HEPATOLOGY

AASLD AMERICAN ASSOCIATION FOR THE STUDY OF LIVER DISEASES

HEPATOLOGY, VOL. 70, NO. 3, 2019

LIVER BIOLOGY/PATHOBIOLOGY

Long-Read RNA Sequencing Identifies Alternative Splice Variants in Hepatocellular Carcinoma and Tumor-Specific Isoforms

Hui Chen,[1*] Feng Gao,[4*] Mian He,[1*] Xiao Fan Ding,[1] Aikha M. Wong,[1] Siu Ching Sze,[1] Allen C. Yu,[5] Tingting Sun,[1] Anthony W-H. Chan,[1] Xin Wang,[4] and Nathalie Wong[1-3]

Alternative splicing (AS) allows generation of cell type–specific mRNA transcripts and contributes to hallmarks of cancer. Genome-wide analysis for AS in human hepatocellular carcinoma (HCC), however, is limited. We sought to obtain a comprehensive AS landscape in HCC and define tumor-associated variants. Single-molecule real-time long-read RNA sequencing was performed on patient-derived HCC cells, and presence of splice junctions was defined by SpliceMap-LSC-IDP algorithm. We obtained an all-inclusive map of annotated AS variants and further discovered 362 alternative spliced variants that are not previously reported in any database (neither RefSeq nor GENCODE). They were mostly derived from intron retention and early termination codon with an in-frame open reading frame in 81.5%. We corroborated many of these predicted unannotated and annotated variants to be tumor specific in an independent cohort of primary HCC tumors and matching nontumoral liver. Using the combined Sanger sequencing and TaqMan junction assays, unique and common expressions of spliced variants including enzyme regulators (ARHGEF2, SERPINH1), chromatin modifiers (DEK, CDK9, RBBP7), RNA-binding proteins (SRSF3, RBM27, MATR3, YBX1), and receptors (ADRM1, CD44v8-10, vitamin D receptor, ROR1) were determined in HCC tumors. We further focused functional investigations on ARHGEF2 variants (v1 and v3) that arise from the common amplified site chr.1q22 of HCC. Their biological significance underscores two major cancer hallmarks, namely cancer stemness and epithelial-to-mesenchymal transition–mediated cell invasion and migration, although v3 is consistently more potent than v1. *Conclusion:* Alternative isoforms and tumor-specific isoforms that arise from aberrant splicing are common during the liver tumorigenesis. Our results highlight insights gained from the analysis of AS in HCC. (Hepatology 2019;70:1011-1025).

Alternative splicing (AS) is a posttranscriptional process that allows generation of alternative mRNA transcripts crucial to normal development and contributes to proteome complexity of mammalian genomes. AS can also promote cancer growth and survival. Through expression switch from canonical isoform to aberrant isoform,[(1)] the production of noncanonical and cancer-specific mRNA transcripts can lead to inactivation of tumor suppressors or activation of oncogenes and thus trigger cancer signaling

Abbreviations: A3, alternative 3' splice site; A5, alternative 5' splice site; aa, amino acid; AS, alternative splicing; CCS, circular consensus sequencing; cDNA, complementary DNA; EMT, epithelial-to-mesenchymal transition; FPKM, fragments per kilobase of transcript per million mapped reads; GO, gene ontology; HCC, hepatocellular carcinoma; IDP, isoform detection and prediction; IR, intron retention; MIHA, immortalized human hepatocyte cell line; mTORC1, mammalian target of rapamycin complex 1; NMD, nonsense-mediated decay; ORF, open reading frame; RhoGEF, Rho guanine nucleotide exchange factors; Rho-GTP, Rho-guanosine-5'-triphosphate; RI, retained intron; SE, skipping exon; SF, splicing factor; SGS, second-generation sequencing; SMRT, single-molecule real-time; VDR, vitamin D receptor.

Received June 25, 2018; accepted December 28, 2018.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.30500/suppinfo.

**These authors contributed equally to this work.*

Supported by a Theme-based Research Scheme from the Hong Kong Research Grants Council (Ref. No.: T12-403/11), in part by Collaborative Research Funds from the Hong Kong Research Grants Council (Ref. No.: C4041-17 and C7019-16E) and the VC's discretionary fund from the Chinese University of Hong Kong.

The data that support the findings of this study are available from the authors upon reasonable request. Complete Dataset of SMRT Sequenced Long Reads and Illumina Short Reads Data of MIHA, HKCI-2, HKCI-C1, HKCI-C2, and HKCI-C3 can be accessed at European Genome-phenome Archive (accession ID: EGAS00001002697). Illumina Short Reads Data of HKCI-4, HKCI-9, HKCI-11, and HKCI-5A can be accessed at Sequence Read Archive under accession IDs SRP023539 (HKCI-4, 9, and 11) and SRA061758 (HKCI-5A).

pathways.[1] It is therefore not surprising that dysregulation of mRNA splicing can contribute to almost all hallmarks of cancer, including sustained proliferation, cell invasion and metastasis, and stemness.

Hepatocellular carcinoma (HCC) is the third leading cause of cancer mortality worldwide. Cumulating studies have unveiled genetic alterations harbored within HCC tumors.[2] Earlier genome-wide analyses by our group, and others, have shown common genomic imbalances in HCC, among which gain of chr.1q21-23 is considered an early initiating event.[3] More recent next-generation sequencing revealed somatic mutations in driver genes of HCC, which shed insight on disease mechanism and potential therapeutic targets.[4] Aberrant RNA splicing has also been implicated in the liver carcinogenesis. Studies on individual genes, including *CD44*[5] and *CDH17*,[6] suggested that the AS isoform expressed correlated with tumor stage, early recurrence, and shorter patient survival. Although interest in defining cancer-associated AS continues to increase, the specific biological function harnessed by many of the AS variants in conferring oncogenic advantages remained minimally investigated. To date, there are only a handful of such studies described in HCC.[7-9] Moreover, although many observational studies have illustrated presence of aberrantly spliced genes in human cancers, the AS landscape of human HCC remains largely unexplored.

Even though studies have demonstrated that anomalous AS of mRNA precursors plays important roles in cancer development,[1] this mechanism has largely been undermined because of technical limitations in systematic analysis. Nonetheless, over the past decade, technologies for genome-wide study of AS events have evolved. The low-sensitivity and low-reliability probed-based isoform microarray has been replaced by high-throughput high-sensitivity and low-error rate RNA-seq of second-generation sequencing (SGS). However, RNA-seq of SGS encounters severe limitations by its short reads generated, usually 101 bp, in reconstructing the full-length transcript and identifying isoforms with complex AS events. Single-molecule real-time (SMRT) sequencing, also referred to as third-generation sequencing, overcomes these limitations by producing long reads >10 kb in length, which encompass virtually all human transcripts from the 5' end to their poly-A.[10] However, intact RNA of high quality is prerequisite to the generation of full-length complementary DNA (cDNA) libraries for SMRT sequencing. More recently, hybrid sequencing is considered the ideal approach to study global AS events by taking advantage and integrating full-length transcript detection from long reads and improved base calling accuracy by short reads correction of errors.[11]

In this study, we undertook the challenge to profile the entire transcriptome composition of

View this article online at wileyonlinelibrary.com.
DOI 10.1002/hep.30500

Potential conflict of interest: Nothing to report.

ARTICLE INFORMATION:

From the [1]Department of Anatomical and Cellular Pathology; [2]State Key Laboratory in Translational Oncology; [3]State Key Laboratory of Digestive Disease, Sir YK Pao Centre for Cancer, The Chinese University of Hong Kong, Shatin, Hong Kong, China; [4]Department of Biomedical Sciences, City University of Hong Kong, Kowloon Tong, Hong Kong, China; [5]School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, China.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Xin Wang, Ph.D.
Department of Biomedical Sciences
City University of Hong Kong
Tat Chee Ave, Kowloon Tong, Hong Kong, China
E-mail: xin.wang@cityu.edu.hk
Tel: +1-852-3442-2367
or

Nathalie Wong, DPhil
Department of Anatomical and Cellular Pathology
The Chinese University of Hong Kong
Prince of Wales Hospital
30-32, Ngan Shing St, Shatin, Hong Kong, China
E-mail: natwong@cuhk.edu.hk
Tel: +1-852-3505-1128

isoforms expressed in HCC. We deployed low-passage patient-derived HCC cultures from eight cases, and compared results with immortalized hepatocyte line and normal liver. SMRT isoform sequencing and SGS RNA-seq were applied on the same specimen, and computed transcript discovery and reconstruction based on the combined hybrid sequencing information from error-corrected long-reads and short-read counts. A comprehensive map of annotated AS variants was obtained. In addition, we discovered unannotated spliced isoforms that are not reported in any database and corroborated many of these predicted unannotated and annotated variants to be tumor specific from rigorous validations in primary HCC tumors and matching nontumoral liver. Although the number of AS events observed in different cancer types has grown in recent years with the use of advanced technologies,[12] relatively few reports have embarked on functional studies. Here, we focused functional investigations on *ARHGEF2* variants (v1 and v3) that arise from the common amplified site chr.1q22 of HCC and confirmed their biological significance in underscoring two major cancer hallmarks, namely cancer stemness and epithelial-to-mesenchymal transition (EMT)-mediated cell invasion and migration.

Materials and Methods

CELL CULTURE

Patient-derived HCC cell cultures (HKCI-2, 4, 5A, 8, 9, 11, C1, C2, and C3) were maintained in AIM-V medium supplemented with 10% fetal bovine serum (FBS; Life Technologies). Hep3B, immortalized human hepatocyte cell line (MIHA), and L02 were maintained in Dulbecco's modified Eagle's medium supplemented with 10% FBS.

PATIENTS SPECIMENS

Paired HCC tumor and adjacent nontumoral liver tissues were collected from patients who underwent curative surgery at Prince of Wales Hospital, Hong Kong. Human sample collection protocol was approved by The Joint Chinese University of Hong Kong – Hospital Authority New Territories East Cluster Clinical Research Ethics Committee. Informed consent was obtained from each patient recruited. Diagnosis of HCC was confirmed by histology.

SMRT SEQUENCING

Total RNA was extracted using RNeasy Mini Kit (Qiagen). First-strand cDNA synthesis was performed using the SMARTer PCR cDNA Synthesis Kit (Clontech). Single-stranded cDNAs were amplified for 11 cycles by PCR using KAPA HiFi polymerase (Kapa Biosystems). Amplified cDNA was loaded on the BluePippin System for size selection. cDNAs with non-size selection or separate size fractions (1-2 kb, 2-3 kb, 3-6 kb, and 5-10 kb) were subjected to SMRTbell library construction using the DNA Template Library Preparation Kit (Pacific Biosciences). Template libraries were sequenced on a Pacific Biosciences RS II sequencer using the DNA Sequencing Kit (Pacific Biosciences). Basic quality information of PacBio SMRT-seq data was obtained by aligning long reads to human reference genome (hg19) with Genomic Mapping and Alignment Program and examined for quality by AlignQC[13] against reference transcriptome of RefSeq gene annotation (Supporting Table S1).

RNA SEQUENCING

Integrity of RNA extracted from cell lines was examined with the Bioanalyzer 2100 (Agilent). cDNA libraries were prepared using the Illumina TruSeq RNA Sample Preparation kit (Illumina). The median insertion size of each library was around 300 nucleotides. Library quality was measured on Agilent 2100 Bioanalyzer. Paired-end libraries were sequenced on the Illumina HiSeq 2000 platform (2 × 100 nucleotide read length) to obtain at least 100 million total reads per sample. FASTQ files were generated by CASAVA.

Bioinformatic analysis and experimental procedures of other supporting verifications and functional assays are described in Supporting Methods.

Results

LANDSCAPE OF ALTERNATIVE SPLICING VARIANTS

SMRT sequencing captured full-length transcripts (transcription start sites, splice sites, and poly-A sites) in eight patient-derived HCC cases and immortalized human hepatocyte MIHA that generated a total of 2,731,554 circular consensus sequencing (CCS) reads with consensus of full-pass cDNA molecules and 5,430,909 partial-CCS

reads that support quality consensus sequence from overlapping reads to CCS. Together, the CCS and partial-CCS reads provided confidence in mapping alignment and high coverage read data. We obtained CCS reads with median of 2,395 bp (quartiles 1,681-3,833 bp) (Supporting Fig. S1A). The length of partial-CCS reads, on the other hand, have a median of read length at 2,916 bp (quartiles 1,695-4,556 bp). In 99% of long reads, the ratio of continuous long-read length to CCS read length was 2 to 15 (Supporting Fig. S1B), suggesting that the majority of sequences passed each original cDNA molecule at least twice, which on overlaid built consensus sequences of high accuracy.

We used hybrid sequencing pipeline, the SpliceMap-LSC-IDP algorithm,[14] to detect AS events and define isoforms. Aligned to hg19 reference genome, 84% error-corrected long reads showed high consensus mapping quality (Supporting Fig. S1C). Isoforms were detected and predicted by isoform detection and prediction (IDP), which uses both junction detections and alignment of error-corrected long reads to establish reliability in statistical modeling (Supporting Table S2). Using stringent criteria (see Supporting Methods), we identified 8,990 full-length transcripts (from 7,250 genes), including 8,292 annotated coding transcripts, 120 long noncoding RNAs, 216 noncoding transcripts, and 362 unannotated isoforms of known genes (Fig. 1A). The latter unannotated isoforms denote unannotated splice junctions that have not been previously reported in RefSeq or GENCODE. Among the annotated genes, 1,414 genes (coding and noncoding) transcribed two or more dominantly expressed isoforms (Fig. 1B). These variants arose from either AS or alternative transcription initiation/termination (Supporting Fig. S2A). Interestingly, 233 variants of our newly discovered unannotated AS (61.6%) represent another alternative isoform(s) of 223 known genes (Fig. 1C).

Among the 1,414 genes with splice variants, 83.8% (1,185 genes) expressed only annotated isoforms. The most frequent AS types found in annotated isoforms were skipping exon (SE), followed by alternative 5' and 3' splice sites (A5 and A3). SE accounted for more than a third of these variants (40%), and A5 and A3 each accounted for 18%. Retained intron (RI), on the other hand, only contributed to 10% of AS events (Fig. 1D). The annotated isoforms contain open reading frames (ORFs) of median 1,134 bp (quartiles 720-1,677 bp) (Fig. 1E). To evaluate the role of proteins derived from annotated isoforms, we performed hallmark pathway analysis using gene set enrichment analysis (GSEA). Interestingly, we observed enrichment of many cancer-related pathways, with the top-ranked pathways highlighting mammalian target of rapamycin complex 1 (mTORC1) signaling, G2M checkpoint, EMT, and the DNA repair process (Fig. 1F). The genes involved in mTORC1 signaling pathway are mostly metabolically related, such as pyruvate dehydrogenase kinase 3 (*PDK3*), aldolase, fructose-bisphosphate A (*ALDOA*), succinate dehydrogenase complex subunit C (*SDHC*) (glycolysis), isopentenyl-diphosphate delta isomerase 1 (*IDI1*), and star related lipid transfer domain containing 4 (*STARD4*) (cholesterol metabolism), but also included some related to cell cycle, such as cyclin-dependent kinase inhibitor 1A (*CDKN1A*). G2M checkpoint included important cell cycle controlling genes, such as *MYC*, cell division cycle 20 (*CDC20*), aurora kinase A (*AURKA*), cell division cycle 25B (*CDC25B*), cell division cycle 27 (*CDC27*), and CDC28 protein kinase regulatory subunit 1B (*CKS1B*). Splicing variants of these genes might represent an alternate mechanism for regulating these cancer-related pathways in HCC.

UNANNOTATED AS VARIANTS

We further characterized the unannotated AS isoforms by comparing them to the annotated isoforms, in terms of their transcript lengths, causative AS events, ORF length, and expression levels. On average, unannotated isoforms had shorter transcript length than annotated ones (Supporting Fig. S2B), which might be due to a large proportion (52%) of unannotated isoforms having a late transcription initiation and early transcription termination (Supporting Fig. S2C). On comparing the type of AS event, unannotated isoforms showed a higher percentage of RI but less SE than annotated isoforms (Fig. 1D). The overall shorter transcript despite frequent RI could be explained by the fact that intron retention is known to cause nonsense-mediated decay (NMD) and mRNA instability by the introduction of premature termination codons (PTCs).[15] The combined effects of RI and late transcription initiation/early transcription

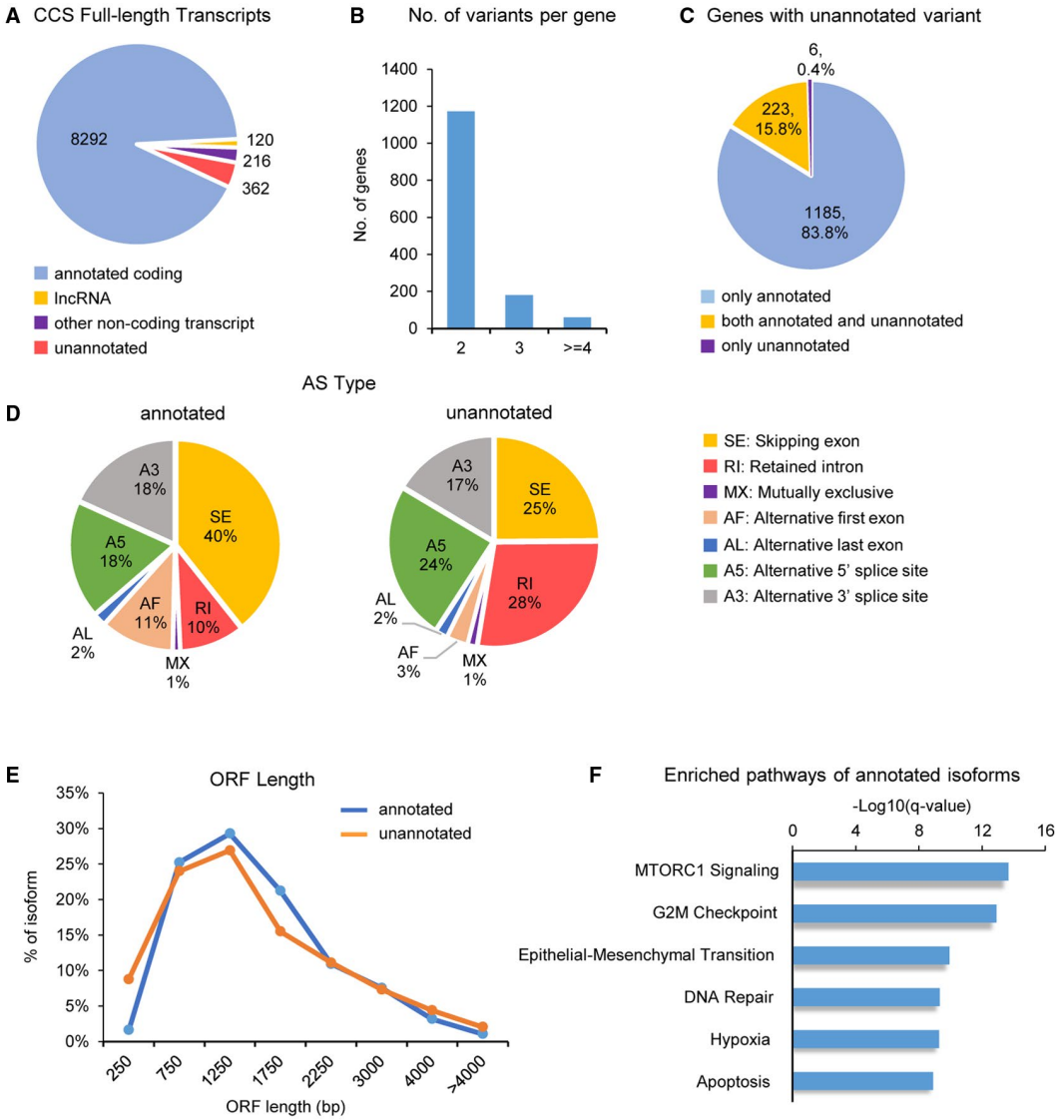

FIG. 1. RNA splicing landscape. (A) Proportion of different transcript categories identified by hybrid sequencing. (B) Number of variants per gene. (C) Proportion of genes with annotation and unannotated variants. (D) Distribution of AS types in annotated and unannotated variants. (E) Distribution of ORF lengths. Annotated and unannotated isoforms have similar ORF length distribution (Median: annotated 1,134 bp, unannotated 1,044 bp). (F) Gene set enrichment analysis shows enrichment of cancer-related pathways in genes with annotated isoforms.

termination within a single transcript could be another contributory factor. Indeed, such co-occurring events could be detected in 45.6% (36/79) of isoforms with RI.

To predict the potential effect of AS on protein translation, we defined the ORF of unannotated isoforms using ORF Finder. The unannotated isoforms had a similar ORF length distribution compared with annotated isoforms, with the median length at 1,044 bp (quartiles 575-1742 bp) (Fig. 1E). For most of the unannotated isoforms (70%), their ORFs were changed, but remained in-frame compared with their annotated counterparts (Fig. 2A), suggesting that they might be translated into a paired and even functional variant. Only 15% of the unannotated isoforms had frameshift (Fig. 2A), which might result in untranslatable mRNA or nonfunctional proteins. The changes in ORF led to early stop codons in 33% of unannotated isoforms (Fig. 2B). We found that the majority (94%) of AS events ensuing an in-frame ORF

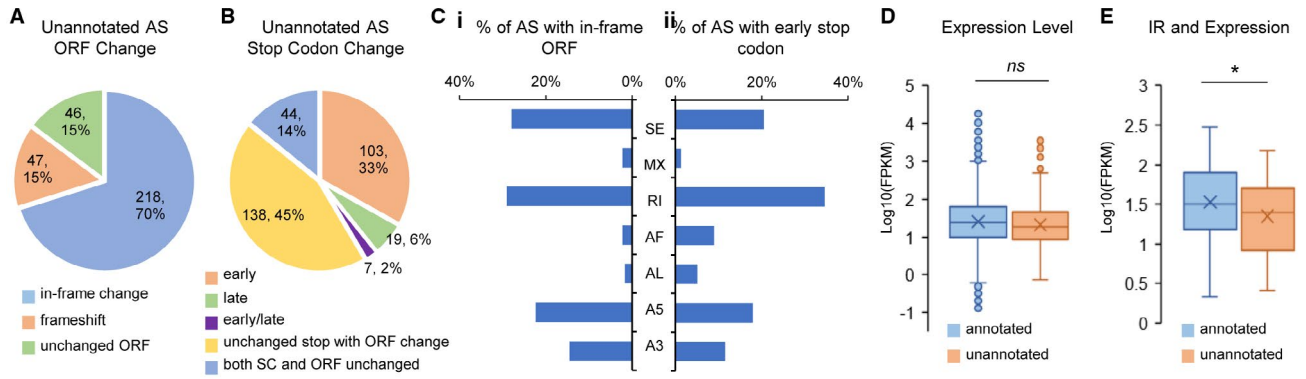

FIG. 2. ORF changes and expression level of unannotated variants. (A) Percentage of isoforms with ORF change among unannotated isoforms. (B) Percentage of stop codon types in unannotated isoforms. SC denotes stop codon. (C) Percentage of AS types with in-frame ORF change and early stop codon events in unannotated isoforms. (i) Four major types of AS (SE, RI, A5, and A3) account for 75% of in-frame ORF changes. (ii) RI accounts for 35% of all early stop codon events. (D) Expression level of isoforms. No significant difference in expression level between annotated and unannotated isoforms ($P = 0.357$, unpaired t test). (E) Pairwise comparison for expression level of annotated and unannotated isoforms with retained introns of the same genes. Unannotated isoforms with retained intron have significantly lower expression compared with their annotated counterparts (Mean FPKM: annotated 59 vs. unannotated 37, $P = 0.02$ by paired t test).

were SE, RI, A5, and A3 (Fig. 2Ci). Nonetheless, RI had stronger association with early stop codon when compared with other AS types, which accounted for more than one third of early stop codon events (Fig. 2Cii). Because intron retention (IR) is known to cause NMD of mRNA by introducing PTC,[15] we further examined whether high frequency of IR in unannotated isoforms affected their expression level. We found that the expression level of unannotated isoforms was not significantly different from annotated isoforms ($P = 0.357$, unpaired t test) (Fig. 2D). However, when compared pairwise, the unannotated isoforms that underwent IR had significantly lower expression than their annotated counterpart without IR ($P = 0.02$, paired t test) (Fig. 2E), suggesting that IR likely led to reduced expression through NMD.

Recurrent detections of the 362 unannotated isoforms were common, with 108 isoforms identifiable in three or more patient-derived HCC lines. Although less frequent than the annotated isoforms, such recurrent frequency suggested these unannotated isoforms could be dominantly expressing variants as well (Fig. 3A). This led us to speculate the possible cause underlying the generation of these unannotated isoforms in HCC. We postulated dysregulation of splicing factors (SFs) could be a contributory factor and examined the correlation between expression of SFs and the incidence of AS events. We found SFs that strongly correlated with two major AS events of the unannotated isoforms, namely RI and SE (Fig. 3Bi,Bii). Interestingly, most RI-positively correlated SFs were negatively associated with SE, and vice versa. For instance, up-regulation of LSM4, U2 small nuclear RNA auxiliary factor 1 (U2AF1), small nuclear ribonucleoprotein U1 subunit 70 (SNRNP70), and dead-box helicase 41 (DDX41) positively correlated with RI, but negatively correlated with SE. On the other hand, down-regulation of serine and arginine rich splicing factor 1 (SRSF1), serine and arginine rich splicing factor 2 (SRSF2), SNW domain containing 1 (SNW1), and small nuclear ribonucleoprotein D1 polypeptide (SNRPD1) correlated with SE, but negatively correlated with RI (Fig. 3Bi,Bii). Our analysis suggested that a subgroup of deregulated SFs was linked to each AS event of RI and SE, and would seem mutually exclusive.

To systemically evaluate the functions of unannotated isoforms expressed, we performed gene ontology (GO) enrichment analysis for their molecular function. We found enriched GO categories for RNA binding, enzyme binding and regulation, transcription activity and chromatin binding, kinase binding, and protease regulation (Fig. 3C). To corroborate existence of many of these predicted unannotated transcripts in HCC, we carried out validation studies in paired primary tumor and adjacent nontumoral liver, in

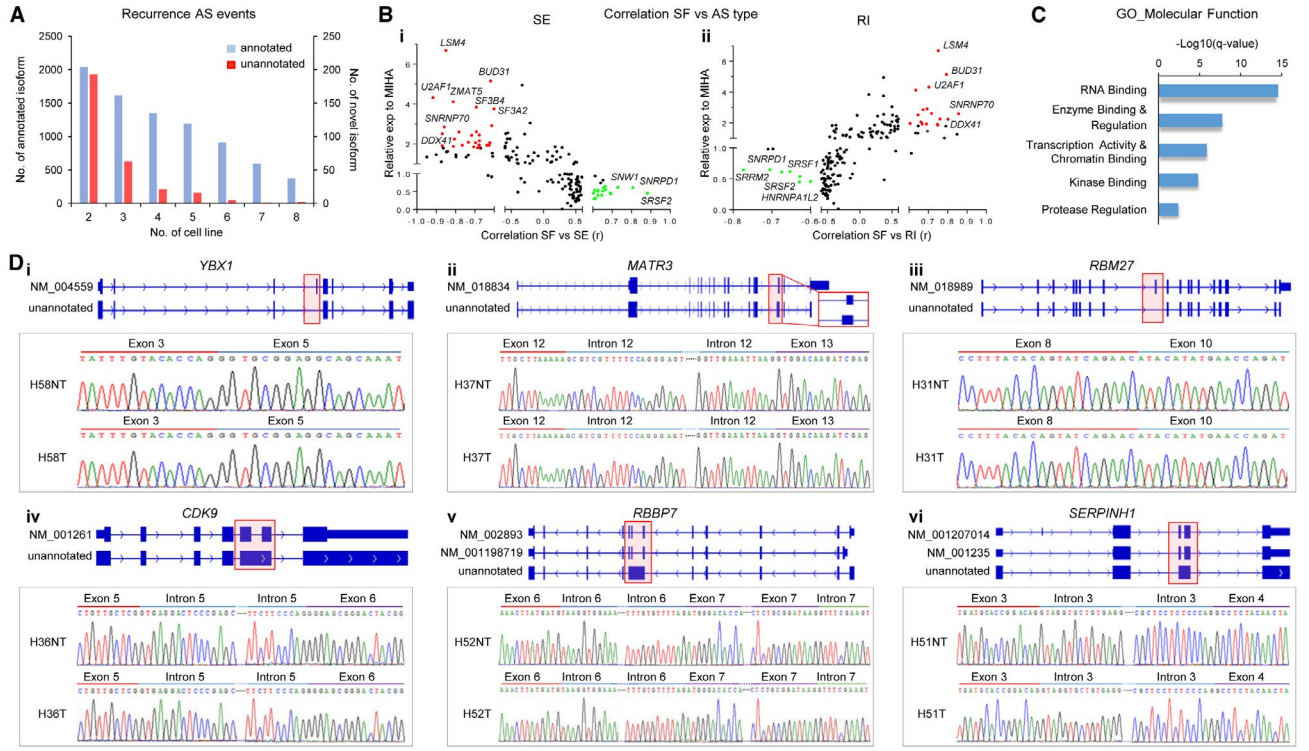

FIG. 3. IDP predicted unannotated isoforms verified in HCC tumors and nontumoral livers. (A) Recurrent incidence of annotated and unannotated isoforms in patient-derived HCC cell cultures. (B) Correlation between expression of splicing factors and frequency of skipping exon (i) and retained intron (ii). X axis: Pearson correlation between FPKM of SFs and percentage of genes with SE or RI within the multi-isoform genes across MIHA and eight patient-derived HCC lines. Y axis: relative FPKM of SFs in patient-derived HCC lines compared with MIHA. Red dots: SFs with 1.8-fold up-regulation in HCC lines. Green dots: SFs with 0.6-fold down-regulation in HCC lines. (C) Molecular function of unannotated isoforms indicated by gene ontology analysis. (D) Verification of unannotated isoforms of *YBX1, MATR3, RBM27, CKD9, RBBP7*, and *SERPINH1* by Sanger sequencing in HCC tumors (T) and paired adjacent nontumoral liver tissues (NT). Maps of unannotated isoforms and their annotated counterparts shown with exon arrangement and AS sites highlighted (red rectangle).

addition to IDP-determined patient-derived cell cultures. Within the top three GO categories, candidate variants were selected based on fragments per kilobase of transcript per million mapped reads (FPKM) levels and published literature on the annotated counterpart. Unannotated isoforms of RNA-binding proteins (Y-box binding protein 1 [*YBX1*], matrin 3 [*MATR3*], RNA binding motif protein 27 [*RBM27*]), chromatin modifier and transcription regulator (cyclin dependent kinase 9 [*CDK9*], RB binding protein 7 [*RBBP7*]), and enzyme regulator (serpin family H member 1 [*SERPINH1*]) were assessed. All isoforms were detected by quantitative RT-PCR with TaqMan probes specific to the new splice junction. Interestingly, these new variants of *MATR3, RBM27*, and *RBBP7* displayed significant up-regulation in tumor relative to nontumoral liver, whereas *YBX1* showed common down-regulation in HCC ($P < 0.001$, paired t test) (Supporting Fig. S3A). RT-PCR products were also Sanger sequenced, which further confirmed IDP predictions and the presence of these unannotated isoforms in HCC (Fig. 3D; Supporting Fig. S3B).

TUMOR-SPECIFIC ISOFORMS

To define HCC-specific AS variant, we filtered out organ-related transcripts and isoforms common in normal liver, MIHA, and patient-derived HCC cells. A total of 2,057 transcripts with FPKM above null in two or more patient-derived HCC lines were defined as preferentially and recurrently expressed in patient-derived HCC cells (Fig. 4A; Supporting Table S3). The extent to which AS produced tumor-specific isoforms was further analyzed

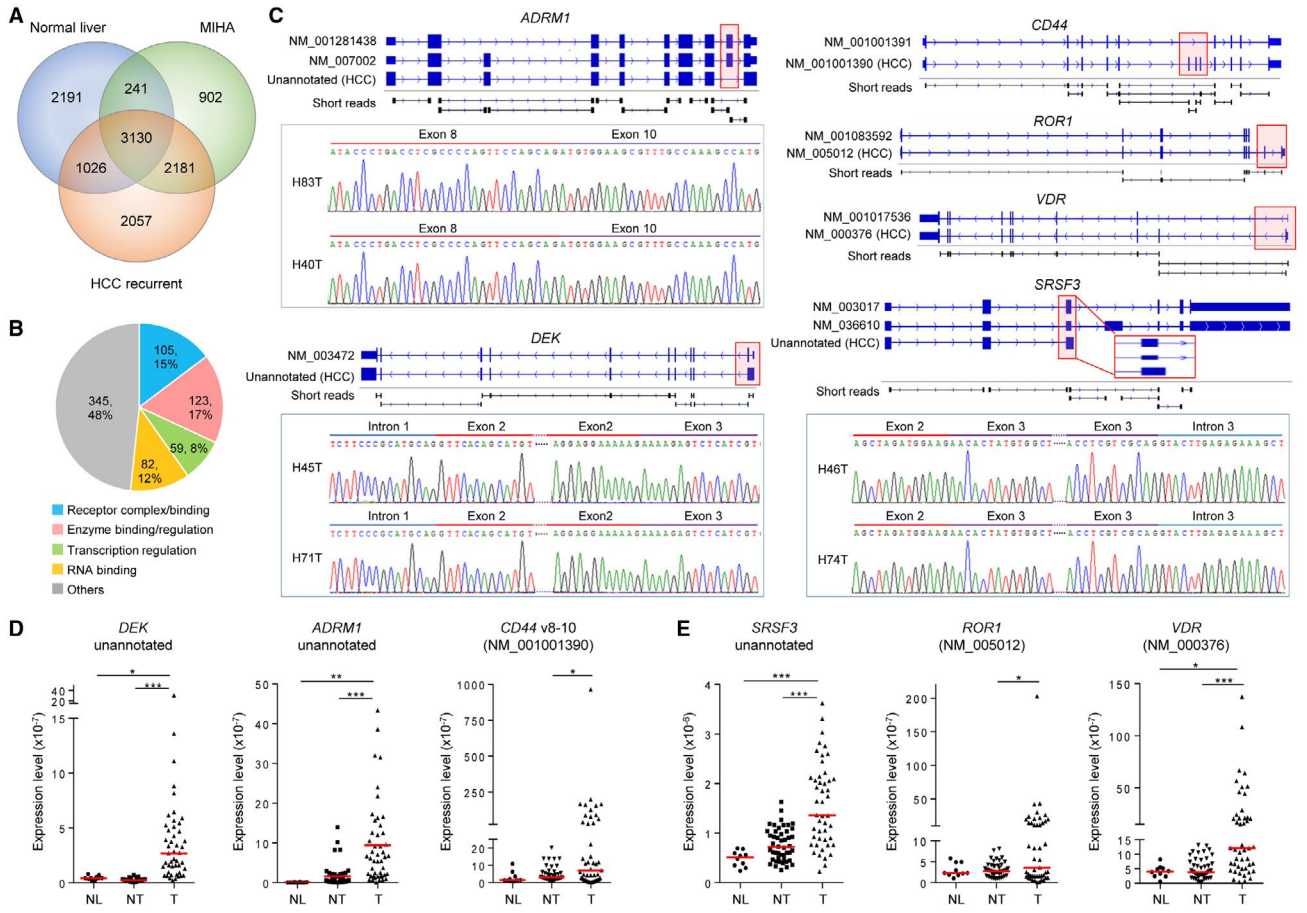

FIG. 4. HCC-specific isoforms. (A) Venn diagram on the number of common and specific isoforms in MIHA, normal liver, and patient-derived HCC cells. (B) GO of recurrent tumor-specific AS isoforms. (C) Maps of HCC-specific isoforms shown with exon arrangement and AS sites highlighted (red rectangle). Sanger sequencings show unannotated junctions of *SRSF3*, *DEK*, and *ADRM1* in HCC tumors. (D) Quantitative PCR detected significant up-regulation of unannotated DEK, ADRM1, and *CD44v8-10* isoforms in HCC tumor (T) compared with nontumoral adjacent liver (NT) and normal liver (NL). NL versus NT, NL versus T by unpaired *t* test. NT versus T by paired *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (E) Quantitative PCR indicated progressive up-regulation of isoforms *SRSF3*, *ROR1*, and *VDR* from NL to T. Comparisons between groups were performed by one-way analysis of variance with linear trend test and Tukey post-hoc test. Trend test: *SRSF3* and VDR, $P < 0.001$; ROR1, $P < 0.05$. Tukey test: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

in these 2,057 transcripts. We found 714 genes exhibited two or more AS variants that totaled to 892 HCC-specific isoforms, with 166 being unannotated. Further GO analysis categorized 51.7% of these genes (369/714 genes) to be enriched for enzyme binding/regulation, receptor complex/binding, RNA binding, and transcription regulation (Fig. 4B; Supporting Table S3).

We next assessed the tumor specificity of these AS variants in a series of primary HCC tumors and their paired adjacent nontumoral liver (n = 47), and normal livers (n = 10). Candidate isoforms from GO categories were selected based on their high FPKM expressions. These included receptor complex (adhesion regulating molecule 1 [*ADRM1*], *CD44v8-10*, receptor tyrosine kinase like orphan receptor 1 [*ROR1*], vitamin D receptor [*VDR*]), chromatin modifier (*DEK*), RNA-binding protein (serine and arginine rich splicing factor 3 [*SRSF3*]), and enzyme regulator (Rho/Rac guanine nucleotide exchange factor 2 [*ARHGEF2*]). Given the importance of receptor splice variants in carcinogenesis, more candidates were selected from this category. We validated the presence of these isoforms by Sanger sequencing in the same patient-derived HCC cells from which they were first identified by

IDP and also in primary HCC tumors (Fig. 4C; Supporting Fig. S3B). Studies on enzyme regulator *ARHGEF2* are detailed in the next section. Fig. 4C shows examples of CCS and short reads alignment on HCC-specific isoforms.

Quantitative PCR using TaqMan junction probe confirmed all variants to be significantly up-regulated in HCC tumors compared with their nontumoral liver ($P < 0.05$, paired t test) (Fig. 4D). Notably, negligible expressions of DEK and ADRM1 variants were consistent in all normal livers, and in most cases for CD44v8-10. This in turn reflected distinct and specific expressions of DEK, ADRM1, and CD44v8-10 variants in HCC tumors. Compared with paired nontumoral liver, we observed a median overexpression of DEK at 10.73-fold (quartiles, 5.66-24.03 folds) and ADRM1 at 8.33-fold (quartiles, 2.36-32.45 folds). For both DEK and ADRM1, more than 10-fold up-regulation could be readily detected in ~50% of cases, with few cases reaching as high as more than 100-fold. Although less pervasive, 10-fold up-regulation of CD44v8-10 could be found in ~28% of cases with a median fold gain suggested at 2.23 (quartiles, 0.66-10.52 folds). The curtailed effect, despite negligible expressions in normal livers, was attributed to these AS variants that were detectable in adjacent nontumoral livers as well, albeit at much lower levels than tumors. Similarly, a stepwise progressive increase in the expression of SRSF3, ROR1, and VDR variants from normal liver and nontumoral liver adjacent to HCC was also evident (one-way analysis of variance test $P < 0.05$) (Fig. 4E). Because HCC-adjacent nontumoral livers are invariably cirrhotic or fibrotic, they are well recognized as the premalignant state of HCC. Our findings might thus have implications for these AS variants in predisposing risk to HCC development.

BIOLOGIC AND CLINICAL SIGNIFICANCE OF *ARHGEF2* VARIANTS

We focused downstream investigations on an enzyme regulator, ARHGEF2 (variants v1 and v3) (Fig. 5A), which is a Rho guanine nucleotide exchange factor (RhoGEF) that plays essential role in activating oncogenic RhoA signaling in HCC.[16] Moreover, both variants are expressed from chr.1q22 locus, which has been implicated in the initiation stage of

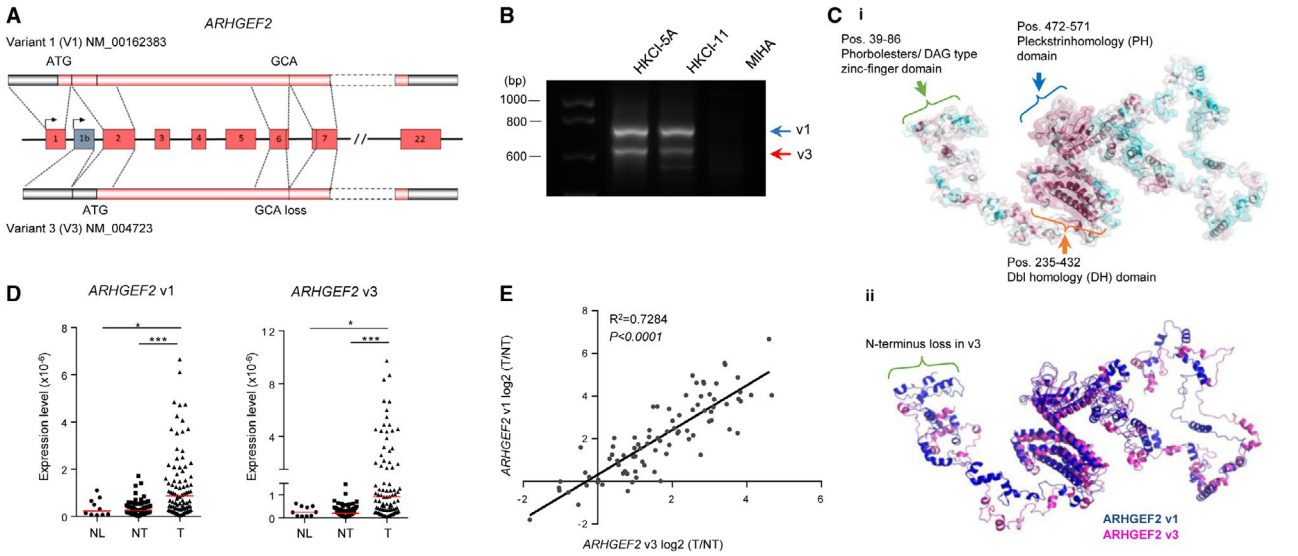

FIG. 5. Expression of *ARHGEF2* variants in patients with HCC. (A) Exon arrangement of *ARHGEF2* variants. (B) Verification of IDP predicted *ARHGEF2* v1 and v3 in patient-derived HCC lines by RT-PCR. Supporting Fig. S4 shows confirmatory Sanger sequencing. (C) 3D structure prediction of v1 and v3 proteins. (i) Structure and functional domains of v1 protein. (ii) Superimposed image of v1 and v3 structures. Blue: v1; Magenta: v3. (D) *ARHGEF2* v1 and v3 mRNA expression in normal liver (NL), adjacent nontumoral liver (NT), and HCC tumor (T) by quantitative RT-PCR. NL: n = 10; NT: n = 87; T: n = 87. NL versus NT, NL versus T: unpaired t test. NT versus T: paired t test. *, $P < 0.05$; ***, $P < 0.001$. (E) Expression of v1 and v3 correlated positively in HCC tumors. Pearson correlation, $R^2 = 0.7284$, $P < 0.0001$.

liver carcinogenesis.[3] Initial RT-PCR corroborated expression of both *ARHGEF2v1* (NM_00162383) and *ARHGEF2v3* (NM_004723) isoforms in patient-derived HCC cells but not MIHA (Fig. 5B). Sanger sequencing of PCR products further confirmed IDP prediction in the alternate first exons used by v1 and v3 that lead to their difference in 5'untranslated region and downstream AUG start codon (Supporting Fig. S4). In addition, v3 also harbors a 3-base deletion (GCA) from an alternative 3'splice at the boundary of exon 6 and exon7 that corresponds to one amino acid (aa; alanine) loss (Supporting Fig. S4). Hence, *ARHGEF2v1* encodes a protein of 986 aa whereas *ARHGEF2v3* encodes a shorter protein of 959 aa. Structurally, the ARHGEF2 protein is well characterized with highly conserved central domains, including a Dbl homology domain (microtubule binding) and a Pleckstrin homology domain (GEF activity) (Fig. 5Ci). According to their predicted crystal structures, these core domains are maintained in both v1 and v3, despite v3 having one aa less at position 194 (Fig. 5Cii). The change in first exon, however, results in a loss of 81 bases (27 aa) at the N-terminus of v3 compared with v1, which likely corresponded to an uncapped phorbol esters/DAG zinc finger domain in v3 (Fig. 5Cii).

To establish relevance of ARHGEF2 v1 and v3 in patients with HCC, we examined expression of v1 and v3 isoforms in an independent cohort of 87 primary HCC and paired nontumoral liver, and 10 normal livers, by quantitative TaqMan assays. The clinicopathological characteristics of patients are summarized in Supporting Table S4 and detailed in Supporting Table S5. Frequent up-regulations of both v1 and v3 in tumors could be readily detected (Fig. 5D). Compared with their matching nontumoral liver, v1 showed a median up-regulation of 3.05-fold (quartiles, 1.65-7.13 folds), and up-regulation of v3 at a median 4.07-fold (quartiles, 1.39-11.34 folds). Overall, both v1 and v3 appeared to be up-regulated in ~51% of HCC according to their median factor cutoff, and overexpressed by >10-fold in 13.8% for v1 (12/87 cases) and 28.7% for v3 (25/87 cases; $P < 0.0001$) (Fig. 5D). Notably, expressions of v1 and v3 showed strong linear correlation, suggesting common coexpression of these variants in HCC (Fig. 5E). Our data also suggested normal livers showed low or negligible levels of both ARHGEF2 v1 and v3 with no significant difference relative to tumor-adjacent nontumoral livers (Fig. 5D). Notably, some nontumoral livers showed increased expressions of v1 and/or v3 compared with normal livers. We further conducted gel-based RT-PCR, which concurred in suggesting some nontumoral livers displayed elevated expression of v1 and v3, albeit at lower levels than matching tumor (Supporting Fig. S5Ai,B,C). Consistent with TaqMan assay, v1 and v3 expressions were undetectable in gel-based analysis of normal livers (Supporting Fig. S5Aii). Because adjacent nontumoral liver is often considered the precancerous lesion of HCC, our finding may have implication for v1 and v3 expressions in the early carcinogenetic changes.

Using the median fold change as threshold, we defined HCC tumor with high or low expressions of v1 and v3. In correlative analyses with clinicopathologic features of tumors, we found that cases with high v1 and v3 expressions showed significant association with the pathologic presence of microvascular invasion ($P = 0.022$, Fisher's exact test), which is a strong independent predictor for HCC dissemination and metastasis.[17] (Fig. 6A) High v1 expression was also found to be significantly associated with advanced tumor grades 2 and 3 ($P = 0.013$, Pearson chi-squared test), although a trend was suggested for v3 (Fig. 6B). In Kaplan-Meier analysis, high expressions of both v1 and v3 correlated with shorter disease-free survival of patients compared with cases with both low expressions. The group with either v1 or v3 high expression fell in between the both-high and both-low groups ($P = 0.043$, log-rank test) (Fig. 6C). Our findings would suggest that both v1 and v3 contribute to poorer disease-free survival and that their effects might be additive.

To determine whether alternatively spliced isoforms can have specific biological functions, we focused *in vitro* and *in vivo* investigations on ARHGEF2v1 and ARHGEF2v3. Our studies centered on elucidating the functional properties of ARHGEF2 variants in conferring oncogenicity in HCC cells and the difference, if any, between v1 and v3. To ensure endogenous ARHGEF2 did not interfere with functional readouts, *ARHGEF2* knockout HKCI-8 and Hep3B were generated by CRISPR/Cas9-mediated genome editing (Supporting Fig. S6A-C). Re-expression of v1 or v3 in ARHGEF2-deficient cells was restored by ectopically expressing cDNA of each isoform in same cell line (Fig. 7A; Supporting Fig. S7A).

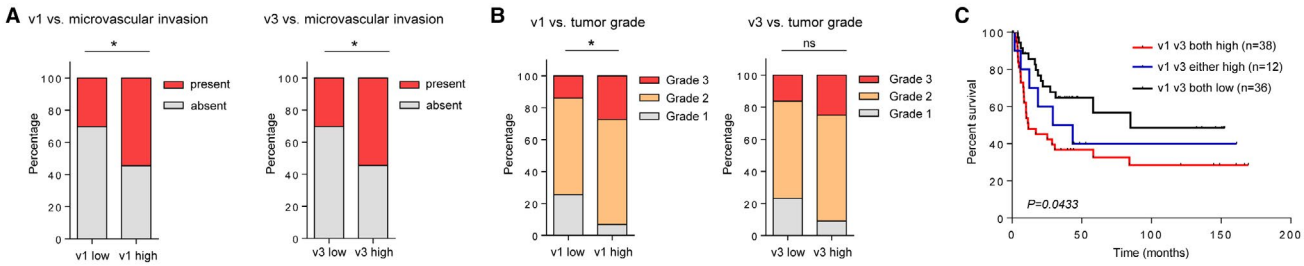

FIG. 6. Clinicopathological correlations of ARHGEF2 variants in patients with HCC. (A) Pathological correlation of v1 and v3 expressions with microvascular invasion. *, $P < 0.05$, by Fisher's exact test. (B) Pathological correlation of v1 and v3 expressions with tumor grade. Pearson chi-squared: v1, $P = 0.035$; v3, $P = 0.164$. (C) Kaplan-Meier survival plot of v1 and v3 expression status versus disease-free survival in 87 patients with HCC. Log-rank (Mantel-Cox) test: $P = 0.043$; log-rank test for trend: $P = 0.0123$. In (A-C), patients who exhibited v1 and v3 mRNA expression above the median level are regarded as v1 and v3 high.

We also overexpressed v1 or v3 in hepatocyte lines MIHA and L02, which endogenously showed undetectable ARHGEF2 expression (Fig. 7A; Supporting Fig. S7A). Successful generation of stably expressing isoforms was confirmed by RT-PCR and immunoblotting and Sanger sequenced to establish the 5' difference and alternative 3'splice site at the exon 6-7 boundary (Supporting Fig. S7B).

Contrary to growth promoting function of many cancer-associated genes, our data suggested that neither v1 nor v3 exerted an effect on cell viability (Supporting Fig. S7C). This was consistent in three cell lines in repeated experiments. Instead, both v1 and v3 augmented cell migration toward chemoattractant (Fig. 7B; Supporting Fig. S8A) and cell invasion through Matrigel (Fig. 7C; Supporting Fig. S8B). However, effect of v3 was constantly more profound than v1, suggesting that although both variants played a role in promoting cell migration and invasion, v3 was distinctively more potent. To affirm functional effects, we knocked down v1 and v3 in overexpressing cell lines, which readily reversed the pro-migration (Supporting Fig. S9A,B) and pro-invasion (Supporting Fig. S9C,D) effects of v1 and v3. Because RhoA is a common substrate of ARHGEF2,[18] we sought to investigate if RhoA signaling underlines the cell motility functions of v1 and v3. When active Rho-guanosine-5'-triphosphate (Rho-GTP) levels were assayed, only v1-expressing cells but not v3-expressing cells demonstrated an increase in RhoA activity compared with vector control (Fig. 7D).

We next explored the possible involvement of EMT in underscoring the ARHGEF2v3 effects. In L02, v3 overexpression induced a morphological change from polygonal to spindly cell shape, which is an indicator of EMT change (Supporting Fig. S10A). Two hallmark EMT proteins, E-cadherin and vimentin, are tightly controlled during the process of EMT through transcriptional regulation.[19] Our results showed notable down-regulation of epithelial adhesion molecule E-cadherin and up-regulation of mesenchymal marker vimentin in v3-expressing HCC cell lines and L02 (Fig. 7E; Supporting Fig. S10B). Correspondingly, we found up-regulation of EMT transcription factors Snail1/Slug and ZEB1 in v3-expressing HCC cell lines (Fig. 7E) and L02 (Supporting Fig. S10B). These changes, on the other hand, were less apparent in v1-expressing cells (Fig. 7E; Supporting Fig. S10B). In accordance with reversing migration and invasion functions, knockdown of ARHGEF2 variants also reversed EMT markers expression induced from v1 and v3 expression (Supporting Fig. S11A,B). Our results would hence suggest that the high migratory and invasive abilities of v3-expressing cells were likely directed by a more intense induction of EMT.

Cancer stemness holds importance in tumor initiation and tumor survival through endowing cells with self-renewal ability.[20] We found both v1- and v3-expressing HCC cells exhibited strong clonogenic ability when seeded at low density (Fig. 8A). Although the ability to form colonies from low-density culture was less evident in MIHA, both v1- and v3-expressing MIHA cells exhibited capabilities to form spheres of increased size and numbers on ultralow attachment plates (Fig. 8B). The low-adhesion cultures also showed high sphere-forming efficiency in Hep3B where overexpression of v3, and to a lesser extent v1, showed an increase in number and size of spheres formed

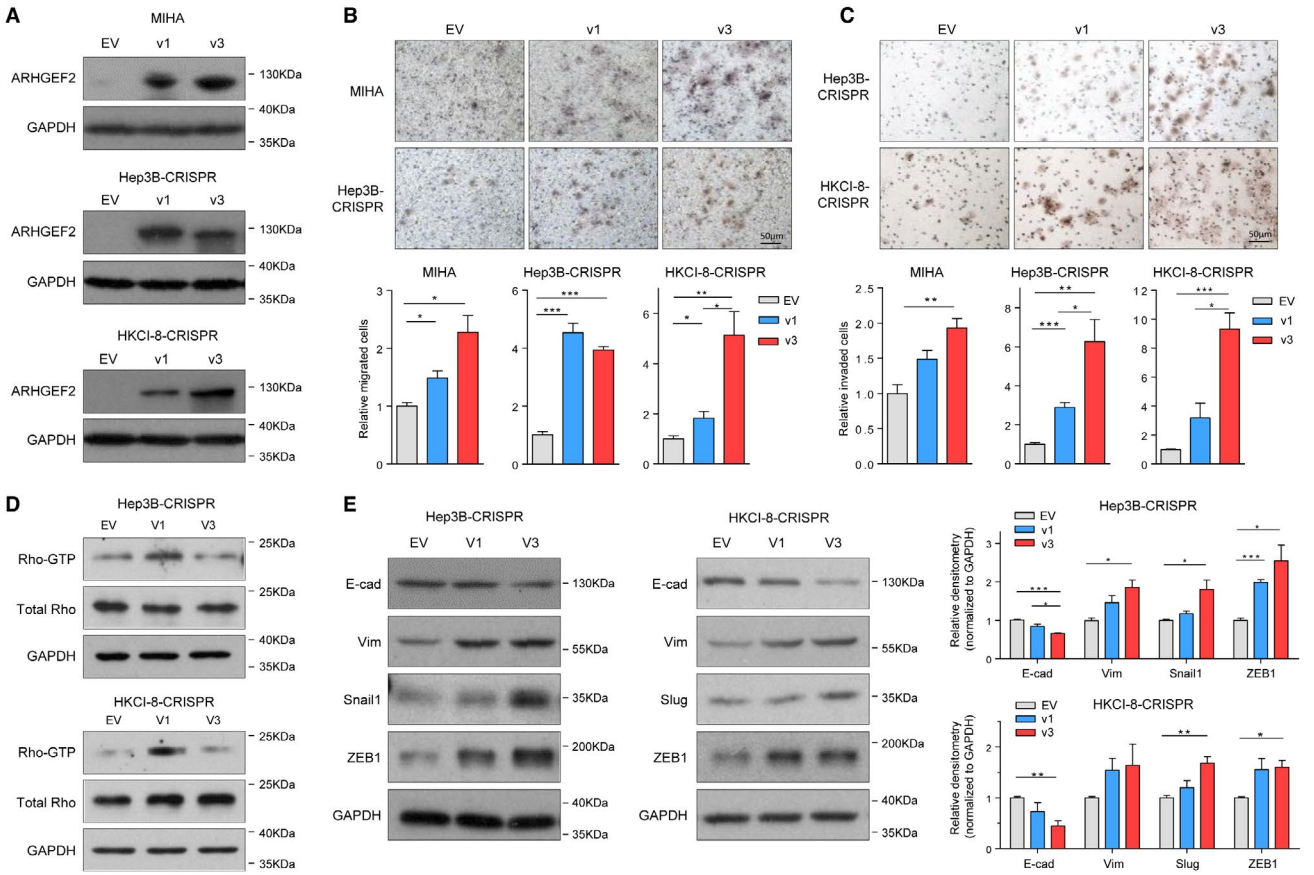

FIG. 7. ARHGEF2 v1 and v3 enhance cell migration and invasion. (A) Overexpression of ARHGEF2 v1 and v3 in MIHA, Hep3B-CRISPR, and HKCI-8-CRISPR shown by western blot. (B) Effect of v1 and v3 overexpression on cell migration. V3 significantly promoted cell migration in MIHA, Hep3B-CRISPR, and HKCI-8-CRISPR, whereas significant cell migration promoted by v1 was only detected in Hep3B-CRISPR (mean ± SEM from approximately three to four independent experiments). Low-level gains in MIHA and HKCI-8-CRISPR were suggested. (C) Effect of v1 and v3 overexpression on Matrigel cell invasion. V3 significantly augmented cell invasion of MIHA, Hep3B-CRISPR, and HKCI-8-CRISPR, whereas v1 promoted low-level gains in three cell lines (mean ± SEM from approximately three to four independent experiments). (D) Effect of v1 and v3 overexpression on Rho activity in Hep3B-CRISPR and HKCI-8-CRISPR. Overexpression of v1, but not v3, increased Rho-GTP level in both cell lines. Rho activity is represented by ratio of Rho-GTP: total Rho. (E) Effect of v1 and v3 overexpression on EMT markers in Hep3B-CRISPR and HKCI-8-CRISPR. Western blot showed down-regulation of epithelial marker E-cadherin, and up-regulation of mesenchymal marker vimentin and EMT transcription factors Snail1, Slug, and ZEB1 in v3-overexpressing cell lines. Densitometry analysis of immunoblots from three independent experiments shown. In (B-E): *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ by t test.

(Fig. 8B). HKCI-8 cells appeared as floating aggregates in low-adherent cultures. The tumor-initiating capacity of HKCI-8 expressing variants was subsequently confirmed by inoculating cells into nonobese diabetic severe combined immunodeficient mice, which showed rapid xenograft tumor formation. HKCI-8 v3-expressing cells presented larger xenograft tumors in mice compared with v1 and vector control (Fig. 8C). There was no significant difference between v1 and vector control, indicating that v3 had more profound tumorigenic advantages than v1. We further examined expression of stemness markers in v1- and v3-expressing cell lines. Consistent with our functional analysis, overexpression of v1 and v3 resulted in up-regulation of stemness proteins in all four cell lines (Fig. 8D-F; Supporting Fig. S12A). Corresponding to a more prominent stemness enhancer effect of ARHGEF2v3, v3 appeared to be a stronger inducer of protein markers than v1. To further support the role of v1 and v3 in enhancing stemness, sphere-forming ability was readily inhibited with knockdown (Fig. 8G), which corresponded to down-regulation of stemness markers (Supporting Fig. S13A,B).

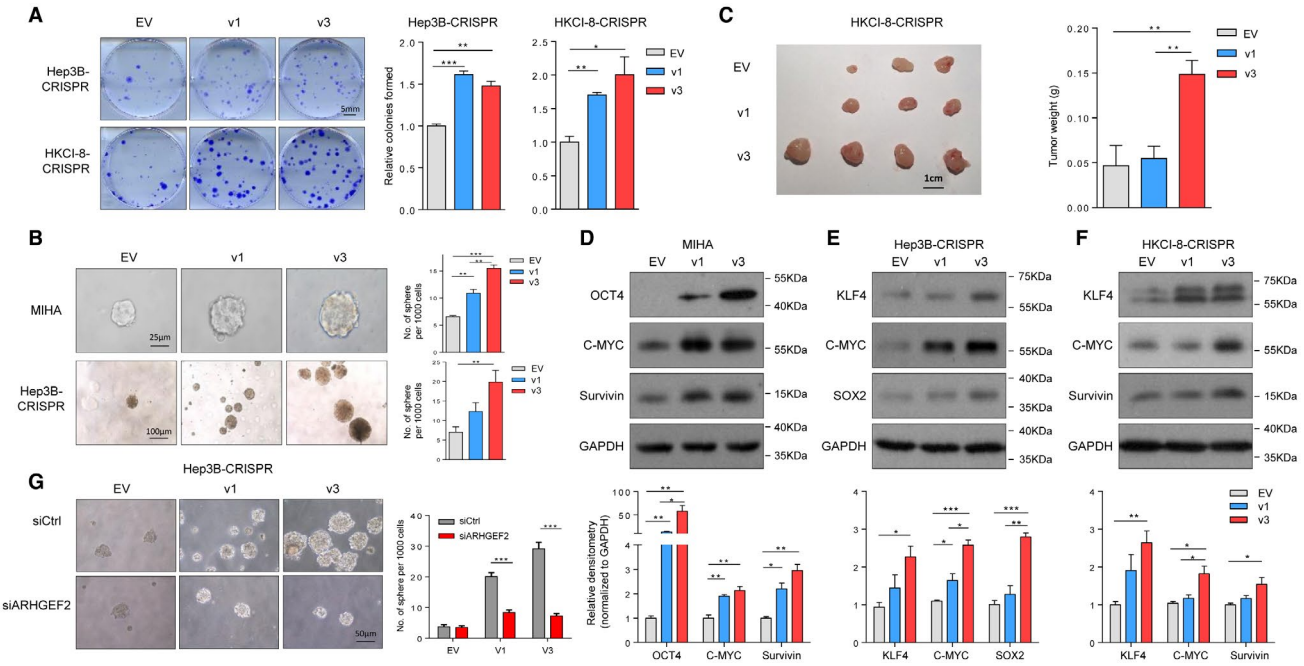

FIG. 8. ARHGEF2 v1 and v3 promote enhance stemness. (A) Effect of v1 and v3 overexpression on colony formation. Both v1 and v3 promoted colony formation in Hep3B-CRISPR and HKCI-8-CRISPR (mean ± SEM from three independent experiments). (B) Effect of v1 and v3 overexpression on sphere formation. V3, and to a lesser extent, v1, promoted sphere formation in MIHA and Hep3B-CRISPR (mean ± SEM from approximately three to four independent experiments). (C) Xenograft formed by subcutaneous injection of vector control (EV), v1 and v3 overexpression HKCI-8-CRISPR into NOD-SCID mice. V3, but not v1 overexpression augmented significantly larger tumors. (D-F) Expression of stemness markers in v1 and v3 overexpression MIHA (D), Hep3B-CRISPR (E), and HKCI-8-CRISPR (F). Western blot showed more potent induction of stemness markers in v3-overexpressing cell lines compared with v1. Densitometry analysis of immunoblots from three independent experiments shown. (G) siARHGEF2 significantly inhibited sphere formation in v1- and v3-overexpressing Hep3B-CRISPR cells compared with siCtrl (mean ± SEM, n = 8). In (A-G): *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ by *t* test.

Discussion

The RNA splicing landscape of HCC has not been fully explored to date, although the influence and impact of alternative splicing is increasingly evident in human cancers. In this study, long-read SMRT sequencing revealed annotated AS isoforms that are common in HCC; however, more excitingly, it also revealed new, previously unidentified splice variants. Analysis of annotated AS isoforms by GSEA indicated that they were highly enriched for the master regulator of cell growth and metabolism (mTORC1 signaling) and regulation of cell cycle checkpoint, including genes such as *MYC, AURKA*, and *CDKN1A* (Fig. 1F). Remarkably, 81.5% of unannotated AS variants exhibited in-frame ORF preservation with their overall distribution of ORF lengths comparable to annotated isoforms (Fig. 1E). GO analysis of unannotated AS variants showed enrichment for molecular functions under the categories of RNA binding, enzyme regulation, and transcription/chromatin activity, which might have implications for these unannotated variants in contributing to cancer development and tumor maintenance. We confirmed existence of a number of predicted new isoforms in tumor and nontumoral liver, and more importantly, many are up-regulated in HCC (Fig. 4D,E; Supporting Fig. S3A). We also explored upstream signals that might drive their expressions. In correlating AS events of unannotated isoforms with SF dysregulations, a number of SF mRNAs appeared to be overexpressed. Few SFs, such as SNRNP70 and U2AF1, showed strong positive correlation with the RI event that might signify the abnormal splicing machinery itself as a major contributory factor. Indeed, both SNRNP70 and U2AF1 are components of major spliceosome that recognize the 5' and 3' splice sites, respectively.[21]

In the oncology field, there is an unmet need for tumor-specific molecules for early cancer detection, diagnosis, and targeting therapy. Their discoveries, however, remain a challenge. Our Venn diagram analysis suggested candidate tumor-specific variants, which we were able to corroborate and quantify by customized splice junction TaqMan probe. HCC-specific expression was highlighted for *ARHGEF2, DEK, ADRM1*, and *CD44v8-10*, which were negligible in normal livers but markedly overexpressed in HCC (Figs. 4D, 5D). DEK plays a role in chromatin topology and transcription.[22] Because DEK epitope can be presented by dendritic cells and stimulate CD8+ T-cell response, it is also considered a tumor-associated antigen.[23] We identified a unannotated isoform with retention of intron 1, which resulted in an extra five aa on the N-terminus. The extra peptide might create a neoantigen that is tumor specific. ADRM1 is a proteasome ubiquitin receptor engaged in ubiquitin–proteasome–dependent protein degradation.[24] The best-known cancer-related function of ADRM1 is its essential role in cell cycle progression.[25] We identified a unannotated isoform of ADRM1 with exon 9 skipped. This splicing results in a frameshift and replacement of the last 69 aa of C-terminal DEUBAD (DEUBiquitinase adaptor) domain by a new 31 aa polypeptide. Although the functional effect of such partial change of DEUBAD domain remains to be determined, the strong up-regulation of this unannotated *ADRM1* isoform is highly suggestive of its potential as a biomarker for HCC. Interestingly, we identified a CD44 isoform, CD44v8-10, to be highly expressed in HCC tumors. The CD44v8-10 isoform differed from the prevalent isoform CD44s at membrane proximal domain by splicing in three extra exons (Fig. 4C). Although we failed to observe isoform switching, because the transmembrane cell-surface receptor of CD44 could be easily recognized by immune cells or cell-selective drug deliveries, tumor specificity of CD44v8-10 might support antigen selection for individualized therapy.

We also identified other tumor-associated candidates S*RSF3*, *ROR1*, and *VDR*, which showed progressive up-regulation from normal liver to premalignant state of liver cirrhosis/fibrosis, and ultimately HCC. It is plausible that these AS variants harbor clonal advantages during the development from precursor lesion to cancer. Of particular interest is the unannotated variant of SRSF3, which showed a significant positive trend of up-regulation ($P < 0.001$, paired t test) (Fig. 4E). SRSF3 is a serine/arginine-rich protein that canonically promotes splicing by recruitment of other splicing factors. SRSF3 protein has two domains. The N-terminal RRM domain recognizes RNA in a sequence-specific manner, whereas the C-terminal RS domain interacts with other proteins.[26] SRSF3 is also an oncoprotein overexpressed in multiple cancer types.[26] High expression of SRSF3 is an unfavorable prognostic predictor in HCC (The Human Protein Atlas). The SRSF3 unannotated variant that we identified has a premature stop codon caused by partial intron 3 retention that would translate a truncated protein. The loss of the second half of RS domain likely affects its interaction with other splicing factors. It is possible that this unannotated isoform might predispose or promote HCC development by affecting splicing of canonical SRSF3 targets.

The extent to which many aberrantly spliced variants contribute to specialized function(s) remains poorly understood. In this study, we specifically addressed this issue by studying variants of the *ARHGEF2* gene. In a separate cohort of HCC specimens obtained from 87 patients, we affirmed common overexpression and tumor dominance of variants v1 and v3. Given the robustness and potential clinical significance of our findings, we centered our subsequent *in vitro* and *in vivo* studies on cancer hallmarks of growth, metastasis, and stemness. Although both v1 and v3 showed enhanced metastatic capabilities and stem cell traits, v3 with a truncated N-terminus consistently demonstrated more potent effects. It is plausible that ARHGEF2 variants might regulate Rho-independent signaling pathways through protein–protein interaction. For instance, ARHGEF2 was shown to enhance oncogenic RAS signaling through its interaction with scaffold protein KSR1, the process of which was independent of the RhoGEF activity.[27] Although v3-specific interacting proteins remain to be defined, our study highlighted that AS allows generation of mRNA transcripts from a single gene that encodes structurally dissimilar protein isoforms of varying functional potency.

The identification and functional establishment of global AS in cancer genome remains a formidable challenge that has been largely unexplored. Limited by the prerequisite of high mRNA quality, we have undertaken AS analysis by long-reads sequencing in

the context of patient-derived HCC cell cultures. A subset of AS isoforms, especially those unannotated variants, was further established in patient-matched HCC specimens and nontumoral liver. Frequent intron retention in unannotated isoforms suggested a possible origin for these aberrant AS in HCC. As shown for variants of ARHGEF2, AS events may have both biological and clinical consequences. Our results underscore the leveraging on next-generation sequencing that could lead to new discoveries of splice variants that may serve as alternative biomarkers and/or molecular targets for development of therapies.

Acknowledgment: We thank the Core Utilities of Cancer Genomics and Pathobiology (CUHK) for providing the facilities and assistance that supported this research. We also thank Pacific Biosciences (PacBio) for their support of SMRT cell sequencing.

REFERENCES

1) Chen J, Weiss WA. Alternative splicing in cancer: implications for biology and therapy. Oncogene 2015;34:1-14.
2) **Shibata T**, **Aburatani H**. Exploration of liver cancer genomes. Nat Rev Gastroenterol Hepatol 2014;11:340-349.
3) Poon TC, Wong N, Lai PB, Rattray M, Johnson PJ, Sung JJ. A tumor progression model for hepatocellular carcinoma: bioinformatic analysis of genomic data. Gastroenterology 2006;131:1262-1270.
4) **Fujimoto A**, **Furuta M**, **Totoki Y**, **Tsunoda T**, **Kato M**, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016;48:500-509.
5) Endo K, Terada T. Protein expression of CD44 (standard and variant isoforms) in hepatocellular carcinoma: relationships with tumor grade, clinicopathologic parameters, p53 expression, and patient survival. J Hepatol 2000;32:78-84.
6) Wang XQ, Luk JM, Leung PP, Wong BW, Stanbridge EJ, Fan ST. Alternative mRNA splicing of liver intestine-cadherin in hepatocellular carcinoma. Clin Cancer Res 2005;11:483-489.
7) Li X, Qian X, Peng LX, Jiang Y, Hawke DH, Zheng Y, et al. A splicing switch from ketohexokinase-C to ketohexokinase-A drives hepatocellular carcinoma formation. Nat Cell Biol 2016; 18:561-571.
8) **Lu X**, **Feng X**, Man X, Yang G, Tang L, Du D, et al. Aberrant splicing of Hugl-1 is associated with hepatocellular carcinoma progression. Clin Cancer Res 2009;15:3287-3296.
9) **Luo ZL**, **Cheng SQ**, **Shi J**, Zhang HL, Zhang CZ, Chen HY, et al. A splicing variant of Merlin promotes metastasis in hepatocellular carcinoma. Nat Commun 2015;6:8457.
10) Gonzalez-Garay ML. Introduction to isoform sequencing using pacific biosciences technology (iso-seq). In: Wu J, ed. Transcriptomics and Gene Regulation. Volume 9. Springer; 2015:141-160.
11) Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 2015;13:278-289.
12) Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R, Levanon EY. Identification of recurrent regulated alternative splicing events across human solid tumors. Nucleic Acids Res 2015;43:5130-5144.
13) Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Res 2017;6:100.
14) Au KF, Sebastiano V, Afshar PT, Durruthy JD, Lee L, Williams BA, et al. Characterization of the human ESC transcriptome by hybrid sequencing. Proc Natl Acad Sci U S A 2013;110:E4821-4830.
15) Jacob A, Smith C. Intron retention as a component of regulated gene expression programs. Hum Genet 2017;136.
16) Cheng IK, Tsang BC, Lai KP, Ching AK, Chan AW, To KF, et al. GEF-H1 over-expression in hepatocellular carcinoma promotes cell motility via activation of RhoA signalling. J Pathol 2012;228:575-585.
17) Zhang X, Li J, Shen F, Lau WY. Significance of presence of microvascular invasion in specimens obtained after surgical treatment of hepatocellular carcinoma. J Gastroenterol Hepatol 2018;33:347-354.
18) Birkenfeld J, Nalbant P, Yoon SH, Bokoch GM. Cellular functions of GEF-H1, a microtubule-regulated Rho-GEF: is altered GEF-H1 activity a crucial determinant of disease pathogenesis? Trends Cell Biol 2008;18:210-219.
19) Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol 2014;15:178-196.
20) Kreso A, Dick JE. Evolution of the cancer stem cell model. Cell Stem Cell 2014;14:275-291.
21) **Dvinge H**, **Kim E**, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer 2016;16:413-430.
22) Pease NA, Wise-Draper T, Privette Vinnedge L. Dissecting the potential interplay of DEK functions in inflammation and cancer. J Oncol 2015;2015:106517.
23) Zheng J, Kohler ME, Chen Q, Weber J, Khan J, Johnson BD, et al. Serum from mice immunized in the context of Treg inhibition identifies DEK as a neuroblastoma tumor antigen. BMC Immunol 2007;8:4.
24) **Qiu XB**, **Ouyang SY**, Li CJ, Miao S, Wang L, Goldberg AL. hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. EMBO J 2006; 25:5742-5753.
25) Randles L, Anchoori RK, Roden RB, Walters KJ. The proteasome ubiquitin receptor hRpn13 and its interacting deubiquitinating enzyme Uch37 are required for proper cell cycle progression. J Biol Chem 2016;291:8773-8783.
26) Corbo C, Orru S, Salvatore F. SRp20: an overview of its role in human diseases. Biochem Biophys Res Commun 2013;436:1-5.
27) **Cullis J**, **Meiri D**, Sandi MJ, Radulovich N, Kent OA, Medrano M, et al. The RhoGEF GEF-H1 is required for oncogenic RAS signaling via KSR-1. Cancer Cell 2014;25:181-195.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.30500/suppinfo.