon boundaries) of the host arm. For the same set of junctions over the viral arm, enrichments were observed on three junctions located at 282 nt., 458 nt. and 2449 nt., instead of the major DR1 breakpoint at ~1824 nt., and observed in a global manner. Moreover, we have revealed a common but novel pre-mRNA splicing mechanism of HBV intronic integration, generating potential oncogenic HBV-human chimeras in HCC.

In the present study, more HBV-human junctions were found in NT-Ls (413 of 507, about 81%) than in T-L tissues (94 of 507,

about 19%), which is in agreement with previous reports [20,15]. However, our observation was different from a recent HBV integration study using whole-genome deep sequencing [21] and study of Sung *et al.* [9], which showed that the number of HBV integration sites was higher in tumor derived tissues than that in NT livers. This may be due to the technical differences between transcriptome sequencing and that of whole-genome sequencing employed by other HBV integration studies. The higher expression of cis/trans-activated chimeric transcripts present in the HBV integrated hepatocytes has enabled detection by transcriptome sequencing. Furthermore, a previous study also provided an explanation for similar observations [4] where viral integration is shown to occur during HBV chronic infection before HCC initiation. Hepatocytes with viral integration expanded during chronic hepatitis and some cells took part in the formation of focal proliferative lesions and gained growth advantages, therefore, leading to clonal expansions during tumorigenesis. This echoed with our observation that the average sequencing read depth covering the HBV-human junctions was higher in HCCs (49.6) than in NT-Ls (25.4) as a result of clonal expansion.

By confining our search of HBV-human chimeric junctions to human splice sites, we showed that in *CCNA2*, the foreign splicing events at two HBV sites (282 nt. and 458 nt.) were capable of introducing an insert without frame-shift between the exon 2

and exon 3. Of interest, our study showed insertion of new splice acceptor (AG) from HBV 282 nt. and splice donor (GT) from HBV 458 nt. in the integrated fragment of *HBs* into the *CCNA2* gene, and GT-AG splice pair is a canonical splice site previously reported in mammalian genomes [22]. In spite of the randomness of HBV integration into human genomic DNA, it is worth noting that as long as the inserted viral sequence contains the two HBV foreign splice sites, this post-transcriptional pre-mRNA splicing mechanism allows the integrated HBV sequence to generate an identical fusion transcript regardless of the precise breakpoints of both the human and HBV genomes. In fact, splicing events have been widely reported in HBV genome [23,24], and are involved in the post-transcriptonal control of HBV gene expression [25] and generation of truncated HBV protein [24]. However, the role of pre-mRNA splicing in generating HBV-human chimeric protein was not previously appreciated. Also, there have been limited reports on the link between intronic HBV integration and HCC development. Previous findings have mostly focused on integrations in host exonic and promoter regions, which are believed to exert a greater influence on protein function and transcriptional regulation [9,15]. As a result, the impact of intronic integration on coding transcripts has been underestimated. Given the total intronic region is more than ten times longer than the exonic regions, our study has greatly highlighted the significance of HBV intronic integration by showing intronic integrations could alter the sequence of protein coding genes and probably the protein conformation, which greatly enhances the capability of HBV in affecting human protein functions.

As far as we understand, this is the first report of the identification of this novel chimera A2S. CCNA2 is a cell cycle regulatory protein which acts as a regulatory subunit of CDKs. Genomic integration events were recurrently found within intron 2 of *CCNA2* in previous studies [15]. Early in 1990, Wang *et al.* first reported the existence of an HBV-fused *CCNA2* in HBV-associated HCC [17,26]. Later, the oncogenic property of the HBV-altered *CCNA2*, in which N-terminal part was replaced by the HBV envelope protein, was further demonstrated by Berasain *et al.* They showed that rat kidney cell was transformed by the disrupted *CCNA2* due to abnormal cellular localization of the protein [27]. These studies have provided evidence that HBV promotes cellular transformation through integration into *CCNA2* gene. However, the underlying mechanism of the generation of chimeric protein and the viral-induced carcinogenetic changes were not fully explained in these studies.

In our study, the splicing events introduced by the two splice sites, 282 nt. and 458 nt., were more commonly found in HCC tumors, indicating the splicing events induced by the two splice site hotspots may be implicated in hepatocarcinogenesis by altering the functions of different genes upon HBV fusion. Other than *CCNA2*, the chimeric transcripts listed in Fig. 1C may also lead to the production of HBV-human chimeric proteins. In *NR3C2* and *TERT*, HBV was found to form chimera from intron 2 and spliced with exon 3 at HBV 458 nt. However, for *NR3C2*, the transcription was not in-frame as indicated in the Supplementary Fig. 11A, and acquired a stop codon in exon 3. Therefore, in this splicing event, we may speculate that a truncated HBV-NR3C2 chimeric protein may be produce. For *TERT*, the transcription was in-frame (Supplementary Fig. 11B), and the chimeric transcript could encode a functional chimeric protein. It is also speculated that the active transcription by the fused HBV-S promoter together with the

cryptic splicing could cis-activate *TERT*. Possible functional impact of the HBs-TERT and HBV-NR3C2 chimeras requires further investigation.

HBV integration into the intron 2 of *CCNA2* followed by pre-mRNA splicing was capable of producing a HBV-human fusion product A2S, which may provide a selective advantage in tumorigenesis. In our cohort, we detected recurrent A2S expression in 6 of 48 (12.5%) of the HBV-associated HCCs, of which more than 80% (5 of 6) was found in the tumors only, implicating its potential tumorigenicity. Moreover, it was interesting to find no A2S transcript in the 22 cases of HBV-associated non-HCC cirrhotic livers in this study (Supplementary Fig. 6). This suggests that this event is possibly in close relation with the onset of HCC. In addition, it is noteworthy that the ability of induction of S phase progression of A2S was comparable to overexpression of the wild-type CCNA2. Since upregulation of CCNA2 is highly associated with tumorigenesis in a number of cancers including HCC [12,28,29], dysregulation of CCNA2 may play a direct role in cellular transformation and oncogenesis through the loss of cell cycle control [30]. This result suggests a potential oncogenic role of A2S in hepatocarcinogenesis and more functional analysis should be performed to confirm the tumorigenicity in future studies. Taken together, our findings have emphasized the importance of the A2S chimeric protein in conferring oncogenic advantages in the progression of HCC.

In cyclin degradation, the destruction box in the N-terminus binds with a series of proteins and produces a hybrid that is susceptible to proteolysis through ubiquitin pathway [14]. In line with the findings on potential tumorigenicity of the A2S chimeric protein, we also demonstrated that the A2S protein acquired an altered regulatory cascade when compared to the wild-type CCNA2. In the A2S chimeric protein, the C-terminus functional motifs including the cyclin box motifs, which play a key role in cell cycle regulation [12,31], were undisrupted upon HBV fusion. However, the proper degradation of the protein was hindered by an indirect disruption on the conformation of destruction box located on the N-terminus of CCNA2. As demonstrated, A2S fusion protein was still capable of progressing cell cycle in the S phase. This functional characteristic was distinct from the other previously reported HBV-CCNA2 chimeric protein, in which the N-terminal part was replaced by the HBV envelope protein [17], the destruction box of the protein was directly replaced by HBs protein and its oncogenicity was associated with an abnormal cellular localization in the endoplasmic reticulum membrane instead of cell cycle deregulation. In the present study, together with the finding that A2S chimeric protein had a positive influence in S phase progression, the inability of degradation of A2S protein was likely to ultimately increase the overall expression of A2S protein along the cell cycle progression. We speculated that the accumulation of non-degradable A2S not only favored the progression of S phase in enhancing cell proliferation, but might also disturb the proper mitotic exit in the progression of G2-M phase. A delay of mitotic exit might result in abnormal chromosome segregation and hence chromosomal instability promoting cancer susceptibility.

In summary, this study has highlighted the importance of HBV involvement in the mammalian pre-mRNA splicing machinery, a previously omitted but important mechanism, in the generation of potential tumorigenic HBV-human chimeras. Our study has also identified a novel potential tumorigenic HBV-human chimera A2S, and provided insight on the impact of intronic HBV

Research Article

integration with the gain of splice sites in the development of HBV-associated HCC.

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Authors' contributions

I.O. Ng, C.M. Wong and P.C. Sham designed the study. The manuscript was prepared by Y.T. Chiu, K.L. Wong, C.M. Wong, and I.O. Ng. P.C. Sham. K.L. Wong, P.C. Sham, S.W Choi, W.L Yang, S. Cherny, and D.W. Ho performed the bioinformatics analysis. Y. T. Chiu, K.M. Sze, L.K Chan and J.M. Lee performed the experiments and provided experimental design. K. Man provided human cirrhotic liver tissue samples. Y.T. Chiu, K.L. Wong, S.W Choi, P.C. Sham, C.M. Wong and I.O. Ng analyzed the data.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhep.2016.02.005.

References

[1] Ishikawa T. Clinical features of hepatitis B virus-related hepatocellular carcinoma. World J Gastroenterol 2010;16:2463–2467.

[2] Yuen MF, Hou JL, Chutaputti A. Asia Pacific Working Party on Prevention of Hepatocellular C. Hepatocellular carcinoma in the Asia pacific region. J Gastroenterol Hepatol 2009;24:346–353.

[3] Nagaya T, Nakamura T, Tokino T, Tsurimoto T, Imai M, Mayumi T, et al. The mode of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. Genes Dev 1987;1:773–782.

[4] Murakami Y, Saigo K, Takashima H, Minami M, Okanoue T, Brechot C, et al. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. Gut 2005;54:1162–1168.

[5] Robinson WS, Klote L, Aoki N. Hepadnaviruses in cirrhotic liver and hepatocellular carcinoma. J Med Virol 1990;31:18–32.

[6] Hino O, Kitagawa T, Koike K, Kobayashi M, Hara M, Mori W, et al. Detection of hepatitis B virus DNA in hepatocellular carcinomas in Japan. Hepatology 1984;4:90–95.

[7] Tokino T, Matsubara K. Chromosomal sites for hepatitis B virus integration in human hepatocellular carcinoma. J Virol 1991;65:6761–6764.

[8] Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 2012;44:685–689.

[9] Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 2012;44:765–769.

[10] Zindy F, Lamas E, Chenivesse X, Sobczak J, Wang J, Fesquet D, et al. Cyclin A is required in S phase in normal epithelial cells. Biochem Biophys Res Commun 1992;182:1144–1154.

[11] Pagano M, Pepperkok R, Verde F, Ansorge W, Draetta G. Cyclin A is required at two points in the human cell cycle. EMBO J 1992;11:961–971.

[12] Yam CH, Fung TK, Poon RY. Cyclin A in cell cycle control and cancer. Cell Mol Life Sci 2002;59:1317–1326.

[13] Klotzbucher A, Stewart E, Harrison D, Hunt T. The 'destruction box' of cyclin A allows B-type cyclins to be ubiquitinated, but not efficiently destroyed. EMBO J 1996;15:3053–3064.

[14] Glotzer M, Murray AW, Kirschner MW. Cyclin is degraded by the ubiquitin pathway. Nature 1991;349:132–138.

[15] Ding D, Lou X, Hua D, Yu W, Li L, Wang J, et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. PLoS Genet 2012;8 e1003065.

[16] Lau CC, Sun T, Ching AK, He M, Li JW, Wong AM, et al. Viral-human chimeric transcript predisposes risk to liver cancer development and progression. Cancer Cell 2014;25:335–349.

[17] Wang J, Chenivesse X, Henglein B, Brechot C. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. Nature 1990;343:555–557.

[18] Paterlini-Brechot P, Saigo K, Murakami Y, Chami M, Gozuacik D, Mugnier C, et al. Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. Oncogene 2003;22:3911–3916.

[19] Saigo K, Yoshida K, Ikeda R, Sakamoto Y, Murakami Y, Urashima T, et al. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. Hum Mutat 2008;29:703–708.

[20] Jiang S, Yang Z, Li W, Li X, Wang Y, Zhang J, et al. Re-evaluation of the carcinogenic significance of hepatitis B virus integration in hepatocarcinogenesis. PLoS One 2012;7 e40363.

[21] Jiang Z, Jhunjhunwala S, Liu J, Haverty PM, Kennemer MI, Guan Y, et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. Genome Res 2012;22:593–601.

[22] Burset M, Seledtsov IA, Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res 2000;28:4364–4375.

[23] Huang CC, Kuo TM, Yeh CT, Hu CP, Chen YL, Tsai YL, et al. One single nucleotide difference alters the differential expression of spliced RNAs between HBV genotypes A and D. Virus Res 2013;174:18–26.

[24] Soussan P, Garreau F, Zylberberg H, Ferray C, Brechot C, Kremsdorf D. In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. J Clin Invest 2000;105:55–60.

[25] Sommer G, Heise T. Posttranscriptional control of HBV gene expression. Front Biosci 2008;13:5533–5547.

[26] Wang J, Zindy F, Chenivesse X, Lamas E, Henglein B, Brechot C. Modification of cyclin A expression by hepatitis B virus DNA integration in a hepatocellular carcinoma. Oncogene 1992;7:1653–1656.

[27] Berasain C, Patil D, Perara E, Huang SM, Mouly H, Brechot C. Oncogenic activation of a human cyclin A2 targeted to the endoplasmic reticulum upon hepatitis B virus genome insertion. Oncogene 1998;16:1277–1288.

[28] Chao Y, Shih YL, Chiu JH, Chau GY, Lui WY, Yang WK, et al. Overexpression of cyclin A but not Skp 2 correlates with the tumor relapse of human hepatocellular carcinoma. Cancer Res 1998;58:985–990.

[29] Ohashi R, Gao C, Miyazaki M, Hamazaki K, Tsuji T, Inoue Y, et al. Enhanced expression of cyclin E and cyclin A in human hepatocellular carcinomas. Anticancer Res 2001;21:657–662.

[30] Hunter T, Pines J. Cyclins and cancer. II: Cyclin D and CDK inhibitors come of age. Cell 1994;79:573–582.

[31] Lees EM, Harlow E. Sequences within the conserved cyclin box of human cyclin A are sufficient for binding to and activation of cdc2 kinase. Mol Cell Biol 1993;13:1194–1201.

[32] Chiu YT, Liu J, Tang K, Wong YC, Khanna KK, Ling MT. Inactivation of ATM/ATR DNA damage checkpoint promotes androgen induced chromosomal instability in prostate epithelial cells. PLoS One 2012;7 e51108.