

Validation of next-generation sequencing-derived circulating mRNA variants as potential diagnostic tools in hepatocellular carcinoma and cholangiocarcinoma

DANIEL ZEULINSKI¹, MAAROUF A. HOTEIT², DAVID E. KAPLAN^{2,3}, TINGTING ZHAN⁴,
CATALDO DORIA⁵, TIMOTHY M. BLOCK¹ and AEJAZ SAYEED¹

¹Baruch S. Blumberg Institute, Doylestown, PA 18902, USA; ²Division of Gastroenterology and Hepatology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA; ³The Corporal Michael J. Crescenz Veterans Administration Hospital, Philadelphia, PA 19104, USA; ⁴Division of Biostatistics, Department of Pharmacology and Experimental Therapeutics, Thomas Jefferson University, Philadelphia, PA 19107, USA; ⁵Capital Health Cancer Center, Pennington, NJ 08534, USA

Received February 2, 2026; Accepted June 19, 2026

DOI: 10.3892/or.2026.9196

Abstract. Liquid biopsy can transform oncology practice through early detection and real-time monitoring of cancer progression and response to treatment. Circulating cell-free tumor RNA offers promise as a novel and emerging alternative to circulating cell-free tumor DNA for detecting the presence and understanding the biology of liver tumors. Both circulating DNA and RNA, which predominantly reside in extracellular vesicles, carry evidence of vital germline and somatic mutations. High impact single nucleotide changes, insertions and deletions in circulating RNA (ctRNA) serve a key role in cancer development and progression. Tumor cells exploit the RNA splicing machinery to induce structural and functional changes in RNA not transcribed from DNA, thereby promoting tumor growth. Examination of the landscape of mutations in ctRNA using RNA-sequencing in patients with hepatocellular carcinoma (HCC), cholangiocarcinoma and liver cirrhosis has revealed widespread single nucleotide changes and splicing aberrations associated with patients with cancer. The present study aimed to characterize these next-generation sequencing-derived high-risk circulating

mutated RNA (ctmutRNA) variants, and to develop TaqMan chemistry-based assays for validation of ctmutRNAs in tumor tissues and plasma samples from patients with HCC. The present study identified a short ctmutRNA panel able to distinguish all patients with liver cancer in the discovery cohort, and splicing aberrations, including intron retentions, in tumor tissues were validated. Finally, the present study demonstrated that the assays detecting ctmutRNA analytes in the plasma, if confirmed in larger patient cohorts, could be effectively used to identify patients with HCC. Furthermore, the outstanding questions and challenges that need to be addressed for the development of a robust blood-based HCC surveillance test are discussed.

Introduction

Circulating proteins and circulating DNA (ctDNA) have been used for a long time in cancer surveillance and diagnostics (1,2). Circulating RNA (ctRNA) has been demonstrated to have additional advantages over proteins and DNA in clinical management (3). RNA is dynamically and differentially expressed across diverse differentiated tissues. The tissue specificity of ctRNAs reflects the tumor's tissue of origin which may offer advantages over ctDNA for cancer detection. The diverse nature of RNA expression can also facilitate the classification of cancer subtypes in early disease stages (4). A previous study from our laboratory reported the landscape of the circulating transcriptome in the liver and demonstrated robust upregulation of liver-specific transcripts in the circulation of patients with hepatocellular carcinoma (HCC) (5). Our previous studies have also demonstrated the detection of circulating pathogenic mutations in RNA [circulating mutated RNA (ctmutRNA)] exclusively associated with patients with HCC and cholangiocarcinoma (CCA) (6,7). In the present study, ctmutRNA refers to a circulating variant RNA analyte identified as a high-risk pathogenic variant by the Genome Analysis Tool Kit (GATK) variant caller and the SnpEff annotation tool. Compared with cell-free/ctRNA species

Correspondence to: Dr Aeja Sayeed, Baruch S. Blumberg Institute, 3805 Old Easton Road, Doylestown, PA 18902, USA
E-mail: aeja.sayeed@bblumberg.org

Abbreviations: HCC, hepatocellular carcinoma; CCA, cholangiocarcinoma; LC, liver cirrhosis; SNP, single nucleotide polymorphism; indel, insertion and deletion; ctRNA, circulating RNA; ctDNA, circulating DNA; ctmutRNA, circulating mutated RNA; dPCR, digital PCR

Key words: HCC, CCA, SNP, indel and splicing variants, ctRNA, surveillance and early detection, reverse transcription-quantitative PCR

detectable in all human subjects, ctmutRNA is exclusively detected in the circulation of patients diagnosed with cancer and corresponds to single nucleotide polymorphisms (SNPs), insertion and deletion (indel), and splicing variants (6,7). RNA-sequencing (RNA-seq) analysis in matching whole plasma samples, purified extracellular vesicles (EVs) and tumor tissues has demonstrated concordance of these pathogenic ctRNA mutations in patients with HCC (6,7). The existence of these circulating variants identified using next-generation sequencing (NGS) and various mutation calling tools must be ascertained through in-depth characterization and validation in HCC tumor tissues. Besides high-risk SNP variants, circulating indel variants and splicing variants that are not necessarily detectable using DNA need to be investigated to confirm that they originate from tumors.

HCC surveillance in routine practice involves risk assessment using several clinical and biological parameters to identify high-risk patients among patients with chronic liver disease (8). ctRNA mutations can considerably improve this prediction and facilitate early detection, patient stratification and effective monitoring post-diagnosis (6,7,9).

SNPs constitute ~90% of sequence variants in humans, and thus represent the most predominant type of genetic alteration (10). These are scattered throughout the genome in both coding and non-coding regions and ~500,000 SNPs residing in coding regions are known to be key to human genetic variation (11). Among these, missense SNPs (non-synonymous SNPs) cause amino acid substitutions, leading to structural and functional changes in proteins (12). The number of SNPs that have been identified covering common polymorphisms is >10 million (13). The sheer scale of high-throughput sequencing has fundamentally changed how genetic data are interpreted. At present, relying strictly on a Reference SNP identifier or a broad database classification to denote a benign polymorphism is no longer clinically reliable (14). Several *in silico* tools such as SnpEff, SIFT (<https://pcingola.github.io/SnpEff/>) and Polyphen-2 (15) are used to predict the functional impact of variants on genes and proteins. Analysis of functional SNPs in cancer-associated genes can improve therapeutic selection and patient outcomes.

Aberrant splicing during liver tumorigenesis produces alternative and tumor-specific isoforms (16). Long-read RNA-seq has revealed diverse splicing variants, including in enzyme regulators, chromatin modifiers, RNA-binding proteins and receptors (17). Primary HCC tumors exhibit extensive differential splicing in numerous genes, including those coding for RNA binding proteins, affecting metabolism and other cancer hallmarks. Alternative splicing in genes such as estrogen receptor, HER2, KRAS, P53, BCL2, MET, cyclin D1 and BRCA1 across several cancer types is known to have prognostic relevance (18-20). While requiring validation via NGS and non-NGS methods in larger cohorts, these high-risk variants, such as the variant found in UDP-glucose pyrophosphorylase 2, show strong potential as biomarkers and molecular therapeutic targets.

In the present study, some novel RNA-seq-derived ctmutRNA variants in HCC tumors were characterized and their validity was demonstrated. Furthermore, assays were developed to demonstrate the feasibility of detecting ctRNA

variants in blood as potential biomarkers for surveillance, early detection and precision therapeutics.

Materials and methods

Human subjects. Plasma samples from a small independent cohort of patients with HCC and patients with liver cirrhosis (LC) without HCC were acquired from the University of Pennsylvania (Philadelphia, PA, USA), and the protocol was approved by the Institutional Review Board #7 at University of Pennsylvania [Philadelphia, USA; protocol no. 828260; Federalwide Assurance (FWA) no. 00004028]. Tumor tissues from patients diagnosed with HCC were obtained from Capital Health Cancer Center (Pennington, NJ, USA), and the protocols were approved by the Institutional Review Board at Capital Health Medical Center (Pennington, USA; protocol no. 00002877; FWA no. 00003248). A total of two tumor tissues were obtained from BioChemed Services, a commercial vendor. The tumor tissues investigated for the validation of ctmutRNAs in the present study were the same as those analyzed by RNA-seq in our original study (5). HCC was diagnosed based on either biopsy or typical enhancement patterns on dynamic contrast-enhanced CT or MRI (later codified as meeting Li-RADS 5 criteria) (21). Staging was performed according to the Barcelona Clinic Liver Cancer (BCLC) system (22) and American Joint Committee on Cancer (AJCC) guidelines (23). Normal liver tissue obtained from a separate, healthy individual was used as a control. The normal human liver tissue was a kind gift from Dr Ramilla Philip (Immunotope Inc.). Demographic and clinical details of all patients whose tumors and plasma were investigated are shown in Table I. Table SI reflects sample-to-figure usage, correlating samples to specific figures in this manuscript.

Plasma processing, EV enrichment and RNA analysis. The plasma samples used in the present study were acquired from our collaborators at the University of Pennsylvania. The maximum time from blood draw to centrifugation at 2,000 x g for 15 min at 4°C and plasma extraction was 2 h. Hemolysis was examined by visually evaluating the pink/red color of the plasma after centrifugation at 2,000 x g for 15 min at 4°C. All specimens used in the present study were non-hemolyzed. As previously reported (6), plasma samples with a single freeze-thaw cycle were spun at 2,000 x g for 30 min at 4°C to remove the cellular debris, followed by total RNA extraction using the Qiagen RNeasy Serum/Plasma Kit (Qiagen, Inc.). For EV isolation, plasma samples (1 ml) were spun at 2,000 x g for 30 min at 4°C to remove the cellular debris before purification of EVs using the ExoQuick kit (System Biosciences, LLC). EVs were further purified by ultracentrifugation at 100,000 x g for 2 h at 4°C. Total RNA from EVs was extracted using the miRNeasy Serum/Plasma Kit (Qiagen, Inc.). Total RNA was also extracted from tumor tissues and a normal liver tissue using the mirVana miRNA Isolation Kit (Ambion; Thermo Fisher Scientific, Inc.). Tissue (100 mg) treated with RNAlater-ICE solution (Invitrogen; Thermo Fisher Scientific, Inc.) was homogenized in a glass homogenizer and processed following the kit protocol, and RNA was eluted in 100 µl RNase-free water. In all cases,

Table I. Demographic and clinical details of patients whose tumor tissues and plasma samples were used for validation of circulating RNA variants identified by RNA-sequencing.

Diagnosis	Patient ID	Specimen	Age, years	Sex	Ethnicity	Etiology	Grade/differentiation	AJCC stage	Tumor size, cm
HCC	CH-001	Tissue	69	M	White	NA	G1/well differentiated	pT1aN0M0	1.9
HCC	CH-004	Tissue	58	M	Hispanic	NA	G2/moderately differentiated	pT3N0	8.6
HCC	CH-006	Tissue	76	M	Hispanic	MASH	G1/well differentiated	pT1bN0Mx	3.0
HCC	CH-007	Tissue	59	M	White	HCV/ alcohol	G3,4/poorly differentiated	pTxpNx	2.8
HCC	BC-103	Tissue	72	F	Black	NA	G2/moderately differentiated	I (pT1, pN0)	Not available
HCC	BC-105	Tissue	74	M	White	NA	G2/moderately differentiated	IIIA [pT3a(m), pN0]	Not available
CCA	CH-002	Tissue	62	M	Black	MASH	G3/poorly differentiated	IA (pT1aNxMx)	4.2
CCA	CH-005	Tissue	83	F	White, Hispanic	HBV/ MASH	G2/moderately differentiated	pT2NxMx	15.6
No diagnosis	Normal liver	Tissue	57	M	White	N/A	N/A	N/A	N/A
HCC	CH-033	Plasma	27	M	Hispanic	HBV	G2/moderately differentiated	pTxpNxM1	1.0
HCC	CH-062	Plasma	60	M	White	HCV	G3,4/poorly differentiated	pTx, pNx	0.8
HCC	CH-103	Plasma	70	M	White	HBV/ MASH	-	I (pT1, pN0)	1.8
HCC	CH-138	Plasma	67	M	Hispanic	HCV/ alcohol	-	IVa (pT3NxM1)	8.8
HCC	CH-144	Plasma	77	M	White	HCV	-	IVb (pT1NxM1)	1.0
HCC	UP-032	Plasma	60	M	White	HCV	-	Ib (cT1bN0M0)	3.2
HCC	UP-050	Plasma	56	M	White	HCV	-	IVA (cT4N1M0)	10.9
HCC	UP-075	Plasma	72	M	White	MASH	-	Ib (cT1bN0M0)	2.6
HCC	UP-097	Plasma	70	M	Black	HCV	-	IIIb (cT4N0M0)	6.0
HCC	UP-120	Plasma	79	F	Asian	HBV	-	Ib (cT1bN0M0)	2.1
LC	UP-001	Plasma	73	M	White	HCV	-	-	-
LC	UP-003	Plasma	78	M	White	MASH	-	-	-
LC	UP-008	Plasma	47	F	White	Alcohol	-	-	-
LC	UP-022	Plasma	66	M	White	MASH	-	-	-
LC	UP-023	Plasma	75	F	White	PBC	-	-	-
LC	UP-025	Plasma	68	F	White	Alcohol	-	-	-

All patient specimens were acquired at the time of diagnosis before initiating any therapeutic treatment. HBV, hepatitis B virus; HCV, hepatitis C virus; MASH, metabolism-associated steatohepatitis; HCC, hepatocellular carcinoma; CCA, cholangiocarcinoma; LC, liver cirrhosis; AJCC, American Joint Committee on Cancer; F, female; M, male; N/A, not applicable; NA, non-alcoholic; PBC, primary biliary cholangitis.

RNA purity and integrity were assessed using an Agilent 2100 BioAnalyzer (Agilent Technologies, Inc.).

RNA-seq method used to identify ctmutRNA. Total RNA-seq was not performed in the present study, but was performed previously (6), and the data generated were used to inform the current study. Total RNA from plasma, EVs and tissues was isolated using the aforementioned method to be used for sequencing. The final loading library concentration was 1.8 pM

and this was verified on an Agilent 2100 BioAnalyzer (Agilent Technologies, Inc.). As detailed in our original study (6), RNA-seq was carried out on either an Illumina NextSeq 500 platform (Illumina, Inc.) using the SMARTer® Stranded Total RNA-Seq Kit v2-Pico Input Mammalian (cat. no. 634411; Takara Bio, Inc.) and a high-output flow cell and reagent kit (v2; cat. no. 20024907; Illumina, Inc.) generating 2x75 bp paired-end reads as reported previously (6), or on the Illumina NovaSeq 6000 platform (Illumina, Inc.) using a S1 Flow Cell

and reagents kit (v1.5; cat. no. 20012864; Illumina, Inc.) with a paired end run, 2x150 bp, generating >50 million reads each. Residual Pico v2 SMART adapters on paired-end fastq files were trimmed. Alignment was performed on the current human reference genome assembly (GRCh38.p13) using the STAR splice aware aligner (v2.7.3) (24). Indel local realignment, base quality recalibration and variant calling (haplotype caller algorithm) were performed using Broad Institutes GATK (v4.1.7.0; <https://gatk.broadinstitute.org/hc/en-us>). Variants with a Phred quality (Q) score of ≥ 20 as well as a depth of five or more were considered true positives. A higher Q score indicated higher confidence and lower probability of an incorrect call. Q20 indicates a 1 in 100 chance of error (99% accuracy). Filtration and annotation of high-risk variants was carried out using SnpEff analysis (v4.3t; <https://pcingola.github.io/SnpEff/#snpeff>) to identify high-risk variants in all sample subsets. Tumor-specific high-risk variants (ctmutRNA) were sorted by concordant detection in tumors and plasma from patients with HCC, frequency/recurrence in samples, and lack of detection in normal liver tissue and in the plasma of patients with LC and normal healthy individuals. Raw unprocessed FASTQ files can be accessed via the publicly available National Center for Biotechnology Information (NCBI) Sequence Read Archive database with accession number PRJNA907745 using the following URL link: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA907745>. Additional information on identified variants was obtained using The Cancer Genome Atlas (TCGA; <portal.gdc.cancer.gov>) and the Human Protein Atlas (HPA; <www.proteinatlas.org>).

Testing the performance of ctmutRNA to identify patients with cancer in this small cohort. Table SII summarizes the top 70 high-impact cancer-associated ctRNA variants identified by RNA-seq exclusively in 16 patients with cancer (HCC/CCA) and not detectable in 15 patients without cancer (patients with LC and healthy individuals). The small discovery cohort consisted of patients with various etiologies and clinical stages. The aim of the present pilot study was to identify a short ctmutRNA panel able to distinguish the majority of patients with liver cancer from patients with LC without cancer. An ordered panel framework was used in which signatures were ranked by the number of positive individuals they identified and signatures that did not contribute additional positive identifications beyond those already captured by higher-ranked signatures were removed. This panel of collections is ordered and organized by decreasing diagnostic performance, meaning one may reduce the size of the panel by keeping the collections at the top of the panel and dropping those at the bottom of the panel. The resulting ordered panel allows investigators to construct sub-panels by retaining signatures from the top of the panel and excluding those at the bottom, thereby trading panel size against performance.

The operating characteristics of multiple ordered panels were visualized using operating characteristic curves, which plot sensitivity (the percentage of positive subjects identified) against the number of variant-signatures included and pseudo-receiver operating characteristic (ROC) curves, which plot sensitivity against non-specificity. These curves were termed pseudo-ROC curves because each sensitivity-non-specificity pair corresponded to a distinct ordered sub-panel rather

than to a continuously varying decision threshold. Panels that achieved higher sensitivity with fewer variant-signatures and whose curves approached the upper-left corner of these plots were considered to exhibit superior performance. This is a novel methodology proposed by TZ and currently under review by The Comprehensive R Archive Network.

DNase digestion and cDNA synthesis. RNA (5 μ g) from selected tissue samples, isolated using a mirVana miRNA Isolation Kit (Ambion; Thermo Fisher Scientific, Inc.), was added to a 1.5-ml microcentrifuge tube. A total of 5 μ l DNase I Enzyme (1 U/ μ l; Invitrogen; Thermo Fisher Scientific, Inc.) was added to each sample with 2 μ l 10X DNase I Buffer, 0.5 μ l RNaseOUT (40 U/ μ l) and RNase-free water to bring the total volume to 20 μ l. Samples were heated at 37°C for 30 min. A total of 2 μ l of 50 μ M random hexamers and 2 μ l of 50 mM EDTA were added to each sample, and samples were then heated at 65°C for 10 min and placed on ice. A reverse transcription (RT) reaction mix consisting of 0.3 μ l of 100 mM dNTPs, 0.2 μ l of RNaseOUT, 1 μ l of SuperScript IV RT Enzyme (200 U/ μ l), 3.5 μ l of 0.1 M DTT and 7.25 μ l of 5X RT Buffer (Thermo Fisher Scientific, Inc.) was added to each sample, and samples were heated at 52.5°C for 15 min, followed by heating at 80°C for 10 min. The volume of each sample was adjusted with RNase-free water to 400 μ l to achieve a final concentration of 12.5 ng/ μ l of cDNA.

Reverse transcription-quantitative PCR (RT-qPCR). A total of 4 μ l of tumor tissue cDNA, prepared as aforementioned, at a concentration of 12.5 ng of total converted RNA (50 ng) per μ l was used as the template of the PCR reaction and added to the well of a 96-well PCR plate. A volume of 8 μ l of 2 μ M forward and reverse primer pairs that flanked the target exon-exon junction, whose sequences can be found in Table SIII, was added to the well. Lastly, 12 μ l of 2X PowerUP SYBR PCR master mix (Applied Biosystems; Thermo Fisher Scientific, Inc.) was added to the well, with a final volume of 24 μ l. The plate was sealed with an optically clear film and centrifuged at 500 x g for 2 min at room temperature. PCR was performed on a LightCycler 480 instrument (Bio-Rad Laboratories, Inc.) under the following heating conditions: 95°C for 10 min (enzyme activation), followed by 45 cycles of 95°C for 10 sec (denaturation), 60°C for 30 sec (annealing) and 72°C for 5 sec (extension). A melt curve was generated after the 45th cycle for melt curve genotyping analysis. The $2^{-\Delta\Delta C_q}$ method was used to calculate relative expression (25). β -actin was used as an endogenous reference gene whose primer sequences can be found in Table SIII.

Traditional semi-qPCR. An additional endpoint PCR reaction was performed using the RT-qPCR-amplified tumor and healthy liver cDNA. A volume of 4 μ l of the first PCR product, previously amplified by RT-qPCR as aforementioned, was added to a 200- μ l 8-tube strip tube. A volume of 8 μ l of the same 2 μ M forward and reverse primer pairs that flanked the target exon-exon junction was added to the well. Primer sequences can be found in Table SIII. Lastly, 12 μ l of 2X PCR Master Mix with no fluorescent dye (Thermo Fisher Scientific, Inc.) was added to the well, with a final volume of 24 μ l. PCR was performed on a T100 Thermocycler (Bio-Rad

Laboratories, Inc.) with the following heating conditions: 98°C for 3 min (enzyme activation), followed by 35 cycles of 98°C for 15 sec (denaturation), 60°C for 30 sec (annealing) and 72°C for 45 sec (extension). Subsequently, 2–2.5% agarose gels were prepared using 1X triethylamine and pre-stained with ethidium bromide. A volume of 10–30 μ l of the endpoint PCR product was mixed with 2–6 μ l of 6X loading dye (New England BioLabs, Inc.) and loaded into the well. A volume of 10 or 5 μ l of 50-bp DNA ladder was added to a well adjacent to the samples. Gel electrophoresis was then performed at 100 V for 1 h. The gel was subsequently imaged on a ChemiDoc Touch (Bio-Rad Laboratories, Inc.) or iBright 1500 (Thermo Fisher Scientific, Inc.) and the sizes and densities of the PCR products were compared using Image Lab (v6.0.1; Bio-Rad Laboratories, Inc.). β -actin was used as a reference gene as a loading control and for normalization.

Gel extraction, purification and sequencing. After imaging, cDNA bands of interest were identified. The gel was placed on a UV lightbox and the bands were cut out and placed in a 1.5-ml microcentrifuge tube. Using the Monarch DNA Gel Extraction kit (New England BioLabs, Inc.), the DNA in the band was isolated and eluted according to the manufacturer's protocol. The isolated DNA was sent to Azenta Life Sciences with the appropriate forward primer (Table SIV) for Sanger sequencing. Results were delivered electronically and the returned sequences were compared with the reference sequence using the NCBI Basic Local Alignment Search Tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

SNP variants and digital PCR (dPCR). A volume of 4 μ l of tumor tissue cDNA, generated as aforementioned, at a concentration of 12.5 ng of total converted RNA per μ l (50 ng) was used as the template of the PCR reaction and added to the well of a 96-well PCR plate. Subsequently, 1.25 μ l of 20X TaqMan Assay mix (Applied Biosystems; Thermo Fisher Scientific, Inc.) containing primers and the two allele-specific probes, tagged with fluorescein amidite (FAM) and chloro-phenyl-dichloro-carboxy-fluorescein (VIC) labels, was added to the well with the template. Probe assay information and sequences can be found in Table SIV. A volume of 12 μ l of 2X TaqPath ProAmp PCR Master Mix (Thermo Fisher Scientific, Inc.) was added to the well. Finally, 6.75 μ l of nuclease-free water was added to the well to bring the total reaction volume up to 24 μ l per well. The plate was sealed with an optically clear film and the plate was centrifuged at 500 x g for 2 min at room temperature. PCR was performed on a LightCycler 480 instrument (Roche Diagnostics) under the following heating conditions: 95°C for 10 min (enzyme activation), followed by 40 cycles of 95°C for 10 sec (denaturation), 60°C for 1 min (annealing/extension) and a brief 1 sec pause at 72°C. Endpoint genotyping analysis of the FAM and VIC fluorescence was performed to determine which alleles were present using the Light Cycler 480 operating and analysis software (v1.5; Roche Diagnostics).

For the digital PCR, a volume of 4 μ l of tumor tissue cDNA at a concentration of 12.5 ng of total converted RNA per μ l (50 ng) was used as the template of the dPCR reaction. A total of 0.5 μ l 20X assay mix (Table SIV; Thermo Fisher Scientific, Inc.) was added to the sample along with 2 μ l of 5X

dPCR Master Mix (Thermo Fisher Scientific, Inc.) and 3.5 μ l of PCR-grade water. The total assay solution was mixed and 9 μ l of the solution was added to the dPCR well with 15 μ l of Isolation Buffer (Thermo Fisher Scientific, Inc.). Subsequently, dPCR was performed on the Quant Studio dPCR machine (Thermo Fisher Scientific, Inc.) using the manufacturer's recommended reaction conditions for 40 cycles. Analysis was performed using the QuantStudio Absolute Q analysis software (v6.3.5; Applied Biosystems; Thermo Fisher Scientific, Inc.). The mean partition count for the dPCR reactions was 20,409 $\pm$ 112 and the minimum copy number detected was 1 copy per 50 ng of input cDNA. Positive calls were determined using the automatic thresholding algorithm of the QuantStudio Absolute Q software.

TaqMan probe-based assays for splicing variants. cDNA from tissue used for these assays was prepared as aforementioned. Probes and primers corresponding to indel variants were designed using tools from Thermo Fisher Scientific, Inc. The sequence information for all splicing TaqMan probes and primers (Thermo Fisher Scientific, Inc.) is shown in Table SV. A total of 4 μ l of cDNA, 1 μ l of 20X TaqMan Assay mix (Thermo Fisher Scientific, Inc.), 10 μ l of 2X TaqPath ProAmp Master Mix (Thermo Fisher Scientific, Inc.) and 5 μ l of PCR-grade water were mixed for the reaction. PCR was performed on a LightCycler 480 (Roche Diagnostics) under the following conditions: 95°C for 10 min, followed by 45 cycles of 95°C for 15 sec and 60°C for 1 min. GAPDH (Table SV) was used as an endogenous control gene for the relative expression calculations using the $2^{-\Delta\Delta C_q}$ method.

For the dPCR using these assays, plasma cDNA was prepared as aforementioned. A total of 4 μ l of cDNA, 0.5 μ l of 20X TaqMan assay mix (Table SV; Thermo Fisher Scientific, Inc.), 2 μ l of 5X dPCR Universal DNA Master Mix (Thermo Fisher Scientific, Inc.) and 3.5 μ l of PCR-grade water were combined to make a 10 μ l reaction mix. A total of 9 μ l of the reaction mix was added to the dPCR well. Isolation Buffer (15 μ l; Thermo Fisher Scientific, Inc.) was then added on top of the well. dPCR was performed on the Quant Studio dPCR machine (Thermo Fisher Scientific, Inc.) using the manufacturer's recommended reaction conditions for 40 cycles. Analysis was performed using the QuantStudio Absolute Q analysis software (v6.3.5; Applied Biosystems; Thermo Fisher Scientific, Inc.).

Statistical analysis. Beyond the variant selection process using the aforementioned ordered panel framework, appropriate statistical tools were used for variant validation qPCR and dPCR assays. Data for RT-qPCR assays are presented as the mean $\log_2$ fold-change in expression $\pm$ standard deviation. Data for dPCR assays are presented as individual copy number values with the minimum, first quartile, median, third quartile and maximum shown as a box plot on a $\log_2$ scale. For single-probe qPCR assays, Welch's ANOVA was used to assess the overall difference of the mean $\log_2$ fold change in expression of each target gene in the tumor tissue samples (HCC and CCA tumors; n=6) and the healthy liver tissue sample (n=1). This was followed by Dunnett's T3 multiple comparisons test to compare each tumor sample back to the healthy liver tissue and control for multiple comparisons. Using all of the

Dunnett's T3-adjusted P-values, a Holm-Bonferroni multiple comparisons correction was applied to all assays. Both adjusted P-values for Dunnett's T3 tests and the Holm-Bonferroni test are reported. Hedge's g score was used to calculate the effect size for each assay. Statistical analysis and graphing were performed using GraphPad Prism (v11.0.0; Dotmatics). For the dPCR single-probe assays using plasma samples from patients with HCC and patients with LC without HCC, aggregated ctRNA copy numbers of HCC samples (n=10) and LC samples (n=6) were compared using a Mann-Whitney test to determine statistically significant differences between the two groups for each of the eight assay probe targets. Rank biserial correlation was used to calculate the effect size for each assay. A Holm-Bonferroni multiple comparisons correction was used to determine adjusted P-values for the panel of assays. Mann-Whitney tests, Welch's ANOVAs and Dunnett's T3 multiple test corrections were performed using GraphPad Prism (v11.0.0), and the Holm-Bonferroni corrections were performed using R Statistical Software (v4.3.1; R Core Team). All tests were run in triplicate. Graphs were created using GraphPad Prism (v11.0.0). A mean coefficient of variation (CV) of $5.50 \pm 3.90\%$ was observed in the TaqMan assays with a minimum CV of $2.45 \pm 2.12\%$ and a maximum CV of $12.02 \pm 5.83\%$. $P < 0.05$ was considered to indicate a statistically significant difference.

Results

Diagnostic performance of HCC-specific ctRNA variants identified by RNA-seq. In the present pilot study, a large independent validation cohort was not available to investigate the diagnostic performance of ctmutRNA variants identified in the discovery cohort. Since the original discovery cohort consisted of patients with different etiologies and clinical stages, the present study aimed to determine if a short ctmutRNA panel could distinguish the majority of the patients with cancer. Table SII summarizes the top 70 high-impact cancer-associated ctRNA variants identified by RNA-seq exclusively in 16 patients with cancer (HCC/CCA) and not detectable in 15 non-cancer individuals (LC and healthy). These 70 ctRNA variants were sorted into 54 variant signatures (Fig. 1) enabling identification of identical sets of patients with cancer, resulting in groups of variants that identified the same patients with cancer. Tables SVI and SVII show the 'variants to collections' distribution to explain how the 70 top ctRNA variants were reduced to 54 collections before further analysis. Using the ordered panel methodology in this small feasibility study, a panel of six ctRNA variants was identified with a diagnostic sensitivity of 87.5% and a specificity of 100% (Fig. 1). The sensitivity (87.5%; 14/16 patients with cancer) corresponded to a 95% exact CI of 61.7-98.4%. The specificity (100%; 0/15 non-cancer subjects) corresponded to a 95% exact CI of 78.2-100.0%.

The operating characteristics of multiple ordered panels were visualized using a pseudo-operating characteristic curve, which plots sensitivity (the percentage of positive individuals identified) against the number of variant-signatures included, and another pseudo-ROC curve, which plots sensitivity against non-specificity (Fig. 1). These curves were referred to as pseudo-ROC curves because each sensitivity non-specificity

pair corresponded to a distinct ordered sub-panel rather than to a continuously varying decision threshold. Panels that achieved higher sensitivity with fewer variant signatures and whose curves approached the upper-left corner of these plots were considered to have superior performance.

Characterization of HCC-specific circulating SNPs using tumor tissues. Having identified several high-impact HCC-specific ctRNA variants in a previous cohort of patients with HCC and CCA by RNA-seq, it was decided to investigate if these variants could be confirmed using simple assays in tumor tissues from patients with liver cancer (Table I). Detection of high-impact SNPs in apolipoprotein B (APOB) and transferrin (TF) in HCC tumor tissues from RNA-seq read piles is shown in Fig. S1. The TF SNP was exclusively expressed in all five tumors with no wild-type (WT) allele detected. The APOB SNP variant allele was expressed in all five tumors; however, APOB's variant allele was only exclusively expressed in two tumors, with the other three showing expression of the WT allele. The investigation began using a set of highly recurrent SNPs (Table SII) detected exclusively in the tumors, plasma and exosome fractions from patients with HCC and CCA.

For the investigation of SNPs, duplexed competing hydrolysis probe assays (Thermo Fisher Scientific, Inc.), also referred to as 5' nuclease assays, were used to perform endpoint genotyping on the tumor RNA. HCC-specific SNPs in APOB and TF were examined as these occurred in 100 and 83% of HCC tumor samples analyzed by RNA-seq, respectively. For assay standardization, variable amounts of synthetic cDNA templates corresponding to variant and WT alleles of APOB SNP rs1042034 were tested using SNP TaqMan assays, and consistent results reflecting allelic differentiation were observed with all template amounts (Fig. S2). Although there is no known association between ctRNA variants identified by RNA-seq and corresponding transcript expression, The TCGA database indicated tissue-specific dysregulation of variant-associated transcripts in HCC datasets, underscoring the liver cancer specificity of transcripts harboring these high-impact variants (Fig. 2A). APOB expression has been reported as a potential autonomous predictor of survival in patients with HCC (26). Schematic representations of the SNPs in the last exon of APOB and exon 12 of TF are shown in Fig. 2B. SNP TaqMan assays were carried out using RNA from tumor tissues. When investigating each SNP, two probes corresponding to WT and variant alleles competed to bind the target region and consequently, nuclease activity associated with RNAPol led to the release of the fluorescence tag to reveal the identity of the target allele. To ensure that the assay was working correctly, synthetic DNA containing either the WT or variant sequences was used as a positive control for the WT and variant signal in the TaqMan assay (Fig. 2C). The results demonstrated that the synthetic WT and variant sequence signals were observed close to the y- and x-axis, respectively. Tumor samples tested to examine the SNPs in APOB and TF showed alignment with the variant allele frequencies (VAFs) captured by RNA-seq (6). For example, for APOB, HCC006T exhibited a VAF of 388 variant reads/390 total reads at that locus, which suggested that this tumor was predominantly harboring the variant APOB allele. Conversely, HCC001T exhibited an APOB SNP VAF of 192/635, which suggested

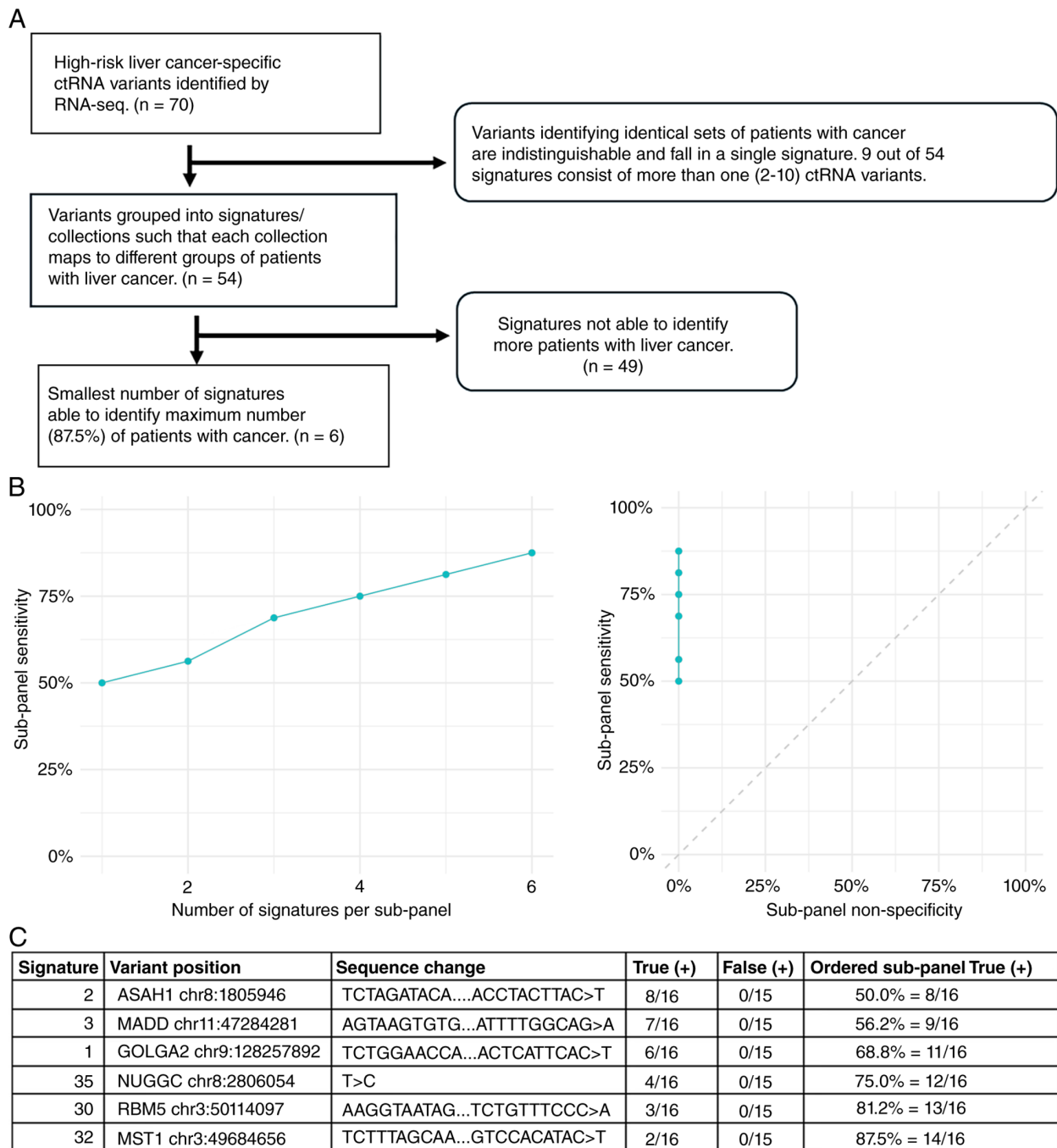


Figure 1. Operating characteristics of a ctmutRNA panel for the identification of subjects with liver cancer with high precision. (A) The flow chart shows the process of reducing 70 ctmutRNA variants identified by RNA-seq to a short panel of six variants able to distinguish 87.5% of patients with liver cancer (n=16). In this scheme, signatures that identified no cancer subjects or were associated with one or more negative control subjects (liver cirrhosis without cancer) were excluded, and the remaining signatures were ranked by the number of positive subjects they identified. Generally, each individual ctRNA variant plotted on the curve represents multiple signatures which identify the same individual patients with cancer (Tables SVI and SVII). To avoid redundancy, signatures that did not contribute additional positive identifications beyond those already captured by higher-ranked signatures were removed. (B) The operating characteristics of multiple ordered panels were visualized using an operating characteristic curve, which plots sensitivity (percentage of positive subjects identified) against the number of variant signatures included, and a pseudo-ROC curve, which plots sensitivity against non-specificity. This curve is referred to as a pseudo-ROC curve because each sensitivity-non-specificity pair corresponds to a distinct ordered sub-panel rather than to a continuously varying decision threshold. Panels that achieve higher sensitivity with fewer variant signatures, and whose curve approaches the upper-left corner of these plots, are considered to have superior performance. (C) The variant annotation and number of true positive patients with HCC captured of the six ctmutRNA signatures in patients with cancer and non-cancer subjects are shown. The ordered-panel framework and methodology are implemented in the R (v4.3.1; R Core Team) package ordPanel, which is currently under review by The Comprehensive R Archive Network. ASAH1, N-acylsphingosine amidohydrolase 1; chr, chromosome; ctmutRNA, circulating mutated RNA; GOLGA2, golgin A2; MADD, MAP kinase-activating death domain; MST1, macrophage stimulating 1; RBM5, RNA binding motif protein 5; RNA-seq, RNA-sequencing; ROC, receiver operating characteristic.

that both WT and variant alleles were detectable in this tumor (data not shown). Investigation of the read piles from the

original RNA-seq data in five tumor tissues highlighted these variant SNPs corresponding to TF and APOB (Fig. S1).

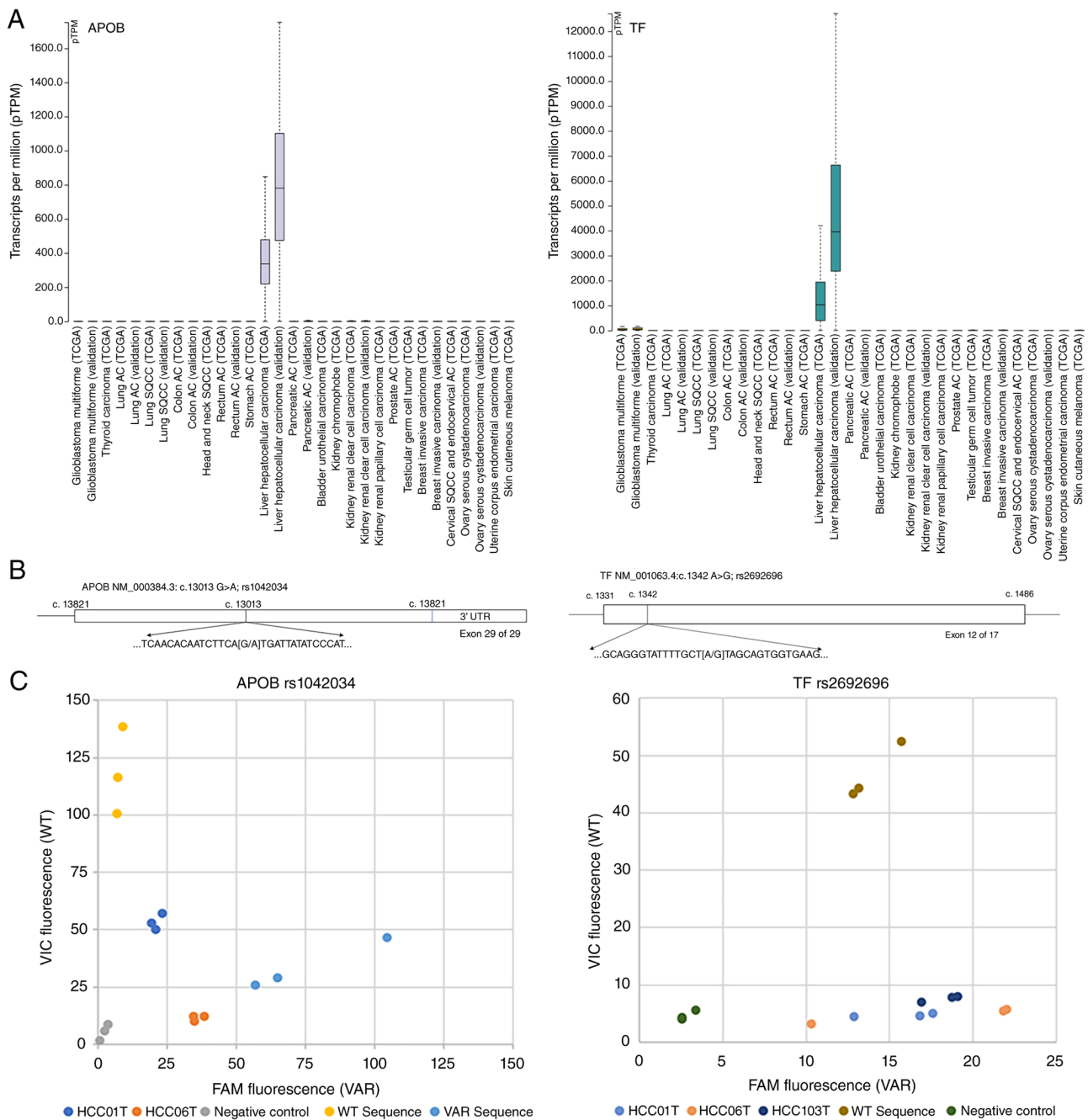


Figure 2. SNP genotyping assay validation using HCC tumor-derived RNA. (A) High-risk, HCC-specific ctRNA SNPs identified in both plasma and tumor tissues of patients with HCC/cholangiocarcinoma by RNA-sequencing corresponded to APOB and TF. Tissue-wide expression analysis of APOB and TF in tumors was performed using TCGA. High expression levels of both APOB and TF were observed in the HCC cohort (362 tumors) and HCC validation cohort (231 tumors), highlighting exclusive liver-specific dysregulation. (B) Schematic representations of the positions of the APOB SNP rs1042034 and the TF SNP rs2692696 based on the Matched Annotation from National Center for Biotechnology Information and European Molecular Biology Laboratory-European Bioinformatics Institute Select APOB and TF mRNA transcripts. The context sequence of the APOB SNP shows the transition from WT (G) to the variant allele (A), and the context sequence of the TF SNP shows the transition from WT (A) to the variant allele (G). (C) Graphical demonstration of discrimination and detection of the WT and VAR alleles using SNP TaqMan assays based on dual-labeled probes. Among the two competitive probes, depending on complementarity, only one will bind and release fluorescence, revealing the identity of the allele in the reaction. (Left) Reverse transcription-quantitative PCR reactions were all performed in triplicate with 50 ng of cDNA for the tumors and 10 pg for the synthetic controls. The allelic discrimination plot for the APOB SNP shows the results for two tumors, HCC001T and HCC006T, with synthetic DNA controls for the WT and VAR allele sequences. (Right) Allelic discrimination plot of TF SNP in three tumors (HCC001T, HCC006T and HCC103T), with a synthetic WT template control. AC, adenocarcinoma; APOB, apolipoprotein B; ctRNA, circulating RNA; FAM, fluorescein amidite; HCC, hepatocellular carcinoma; pTPM, protein-coding transcripts per million; SNP, single nucleotide polymorphism; SQCC, squamous cell carcinoma; TCGA, The Cancer Genome Atlas; TF, transferrin; UTR, untranslated region; VAR, variant; VIC, 2'-chloro-7-phenyl-1,4-dichloro-6-carboxy-fluorescein; WT, wild-type.

With good confidence in the power of this assay, the validation panel was expanded to other cancer-specific SNPs of interest (Table SII). The variant annotations and locations

in the coding regions of liver-specific genes glycosyl-transferase-like domain containing 1 (GTDC1), NUGGC, kallikrein B1 (KLKB1), CD151, phosphoenolpyruvate

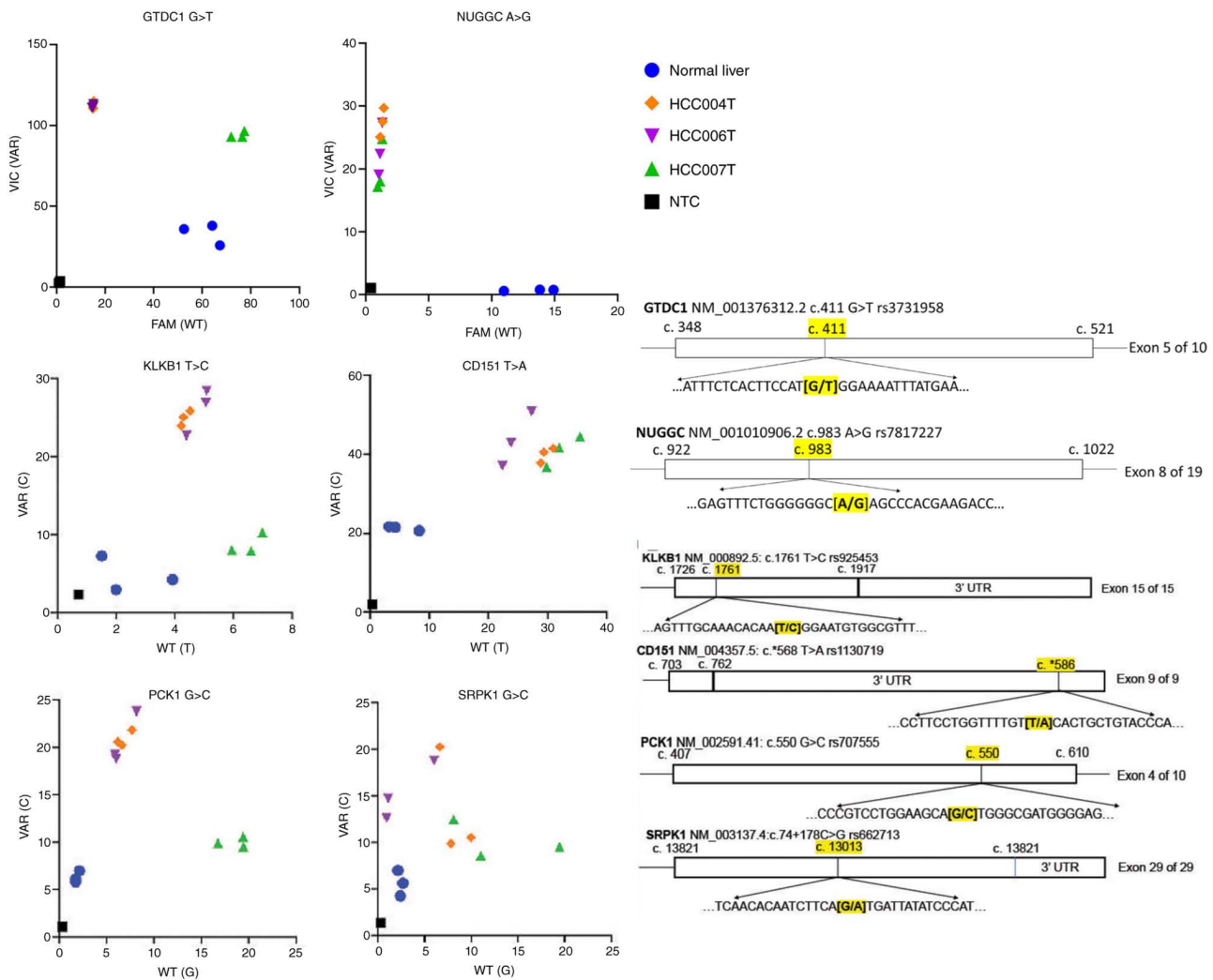


Figure 3. Characterization and validation of HCC-specific SNPs in tissues using endpoint genotyping via RT-PCR. Allelic discrimination plots for six highly recurrent HCC-specific SNPs in three tumors (HCC004T, HCC006T and HCC007T), and in a non-cancerous liver tissue as a control. RT-PCR reactions were all performed in triplicate using dual allele-specific probe TaqMan assays. A total of 50 ng of total RNA converted to cDNA was used as a template. Y-axes represent the fluorescence of the VAR allele and x-axes represent the fluorescence of the WT allele. The right panel shows schematic representations of the SNP positions corresponding to GTDC1 rs3731958, NUGGC rs7817227, KLKB1 rs925453, CD151 rs1130719, PCK1 rs707555 and SRPK1 rs662713 based on the Matched Annotation from National Center for Biotechnology Information and European Molecular Biology Laboratory-European Bioinformatics Institute Select transcripts. FAM, fluorescein amidite; GTDC1, glycosyltransferase-like domain containing 1; HCC, hepatocellular carcinoma; KLKB1, kallikrein B1; NTC, no-template control; PCK1, phosphoenolpyruvate carboxykinase 1; RT-PCR, reverse transcription-PCR; SNP, single nucleotide polymorphism; SRPK1, serine arginine protein kinase 1; UTR, untranslated region; VAR, variant; VIC, 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxy-fluorescein; WT, wild-type.

carboxykinase 1 (PCK1) and serine arginine protein kinase 1 (SRPK1) are shown in Fig. 3. GTDC1 is known to be a potential prognostic gene in liver cancer based on the HPA and a diagnostic gene for colon cancer (27). High NUGGC expression is associated with HCC, prognostic for lung cancer and associated with different patient survival outcomes in various types of cancer as shown in an analysis using the HPA. Serum KLKB1 is known to be a potential prognostic biomarker in HCC (28). CD151, a tetraspanin receptor, is known to drive cancer progression by interacting with integrins and EGFR in non-small cell lung cancer (29) and promotes the invasiveness and angiogenesis of HCC (30). Phosphoenol carboxy kinase is known to serve a role in metabolic reprogramming and immune microenvironment remodeling and has been proposed as a therapeutic target in cancer (31,32). SRPK1, a splicing kinase, has been reported to be a prognostic factor and potential therapeutic

target in various types of cancer, and its upregulation is associated with higher tumor staging, grading and shorter survival (33). A normal liver tissue and three tumor tissues were used to characterize the high-impact HCC-specific SNPs corresponding to these genes using the same TaqMan SNP allelic discrimination assays (Fig. 3). Tumors showed differences in variant allelic frequencies, but all six high-impact variants displayed higher VAR values in the majority of tumors when compared with the normal liver VAR values, validating the results of RNA-seq of ctRNAs in patients with HCC. For KLKB1, the variant allele was detected in HCC004T and HCC006T, but not in HCC007T. These data suggested that even if the expression of a liver-specific transcript may not serve as a biomarker for HCC detection, the allelic discrimination assay employed in the present study was a superior tool to track these potential biomarker SNPs.

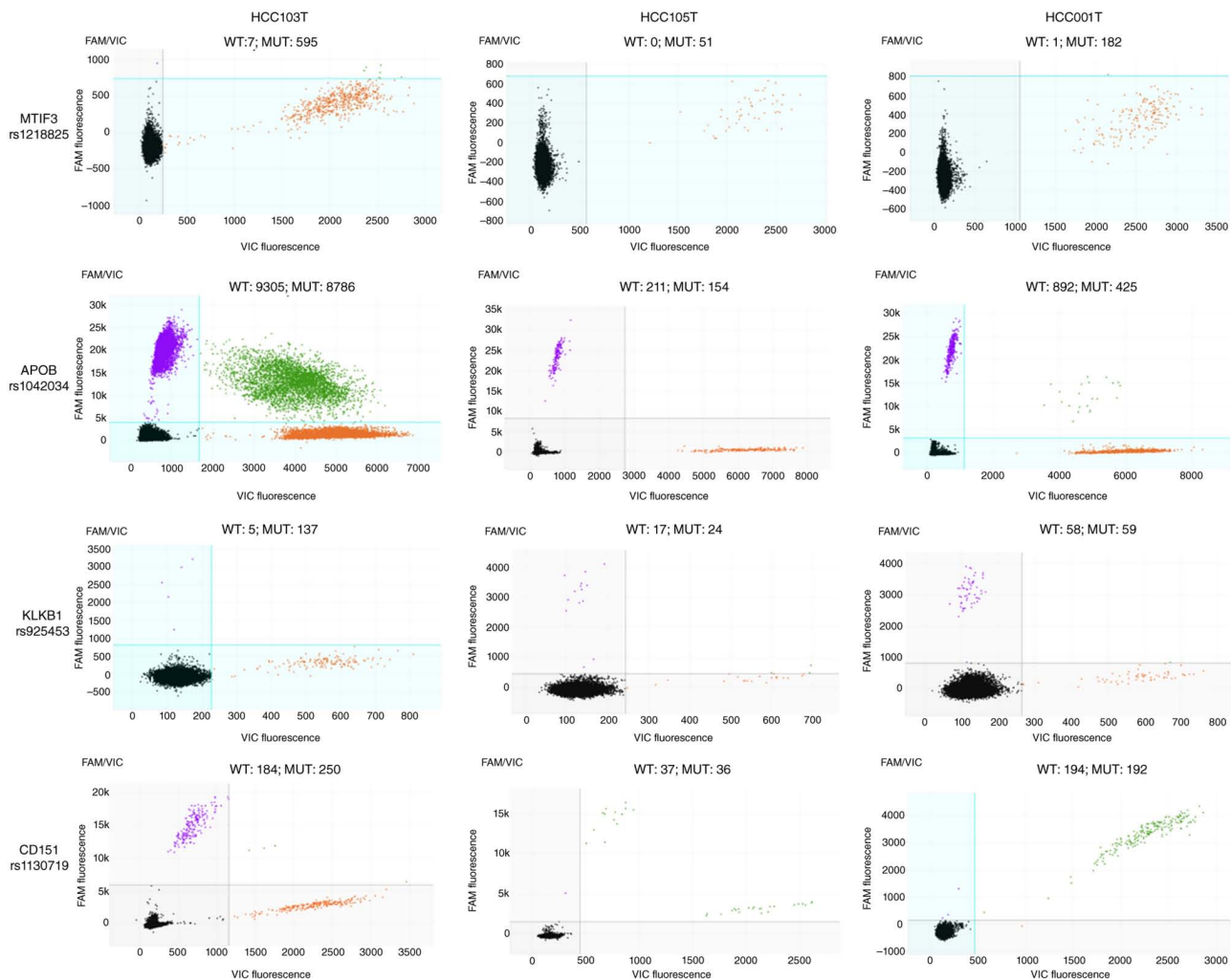


Figure 4. Investigation of HCC-specific high-risk SNPs in HCC tumor tissues using digital PCR enables quantitative analysis of variant and WT alleles. Two-dimensional scatter plots of four HCC-specific SNPs (MTIF3 rs1218825, APOB rs1042034, KLKB1 rs925453 and CD151 rs1130719) across three tumors (HCC103T, HCC105T and HCC001T). Each PCR reaction was split into >20,000 mini chambers, and WT and variant allele numbers were calculated based on the number of wells positive for each labeled probe in each sample. Black dots at the bottom left of each plot represent wells with no reaction. Orange dots represent wells only positive for VIC, purple dots represent wells only positive for FAM, and green dots represent wells positive for both VIC and FAM. In the MTIF3, KLKB1 and CD151 plots, the FAM probe represents the WT allele and the VIC probe represents the MUT allele. In the APOB plots, the FAM probe represents the MUT allele and the VIC probe represents the WT allele. The counts displayed are the MUT and WT allele copy numbers in the total sample loaded (50 ng total RNA). Digital PCR reactions were all performed in triplicate using dual-probe TaqMan assays and 50 ng of total RNA converted to cDNA as a template. APOB, apolipoprotein B; FAM, fluorescein amidite; HCC, hepatocellular carcinoma; KLKB1, kallikrein B1; MTIF3, mitochondrial translational initiation factor 3; MUT, mutant; SNP, single nucleotide polymorphism; VIC, 2'-chloro-7-phenyl-1,4-dichloro-6-carboxy-fluorescein; WT, wild-type.

Validation of SNP variants using digital qPCR. Given the need for non-invasive detection, high sensitivity and quantification of these variant transcripts, a dPCR platform was used to execute the same duplexed allele-specific assay (Fig. 4). The qualitative genotyping assay was turned into a quantitative assay using the dPCR platform. HCC-specific high-risk SNP variants were investigated in three tumor tissues. In addition to the aforementioned investigation of the unique SNPs in APOB, KLKB1 and CD151, a high impact SNP in mitochondrial translational initiation factor 3 (MTIF3) was included in this experiment. Loss of function of MTIF3 may disrupt the fidelity of mitochondrial translation, leading to production of unstable proteins impairing mitochondrial function (34). The MTIF3 specific variant SNP was predominant in all three tumors, while both WT and variant alleles corresponding to APOB, KLKB1 and CD151 were observed. In addition to sensitive detection of variants, this assay allowed

the determination of the exact copy numbers of variant and WT alleles corresponding to these SNPs, as shown in Fig. 4. The copy number counts of APOB in HCC001T were closely associated with the VAF in RNA-seq data, confirming that the use of these TaqMan assays gave accurate results when transferred to a more sensitive platform and that the APOB SNP found in the RNA-seq data was almost certainly a true variant call.

A considerable concern regarding the use of ctDNA variants as biomarkers in blood has been the limit of detection of low-frequency alleles. RNA does not conform to the two copy limit and is produced in larger numbers (5,9). The use of dPCR for the allelic discrimination assays demonstrated a large dynamic range, capturing highly expressed variants, such as variants in APOB, or variants with much lower RNA expression levels, such as variants in KLKB1. The allelic discrimination assay in tumor tissues using dPCR was able

to confirm the occurrence of all these high-impact variants originally detected in ctRNA from patients with HCC.

Validation of splice site indel variants in tumor tissues. The detection of several high-risk circulating splicing variants in ctRNA falling into the category of indel variants has been reported in the present study. The RNA-seq analysis identified a high number of splice site variants with high recurrence in patients with HCC/CCA, which warranted further investigation (Table SII). It can be hypothesized that these structural variants would be equally impactful and pathogenic in terms of functionality of the affected protein in the cancer cell. To validate these specific splice-site mutations, primer pairs were designed to amplify RNA stretches around the exon-intron-exon junctions in endpoint PCR. Due to the higher abundance of RNA isolated from the tumor tissues compared with plasma, the aim was to validate the RNA-seq findings from plasma samples to ensure that the variant calling using the bioinformatics pipeline was accurate. Semi-qPCR assays using three tumor tissues and a normal liver tissue were carried out, corresponding to 15 indel sites. After PCR amplification, these amplified products were separated by agarose gel electrophoresis for semi-quantitation of amplicons (Fig. 5A). Exon-intron-exon junctions where the variants were located on the corresponding mRNA are shown in schematic diagrams in Fig. 5B. Striking differences in amplicon size between the normal liver tissue and the three tumor tissues were demonstrated. Splice site-specific variants in protein phosphatase 1 regulatory subunit 12C (PPP1R12C), vacuolar protein sorting 13 homolog A (VPS13A), protein tyrosine phosphatase receptor type J (PTPRJ), transfer RNA (tRNA) methyltransferase 1 (TRMT1), ubiquitin protein ligase E3 component n-recogin 4 (UBR4), mitogen-activated protein kinase kinase kinase 2 (MAP4K2), solute carrier family 25 member 17 (SLC25A17), macrophage stimulating 1 (MST1), uroporphyrinogen decarboxylase (UROD), pumilio RNA binding family member 2 (PUM2), agrin (AGRN), catenin α like 1 (CTNNAL1), ubiquitin conjugating enzyme E2 Q1 (UBE2Q1) and mitogen-activated protein kinase 2 were detected using this approach, affirming the authenticity of ctRNA variant calls based on RNA-seq data. The PPP1R12C gene encodes a subunit of myosin phosphatase and is a prognostic marker in liver cancer according to the Human Protein Atlas (<https://www.proteinatlas.org/>). The mutations and dysregulated expression of VPS13A in tumors affect prognosis and serve as biomarkers for screening and diagnosis (35). PTPRJ, a receptor protein tyrosine phosphatase, drives VEGF-dependent tumor progression and represents a therapeutic target (36). TRMT1 enzymatically modifies tRNA, and thus, regulates tRNA structure, function and stability (37). UBR4, an E3 ubiquitin ligase, is known to enhance tumor cell proliferation and survival by regulating DNA mismatch repair and the stress response (38). MAP4K2, a mitogen-activated serine/threonine kinase, is upregulated in head and neck cancer and associated with the Hippo pathway affecting autophagy of cellular components under energy stress (39). Another head and neck cancer associated gene, SLC25A17, a mitochondrial carrier, serves a prognostic role in cancer (40). MST1, a Hippo pathway-related tumor suppressor, is known to be involved in multiple myeloma pathogenesis (41). When

cleaved to its 36-kDa form, MST1 suppresses tumor growth by amplifying apoptotic signals and 76% of HCC tumors exhibit the absence of this splicing process to create the 36-kDa protein (42). UROD, a heme biosynthesis enzyme, serves as a prognostic factor in cervical cancer and radiation therapy (43). PUM2 is an oncogenic RNA-binding protein regulating HCC cell proliferation and apoptosis via binding to another tumor suppressor, BTG anti-proliferation factor 3 (44), and has been reported to drive colorectal cancer growth by suppressing p21 (45).

AGRN is a matricellular glycoprotein involved in extracellular matrix signaling, has been reported to promote thyroid, colon and lung cancer, and is known as a therapeutic target (46). CTNNAL1 is known as a prognostic marker in HCC and gastric cancer (HPA). UBE2Q1, a ubiquitin conjugating enzyme, has been identified as a potential prognostic marker for ovarian and digestive cancer (47). As before, normal liver tissue exhibited a single predominant WT product, except for VPS13A and PUM2, for which little to no expression was observed in normal liver tissue. However, using the same splice site-specific primers, tumor tissues either showed amplicons of a different size or altered expression of the WT isoform. While a number of the tumor samples exhibited a similar tumor-specific isoform at the different splice sites, a variety of isoforms of TRMT1, UBR4 and PUM2 were observed across the tumor samples. Unlike the other examples, for the GET3 splice junction site, a single, consistently sized band was observed across normal liver and tumor tissues, showing no splicing differences at this site. In some cases, some ctmutRNA-associated splice site isoforms were observed to be downregulated in a tumor tissue compared with the normal liver. Using both RT-semi-qPCR and RT-qPCR, a loss of expression in several genes, including FASN, POLR2A, TDP2, MADD, ARAF and HYI, was observed (Fig. S3). It is useful to remember that all these high-risk HCC-specific variants were originally identified in the circulation of patients with HCC using RNA-seq and were not found in patients with LC without cancer, highlighting their potential utility as noninvasive biomarkers for HCC surveillance.

Since the previous RNA-seq data showed that a number of HCC-specific variants were also detectable in patients with CCA (Table SII), ctmutRNA variants in CCA tissue were investigated (Fig. 6). Indel variants corresponding to RAF1, coat protein complex I subunit α (COPA), N-acylsphingosine amidohydrolase (ASAH) and serine/threonine-protein kinase A-Raf (ARAF) were investigated. As with the previous set of experiments, the primer pairs were designed to amplify RNA stretches around the exon-intron-exon junctions in endpoint PCR. Semi-qPCR assays using a tumor tissue and a normal liver tissue were carried out targeting the ctRNA splice junction sites and the resulting amplicons were resolved on agarose gels. Fig. 6A shows distinct amplicon products in lanes 2 (normal liver) and 3 (CCA tissue). Lane 1 corresponds to PCR reactions with no template. The panel on the right portrays the variant annotations and schematic representations along exon-intron junctions. As seen in HCC tissues in the aforementioned experiments, the CCA tumor showed differentially sized amplicons corresponding to RAF1, COPA, ASAH and ARAF. Indel variants corresponding to RNA binding motif protein 5 (RBM5) and golgin A2 (GOLGA2) were similarly

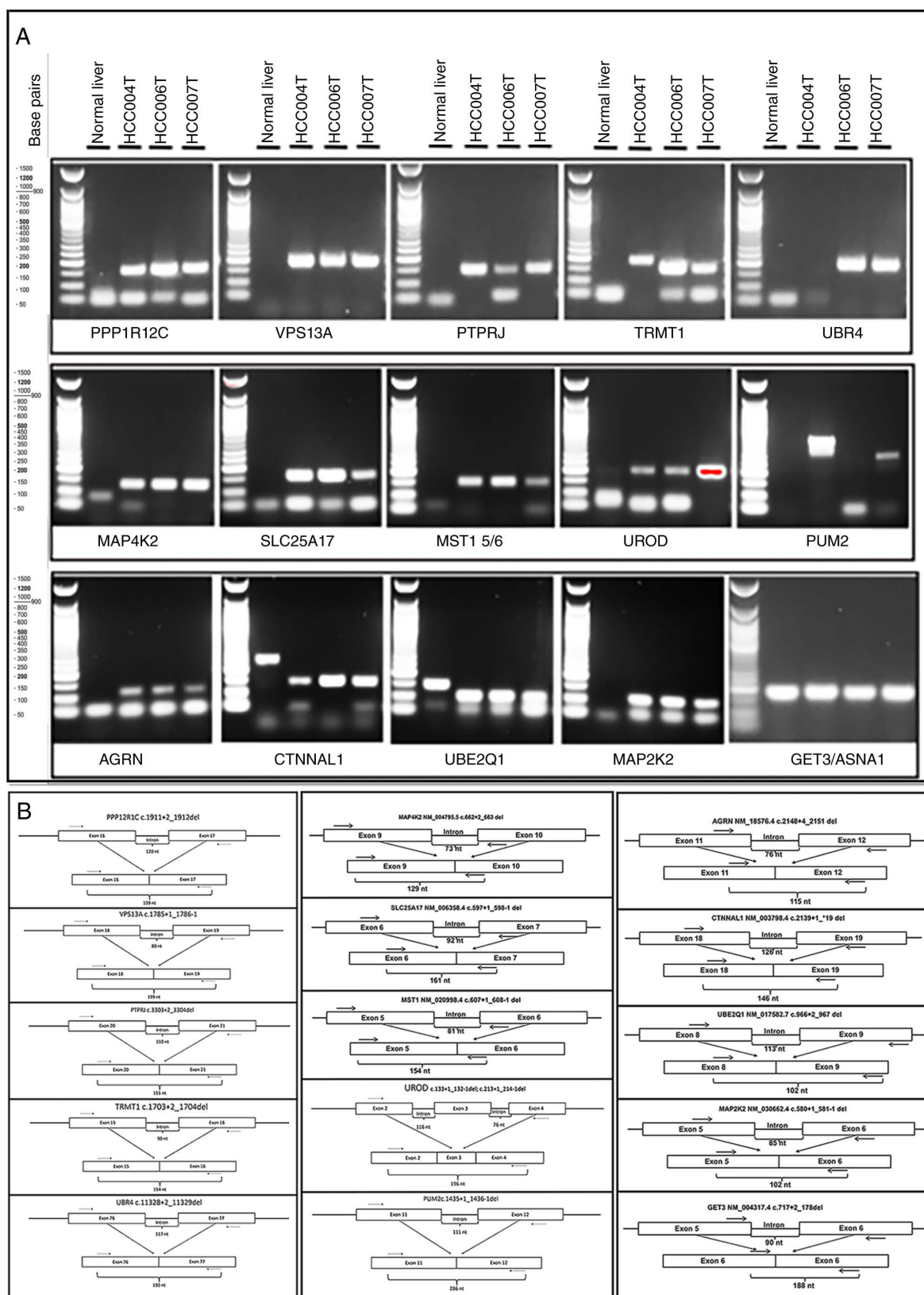


Figure 5. ctRNA indel variants generate tumor-specific, transcribed RNA isoforms visualized using semi-qPCR and agarose gel electrophoresis. (A) Agarose gel images of tumor-specific ctRNA amplicons generated by indel-specific semi-qPCR corresponding to several ctRNA variants identified by RNA-sequencing. Primers encompassing the indel variant sites along exon-intron junctions were used to amplify the cDNA templates from tumor tissues to evaluate the size of the amplicons. A 50-bp ladder was used to determine the size. Appearance of a red streak is caused by pixel saturation (overexposure) when balancing exposure times to visualize both high- and low-abundance bands. This often occurs while optimizing imaging settings to capture finer details and visualize any low-abundance bands across the gel. (B) Schematic representations of the exon-intron junctions and location of the indels around the exon-intron-exon junction. Amplicon sizes expected after normal splicing based on the canonical Matched Annotation from National Center for Biotechnology Information and European Molecular Biology Laboratory-European Bioinformatics Institute Select transcripts are shown. The left and right arrows indicate the relative positions of the forward and reverse primers used for semi-qPCR. ctRNA, circulating RNA; HCC, hepatocellular carcinoma; indel, insertion and deletion; semi-qPCR, semi-quantitative PCR.

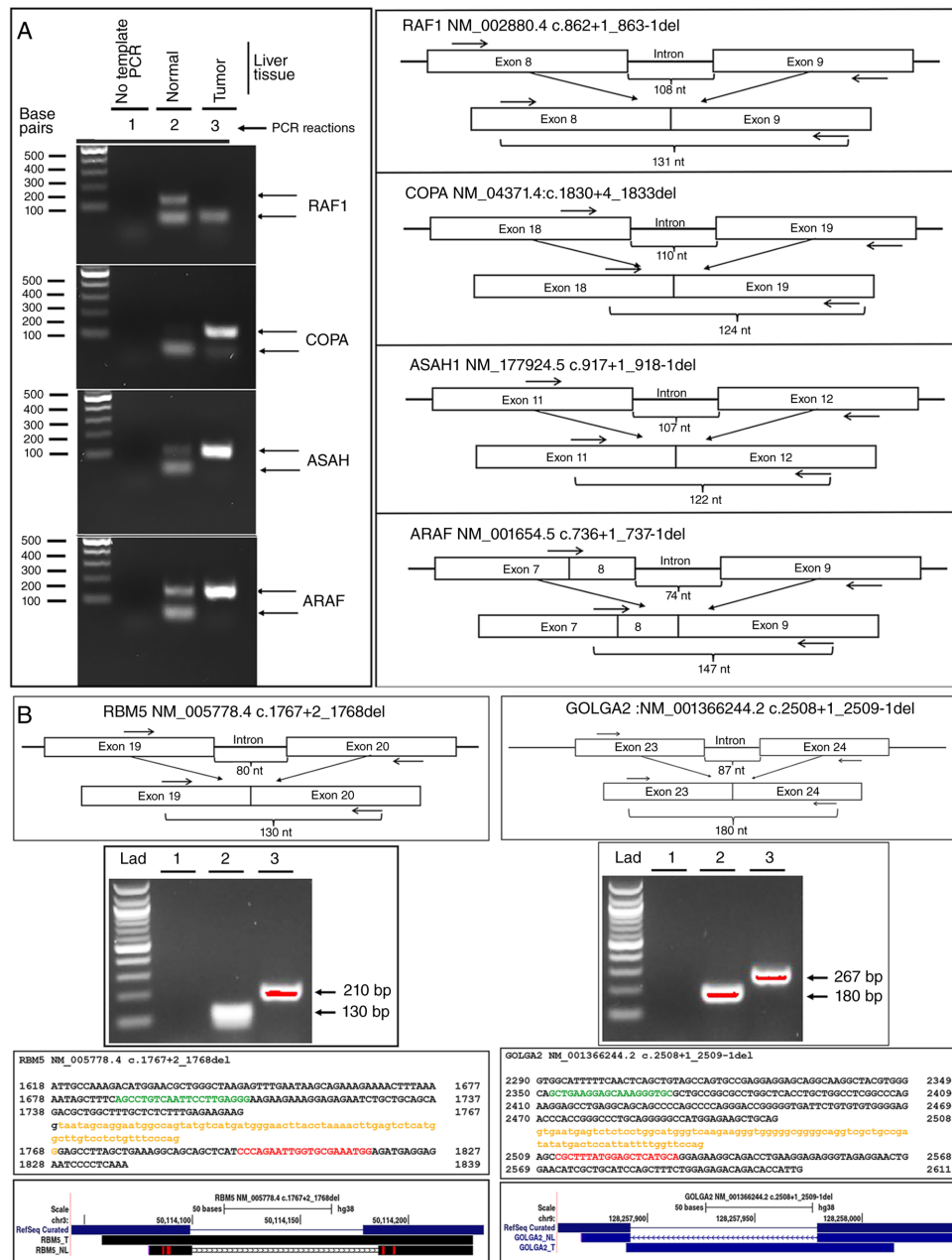


Figure 6. Tumor-specific ctRNA isoforms and intron retentions detected in CCA at exon-intron junctions encompassing indel variants. (A) Previous RNA-seq analysis identified high-risk variants concordant in tumors and plasma samples from patients with CCA. For validation of the RNA-seq variant calls, agarose gel images of PCR-amplified regions with ctRNA indels located at exon-intron junctions corresponding to normal liver and CCA tumor tissue are shown. Lane 1 in the gels represents a no-template PCR. The schematic diagrams on the side show exon-intron junctions corresponding to the indel variants. Arrows indicate the relative positions of the forward and reverse primers used for PCR. Tumor tissue from a patient with CCA (CCA005T) and normal liver tissue previously investigated by total RNA-seq were subjected to semi-qPCR targeting the exon-intron junctions corresponding to ctRNA indel sites in COPA, RAF1, ASAH1 and ARAF transcripts. The gels reflect either a loss of a transcript isoform or an amplicon of increased size. Schematics of the location of the indels around the exon-intron-exon site and predicted sizes with and without possible intron retention are shown. (B) Another two HCC- and CCA-specific indels corresponding to RBM5 and GOLGA2, detected with high frequency by RNA-seq, were investigated in normal liver tissue and CCA tumor tissue. Schematic representations of the exon-intron junctions in RBM5 and GOLGA2 encompassing the high-risk indel sites are shown. The arrows indicate the relative positions of the forward and reverse primers used for PCR. The sizes of the intron and the expected post-splice mature mRNA amplicon sizes are displayed. A 100-bp ladder and a PCR reaction with no template as a control are shown. PCR products corresponding to normal liver tissue and CCA tumor tissue were resolved on a 2% agarose gel. Lane 1 was a no template control reaction, so no amplification was expected, lane 2 is the amplified region in the normal liver sample and lane 3 is the amplified region in the CCA tumor sample. The CCA tumor tissue showed increased amplicon sizes corresponding to indels in RBM5 and GOLGA2, and intron retention could be predicted in both cases. Appearance of red streaks is caused by pixel saturation (overexposure) when balancing exposure times to visualize both high- and low-abundance bands. This often occurs while optimizing imaging settings to capture finer details and visualize any low-abundance bands across the gel. To confirm intron retention, amplicon products from normal tissue and CCA tissue were extracted from agarose gels and subjected to Sanger sequencing (Fig. S4). Reference sequences of the RBM5 and GOLGA2 transcripts are shown with forward and reverse primer locations highlighted in green and red, respectively. The graphic illustrates the Sanger sequencing results. The blue bar represents the Refseq Curated gene sequence for RBM5 and GOLGA2. The RBM5_T and GOLGA2_T bars show Sanger sequencing coverage through the intronic sequence of the gene in the tumor samples. RBM5_NL and GOLGA2_NL bars show missing intron sequences in the normal liver. ARAF, serine/threonine-protein kinase A-Raf; ASAH, N-acylsphingosine amidohydrolase; CCA, cholangiocarcinoma; COPA, coat protein complex I subunit α ; ctRNA, circulating RNA; GOLGA2, golgin A2; indel, insertion and deletion; Lad, ladder; NL, normal liver; RBM5, RNA binding motif protein 5; RNA-seq, RNA-sequencing; semi-qPCR, semi-quantitative PCR; T, tumor.

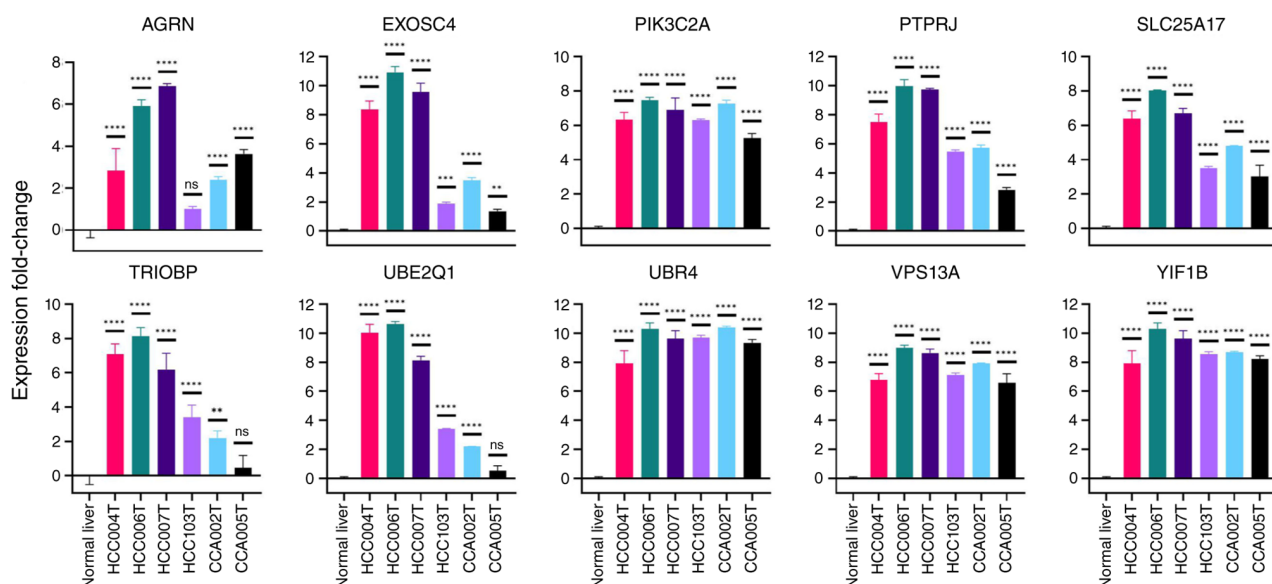


Figure 7. Development of TaqMan chemistry assays against ctRNA indel variants validated by semi-qPCR assays. TaqMan probes corresponding to the indel sites along exon-intron junctions of 10 high-risk HCC/CCA-specific indels were used to develop qPCR assays to test four HCC and two CCA tumor tissues. Normal healthy liver tissue was used as a control to demonstrate the specific detection of indel ctRNA variants. Each reaction was run in triplicate. Welch's ANOVA followed by Dunnett's T3 multiple comparisons test was used to identify statistically significant differences between each sample and the control for each target. ** $P < 0.01$ and **** $P < 0.0001$. Dunnett's T3 adjusted P-values for each sample are reported. Detailed statistical information is provided in Table SVIII. CCA, cholangiocarcinoma; ctRNA, circulating RNA; HCC, hepatocellular carcinoma; indel, insertion and deletion; ns, not significant; qPCR, quantitative PCR.

studied by semi-qPCR in CCA and healthy liver tissue, and longer amplicons were observed in the tumor (Fig. 6B). Primers encompassing the two adjacent exons separated by an intron, as shown in the schematic diagrams, were used to amplify the product using semi-qPCR. The resulting amplicons were resolved on a 2% agarose gel by electrophoresis. Tumor-associated bands of larger size were extracted from the agarose gels and the sequences were studied using Sanger sequencing (Fig. S4). RBM5 appeared to have an insertion of 80 nucleotides, the same size as the size of the intron at the investigated exon junction in the CCA tumor tissue. GOLGA2 appeared to have an insertion of 87 nucleotides, which again was the same size as the intron. Sanger sequencing results confirmed the intron retention of both RBM5 and GOLGA2 in CCA tissue. No such insertions were detected in the PCR amplicons excised from the normal liver (Fig. S4). The Sanger sequencing results were blasted against the published sequence for each gene, and the sequence of the excised band closely matched the canonical exon-intron-exon sequence in NCBI, confirming intron retention as a feature of HCC (Fig. 6B). The results suggested that splicing aberrations in tumors may involve mis-splicing events such as retaining introns and affirmed the validity of the variant calls in ctRNA of patients with HCC using RNA-seq.

Development of TaqMan chemistry-based assays for sensitive detection of splice variants. While semi-qPCR is an effective method for confirming ctRNA variants, it is not suitable as a high-throughput translational assay. To improve the efficiency of detecting these tumor-specific splice variants, TaqMan chemistry-based primer-probe sets were designed to detect these HCC-specific high-risk splice site variants using a qPCR assay. The single-probe assay allowed for variant detection

above the background noise of the WT mRNA sequences. A total of four HCC and two CCA tumor tissues were investigated using variant splice site-specific TaqMan chemistry assays using single probes corresponding to indel/splice junctions. As a qualitative detection system, these variant-specific probes must allow for detection of these HCC-specific ctRNA variants and distinguish HCC tumors from healthy liver tissue (Fig. 7). Using relative expression analysis, the detection of these HCC-specific high-risk variants, which were earlier validated by semi-quantitative endpoint PCR, was demonstrated. For exosome component 4 (EXOSC4), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit α (PIK3C2A), PTPRJ, SLC25A17, UBR4, VPS13A and YIF1B, 1.5 to >10-fold upregulated expression of ctRNA variants was observed in tumors compared with normal tissue. The assays for AGRN, TRIO and F-actin binding protein (TRIOBP) and UBE2Q1 also showed upregulation of ctRNA variants in most of the tumors; however, in HCC103T, AGRN ctRNA's expression upregulation was not statistically significant, and in CCA005T, the upregulation of TRIOBP and UBE2Q1's ctRNA variants was also not statistically significant. Additionally, these 10 assays demonstrated significant upregulation relative to the control baseline following Holm correction (adjusted $P < 0.05$), with 95% confidence intervals that did not cross zero. Effect sizes ranged from large (Hedges' g , 1.17) to extremely large ($g > 8$). UBR4, YIF1B, VPS13A and PIK3C2A exhibited the largest and most precise effects, as reflected by both high standardized effect sizes and confidence intervals with no negative limits (Table SVIII). These results suggested that this approach may be used for testing the occurrence of such variants in plasma samples of patients with chronic liver disease for use as surveillance biomarkers for different types of liver cancer.

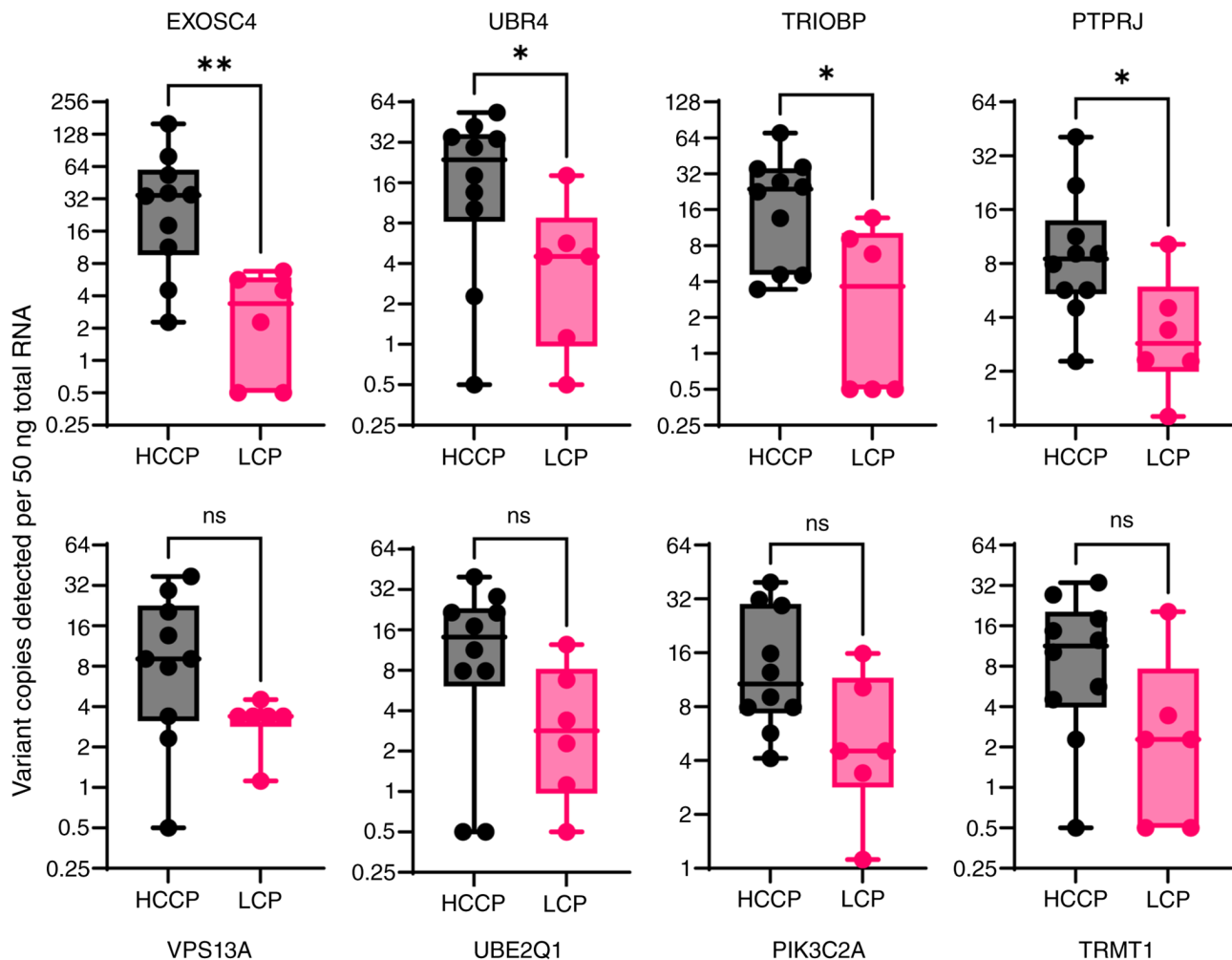


Figure 8. Investigation of plasma samples from a small independent cohort of patients with HCC and LC-without HCC using insertion and deletion-based TaqMan assays and dPCR. RNA was isolated from plasma samples from an independent cohort of patients diagnosed with HCC (n=10) and a control set of patients with LC with no HCC (n=6). Total RNA was converted to cDNA and 50 ng was used as the input for each assay. A dPCR reaction was performed using TaqMan single probe assays for each sample, with a fluorescein amidite reporter dye released when the ctRNA sequence was amplified. The resulting copy numbers of ctRNA for the HCC and LC plasma samples are plotted on a log₂ scale with the value 0.5 copies on the graph representing 0 copies detected to include all data points. Statistical analysis was performed for each target using a Mann-Whitney test to determine significant differences between the HCC and LC samples, followed by a Holm-Bonferroni correction to evaluate statistically significant differences for all the target assays as a complete panel. The box and whisker plots show the ctRNA copy numbers per 50 ng of input cDNA in aggregate for the HCC samples on the left and the LC samples on the right, plotted on a log₂ scale. Mann-Whitney test results: EXOSC4, P=0.007; UBR4, P=0.045; TRIOBP, P=0.031; PTPRJ, P=0.038; VPS13A, P=0.084; UBE2Q1, P=0.086; PIK3C2A, P=0.077; and TRMT1, P=0.091. Holm-Bonferroni correction resulted in the following adjusted P-values: EXOSC4, P=0.056; UBR4, P=0.226; TRIOBP, P=0.214; PTPRJ, P=0.226; VPS13A, P=0.307; UBE2Q1, P=0.307; PIK3C2A, P=0.307; and TRMT1, P=0.307. The P-values in the figure represent the Mann-Whitney test results. *P<0.05, **P<0.01. A detailed statistical summary is provided in Table SIX. Each black and pink dot represents the number of variant allele copies detected in an HCCP or LCP sample, respectively. All cDNA input amounts were normalized to 50 ng per reaction. ctRNA, circulating RNA; dPCR, digital PCR; EXOSC4, exosome component 4; HCC, hepatocellular carcinoma; HCCP, HCC plasma; LC, liver cirrhosis; LCP, LC plasma; ns, not significant; PIK3C2A, phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 α ; PTPRJ, protein tyrosine phosphatase receptor type J; TRIOBP, TRIO and F-actin binding protein; TRMT1, transfer RNA methyltransferase 1; UBE2Q1, ubiquitin conjugating enzyme E2 Q1; UBR4, ubiquitin protein ligase E3 component n-recogin 4; VPS13A, vacuolar protein sorting 13 homolog A.

Transitioning from tumor tissue to blood to test splice variant assays in patients with HCC. Having validated the ctRNA variant-specific assays in liver tumor tissues, the next step was to test the functionality of these assays in plasma from patients with HCC in an effort to develop tools for HCC surveillance using non-invasive detection. TaqMan assays developed against high-impact variants, tested earlier on tumor tissues, were conducted using plasma samples from a small independent cohort of patients with HCC (n=10) and patients without cancer but with LC (n=6). The ctmutRNA variants corresponding to UBR4, UBE2Q1, TRIOBP, EXOSC4, VPS13A, PTPRJ, PIK3C2A and TRMT1 were investigated using

the TaqMan assays and plasma samples from patients with HCC and LC by dPCR. The experiment was performed on an Absolute Q dPCR platform (Fig. 8). The box and whisker plots demonstrated that variant copy numbers of indel targets corresponding to EXOSC4, UBR4, TRIOBP and PTPRJ were significantly upregulated and abundant in patients with HCC compared with patients with LC. Copy numbers of indel variants corresponding to VPS13A, UBE2Q1, PIK3C2A and TRMT1 were higher in patients with HCC, although the expression differences were not statistically significant for all eight genes in this small test cohort (Table SIX). The results of the present study not only affirmed the validity of assays using

plasma samples but also underscored the biomarker value of ctRNA targets for HCC and CCA surveillance. Table SI summarizes the mapping of clinical samples to the specific figures in the present study.

Discussion

Conventional surveillance strategies for HCC exhibit suboptimal performance for early detection and blood-based assays are actively being evaluated (48). Development of circulating nucleic acid-based signatures could facilitate early detection, identify high-risk patients, guide personalized therapeutics and predict tumor relapses. In the post-intervention setting, incorporating mutation profiles into current histological grading and TNM staging (AJCC, BCLC and Cancer of the Liver Italian Program) could enhance the clinical stratification of patients at high risk for tumor relapse (49,50).

Somatic mutations, which are genetic alterations that are not inherited and emerge in non-germline cells during a person's lifetime, have been used for diagnosis and targeted therapy selection (51). Our previous studies and others have reported that ctRNA mutations associated with cancer predominantly comprise SNPs and indels (6,7,9,52). In coding regions, an SNP or indel can be frameshift (FS) or non-frameshift (NFS). FS variants in coding regions that change open reading frames are potentially deleterious (53), while NFS variants where the length is a multiple of three nucleotides only affect the amino acid(s) of the variant. RNA-seq has enabled the identification of numerous SNPs and variations of alternatively processed RNAs (54,55). A multitude of SNPs at splicing junctions and deregulated RNA modifiers or regulators promote splicing aberrations associated with cancer progression. Aberrant RNA splicing is known to result in the production of oncogenic isoforms that promote tumorigenesis and cancer progression (18). Several tumor-specific alternative splice isoforms have been reported to be associated with overall survival in HCC (56-58). While RNA-seq can readily detect SNPs (59), alternative splicing complexity makes indel detection comparatively challenging. The present study attempted to characterize some HCC-associated SNPs and indel variants identified in the circulation of patients with HCC using total RNA-seq, and to develop assays to validate the detection of these ctmutRNA variants in tumor tissues and to finally test these assays using plasma samples from patients with HCC.

RNA-seq-based approaches for capturing multiple and diverse types of mutations simultaneously have been reported previously (60). ctmutRNA variants exclusively associated with patients with HCC have been identified but the role they serve in HCC needs to be further investigated. High-risk mutations in certain transcripts corresponding to GTDC1, MAF1, NUGGC and MST1, which are known to be associated with HCC (TCGA protein atlas database), warrant investigation to determine how these mutations influence HCC. A high-impact HCC-specific SNP in APOB, reported in the present study, is known to be associated with dyslipidemia and ischemic heart disease (61). Similarly, SNPs in TF and KLKB1 are known to be associated with hemochromatosis (62) and venous thromboembolism (63), respectively. Additionally, validated high-risk SNPs in CD151 genes are known to be associated

with lung squamous cell carcinoma (64), while PCK1 has been reported to be a potential therapeutic target in cancer (32). In the present study, high-impact SNPs associated with some splicing factors were reported and confirmed. SRPK1, a serine arginine protein kinase, acts as a regulator of constitutive and alternative splicing by phosphorylating splicing factors rich in Ser/Arg domains. SRPK1 is an oncoprotein involved in PI3K/AKT signaling in HCC (65). Indel variants associated with RBM5 and GOLGA2 leading to intron retention were also observed in the present study. RBM5 is a part of the spliceosome A complex and encodes an RNA binding protein that harbors a structural module for RNA recognition and regulates RNA splicing (66). It could be hypothesized that mutations in splicing factors, such as SRPK1 and RBM5, introduce splicing aberrations in cancer cells, leading to cancer-specific isoforms. Intron retention is considered to contribute to the transcriptional diversity of cancer cells and may serve a role in cancer pathogenesis and drug resistance (67). Furthermore, retained introns can potentially be presented as neoantigens on the surface of cancer cells, which is a potential target for personalized cancer vaccines (68).

Aberrant splicing events commonly detected in the liver include intron retention, exon skipping and alternative usage of the first exon (17,19). In the present study, intron retention in GOLGA2 in liver tumors was reported; however, exon exclusion in GOLGA2 has been reported to contribute to tumorigenesis mediated by PTEN abrogation (69). Compared with normal liver, primary HCC tumors exhibit extensive differential splicing and a number of these splicing aberrations are associated with patient survival (20,70). Alternative splicing in HCC produces novel protein isoforms with distinct, occasionally opposing functions, from canonical counterparts (71,72), affecting genes involved in the cell cycle, DNA repair, metabolism and epithelial-mesenchymal transition (73). Metabolism-related genes, particularly those involved in carbohydrate processing, are most commonly affected (16). Some splicing events are known to be strongly associated with HCC etiology (74) and may represent novel disease biomarkers.

Consistent with observations in the present study, HCC-associated splice variants in RAF1 and ARAF have been previously reported to alter protein structure and function, potentially causing oncogenic effects (75). Similarly, pre-mRNA of COPA is known to undergo alternative splicing to produce different transcripts and protein isoforms (76,77). The splice variant corresponding to the ASAH1 gene has been reported to be dysregulated in cancer and has been proposed as a therapeutic target in cancer (78). In alignment with the findings of the present study, ASAH1 is known to undergo alternative splicing, generating multiple transcript variants (79).

To characterize HCC-specific circulating splicing variants identified by RNA-seq, primers encompassing the corresponding variant splicing junction were designed to evaluate the size of amplicons in tumors and normal liver tissue using semi-qPCR. The variability of PCR amplicon sizes in tumors corresponding to several variant splice junctions was demonstrated and it was hypothesized that the mutations in the splicing factors contributed to the dysregulation of spliceosome machinery and consequent aberrant splicing during

tumor development and progression (80). Altered expression and mutations of splicing factors may generate tumor-specific isoforms in cancer (81). Tumor cells exploit aberrant RNA splicing to develop, grow and progress into therapy-resistant malignancies (82). Therefore, dissecting tumor-specific RNA isoforms and the splicing networks between alternative splice forms and splicing factors may reveal mechanisms underlying tumor development and progression, potentially identifying prognostic biomarkers and therapeutic targets.

One limitation of the present study was the lack of somatic or germline characterization of the ctRNA variants. Without access to paired lymphocytes or matching adjacent liver tissues, it was not possible to evaluate the somatic or germline nature of these variants. Additionally, due to limited plasma volumes, ctmutRNA biomarker performance could not be compared with levels of α -fetoprotein (AFP), AFP-L3 and des- γ carboxyprothrombin (DCP) in the same samples, and therefore, GALAD score correlations were not possible. The GALAD score, derived from sex, age, and levels of AFP, AFP-L3 and DCP, is a highly accurate statistical model used to estimate the probability of developing HCC in patients with chronic liver disease (83). However, these head-to-head comparisons, along with functional characterization of these ctmutRNA variants in larger patient cohorts, are planned for future validation studies.

Liquid biopsy is evolving from signal discovery to deployable and clinically integrated multiparametric decision systems facilitating surveillance, early detection and monitoring (84). The present proof-of-concept study investigated a small cohort of patients to highlight the role of ctRNA variants as potential cancer biomarkers for non-invasive detection of liver cancer. Through a rigorous discovery and validation process, a series of RNA-seq-derived SNP and indel ctRNA variants associated with HCC and CCA were characterized and validated. Additionally, the present study highlighted some novel circulating cancer-specific splicing aberrations in RNAs. The FSs introduced by SNP and indel variants likely caused pathologic changes such as intron retention or exon skipping, resulting in early termination codons and truncated proteins with a considerable role in cancer development and progression. If validated in larger patient cohorts, these novel biomarkers will be useful in early non-invasive detection and effective management of patients with liver cancer.

Acknowledgements

The authors would like to thank Ms. Tina Bergner (Capital Health Cancer Center, Pennington, NJ, USA) for assistance with blood and tumor tissue collection.

Funding

The present study was supported by the Commonwealth of Pennsylvania's Angel Investment Venture Capital Program (grant no. OB-2023-07) and the Commonwealth University Research Enhancement Program (grant no. SAP# 4100095321). The present study was also supported by a NIH NCATS grant to the Medical University of South Carolina (grant no. A00-2219-5010), and private funding sources viz Steve Miller Fund and The Edmund and Carol Blake Foundation.

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

DZ and AS performed the experiments, analyzed the results and created the figures. DZ and AS created the initial draft. TZ carried out statistical analysis and analyzed the diagnostic performance of multiple ordered-panels of variant-collections. TMB, CD, MAH and DEK made contributions to conception and interpretation of data, reviewing the data and revising the initial draft. Each author participated in the work and takes full responsibility for the content. CD and MAH provided clinical samples. AS and DZ confirm the authenticity of all the raw data. All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participate

The present study was approved by the University of Pennsylvania Institutional Review Board #7 (approval no. 828260; Philadelphia, PA, USA) and the Institutional Review Board at Capital Health Medical Center (approval no. 00002877; Pennington, NJ, USA). The patients provided written informed consent. All study procedures were carried out in accordance with the ethical principles outlined in the Declaration of Helsinki. All participants were informed about the purpose of the study and assured of confidentiality and provided written consent prior to participation. Participation was voluntary, and respondents could withdraw at any time without consequence.

Patient consent for publication

Not applicable.

Competing interests

TB is on the board of the Directors of Hepion, a company developing medicines for liver diseases, Merlin, a company developing medicines for cancer, and CIRNA, a startup company owned by the Blumberg Institute and focused on the development of tools for early cancer detection. AS is the CSO for CIRNA. The other authors declare that they have no competing interests.

References

1. Repetto O, Vettori R, Steffan A, Cannizzaro R and De Re V: circulating proteins as diagnostic markers in gastric cancer. *Int J Mol Sci* 24: 16931, 2023.
2. Bittla P, Kaur S, Sojitra V, Zahra A, Hutchinson J, Folawemi O and Khan S: Exploring circulating tumor DNA (CtDNA) and its role in early detection of cancer: A systematic review. *Cureus* 15: e45784, 2023.
3. Zhong P, Bai L, Hong M, Ouyang J, Wang R, Zhang X and Chen P: A comprehensive review on circulating cfRNA in plasma: Implications for disease diagnosis and beyond. *Diagnostics (Basel)* 14: 1045, 2024.
4. Alanazi SA, Alshammari N, Alruwaili M, Junaaid K, Abid MR and Ahmad F: Integrative analysis of RNA expression data unveils distinct cancer types through machine learning techniques. *Saudi J Biol Sci* 31: 103918, 2024.

5. Sayeed A, Dalvano BE, Kaplan DE, Viswanathan U, Kulp J, Jannet AH, Hwang LY, Ertel A, Doria C and Block T: Profiling the circulating mRNA transcriptome in human liver disease. *Oncotarget* 11: 2216-2232, 2020.
6. Block T, Zezulinski D, Kaplan DE, Lu J, Zanine S, Zhan T, Doria C and Sayeed A: Circulating mRNA variants as a potential biomarker for surveillance of hepatocellular carcinoma. *Front Oncol* 12: 963641, 2022.
7. Zezulinski D, Hoteit MA, Kaplan DE, Simeone A, Zhan T, Doria C, Ahmed FY, Roberts LR, Block TM and Sayeed A: Detection of circulating mRNA variants in hepatocellular carcinoma patients using targeted RNAseq. *Liver Cancer* 14: 555-586, 2025.
8. Hunold T and Parikh ND: Risk-stratified surveillance for hepatocellular carcinoma. *Hepatol Commun* 10: e0943, 2026.
9. Kim KH and Yoo BC: Circulating RNA as a functional component of liquid biopsy in cancer: Concepts, classification, and clinical applications. *Int J Mol Sci* 27: 2403, 2026.
10. International HapMap Consortium: The International HapMap Project. *Nature* 426: 789-796, 2003.
11. Collins FS, Brooks LD and Chakravarti A: A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* 8: 1229-1231, 1998.
12. Ng PC and Henikoff S: Accounting for human polymorphisms predicted to affect protein function. *Genome Res* 12: 436-446, 2002.
13. Ronaghi M and Langae T: Single nucleotide polymorphisms: Discovery, detection and analysis. *Per Med* 2: 111-125, 2005.
14. Lai J, Yang J, Gamsiz Uzun ED, Rubenstein BM and Sarkar IN: LYRUS: A machine learning model for predicting the pathogenicity of missense variants. *Bioinform Adv* 2: vbab045, 2021.
15. Adzhubei I, Jordan DM and Sunyaev SR: Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter 7: Unit7.20*, 2013.
16. Lee SE, Alcedo KP, Kim HJ and Snider NT: Alternative splicing in hepatocellular carcinoma. *Cell Mol Gastroenterol Hepatol* 10: 699-712, 2020.
17. Chen H, Gao F, He M, Ding XF, Wong AM, Sze SC, Yu AC, Sun T, Chan AW, Wang X and Wong N: Long-Read RNA sequencing identifies alternative splice variants in hepatocellular carcinoma and tumor-specific isoforms. *Hepatology* 70: 1011-1025, 2019.
18. Chen H, Tang J and Xiang J: Alternative splicing in tumorigenesis and cancer therapy. *Biomolecules* 15: 789, 2025.
19. Li S, Hu Z, Zhao Y, Huang S and He X: Transcriptome-wide analysis reveals the landscape of aberrant alternative splicing events in liver cancer. *Hepatology* 69: 359-375, 2019.
20. Zhu GQ, Zhou YJ, Qiu LX, Wang B, Yang Y, Liao WT, Luo YH, Shi YH, Zhou J, Fan J and Dai Z: Prognostic alternative mRNA splicing signature in hepatocellular carcinoma: A study based on large-scale sequencing data. *Carcinogenesis* 40: 1077-1085, 2019.
21. Lee YT, Wang JJ, Zhu Y, Agopian VG, Tseng HR and Yang JD: Diagnostic Criteria and LI-RADS for hepatocellular carcinoma. *Clin Liver Dis (Hoboken)* 17: 409-413, 2021.
22. Reig M, Forner A, Rimola J, Ferrer-Fàbrega J, Burrel M, Garcia-Criado A, Kelley RK, Galle PR, Mazzaferro V, Salem R, *et al*: BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. *J Hepatol* 76: 681-693, 2022.
23. Liao X and Zhang D: The 8th Edition American joint committee on cancer staging for hepato-pancreato-biliary cancer: A review and update. *Arch Pathol Lab Med* 145: 543-553, 2021.
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR: STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21, 2013.
25. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402-408, 2001.
26. Lin Z, Ji X, Tian N, Gan Y and Ke L: APOB is a potential prognostic biomarker in hepatocellular carcinoma. *Discov Oncol* 15: 28, 2024.
27. Shimono N, Nishimura Y, Kamiguchi H and Nishikawa Y: Remarkable expression in the colon adenocarcinoma of Hmat-Xa, a human mannosyltransferase-like gene, that is homologous to drosophila gene GC15914. *Biosci Biotechnol Biochem* 75: 1451-1455, 2011.
28. Che YQ, Zhang Y, Li HB, Shen D and Cui W: Serum KLKB1 as a potential prognostic biomarker for hepatocellular carcinoma based on data-independent acquisition and parallel reaction monitoring. *J Hepatocell Carcinoma* 8: 1241-1252, 2021.
29. Zhu J, Cai T, Zhou J, Du W, Zeng Y, Liu T, Fu Y, Li Y, Qian Q, Yang XH, *et al*: CD151 drives cancer progression depending on integrin $\alpha 3 \beta 1$ through EGFR signaling in non-small cell lung cancer. *J Exp Clin Cancer Res* 40: 192, 2021.
30. Zhang K, Zhu Z, Jia R, Wang NA, Shi M, Wang Y, Xiang S, Zhang Q and Xu L: CD151-enriched migrasomes mediate hepatocellular carcinoma invasion by conditioning cancer cells and promoting angiogenesis. *J Exp Clin Cancer Res* 43: 160, 2024.
31. Xiang J, Wang K and Tang N: PCK1 dysregulation in cancer: Metabolic reprogramming, oncogenic activation, and therapeutic opportunities. *Genes Dis* 10: 101-112, 2022.
32. Liu N, Zhu XR, Wu CY, Liu YY, Chen MB and Gu JH: PCK1 as a target for cancer therapy: From metabolic reprogramming to immune microenvironment remodeling. *Cell Death Discov* 10: 478, 2024.
33. Nikas IP, Themistocleous SC, Paschou SA, Tsamis KI and Ryu HS: Serine-arginine protein kinase 1 (SRPK1) as a prognostic factor and potential therapeutic target in cancer: Current evidence and future perspectives. *Cells* 9: 19, 2019.
34. Rudler DL, Hughes LA, Perks KL, Richman TR, Kuznetsova I, Ermer JA, Abdululai LN, Shearwood AJ, Viola HM, Hool LC, *et al*: Fidelity of translation initiation is required for coordinated respiratory complex assembly. *Sci Adv* 5: eaay2118, 2019.
35. Zhang XQ and Li L: Biological function and clinical value of VPS13A in pan-cancer based on bioinformatics analysis. *Int J Gen Med* 14: 6825-6838, 2021.
36. Fournier P, Dussault S, Fusco A, Rivard A and Royal I: Tyrosine phosphatase PTPRJ/DEP-1 is an essential promoter of vascular permeability, angiogenesis, and tumor progression. *Cancer Res* 76: 5080-5091, 2016.
37. Chou HJ, Donnard E, Gustafsson HT, Garber M and Rando OJ: Transcriptome-wide analysis of roles for tRNA modifications in translational regulation. *Mol Cell* 68: 978-992.e4, 2017.
38. Cervia LD, Shibue T, Borah AA, Gaeta B, He L, Leung L, Li N, Moyer SM, Shim BH, Dumont N, *et al*: A ubiquitination cascade regulating the integrated stress response and survival in carcinomas. *Cancer Discov* 13: 766-795, 2023.
39. Seo G, Mckinley J and Wang W: MAP4K2 connects the Hippo pathway to autophagy in response to energy stress. *Autophagy* 20: 704-706, 2024.
40. Shi Y, Huang J, Hu Y and Shen Y: Multiomics data analyses to identify SLC25A17 as a novel biomarker to predict the prognosis and immune microenvironment in head and neck squamous cell carcinoma. *BMC Bioinformatics* 24: 269, 2023.
41. Abegunde SO, Grieve S, Alfarra H and Reiman T: MST1 down-regulates TAZ tumor suppressor protein in multiple myeloma and is a potential therapeutic target. *Exp Hematol* 123: 34-45, 2023.
42. Ng YK, Lau WS, Lui VW, Cheng AS, Ng PK, Tsui SK, Cheung YS and Lai PB: Full-length Mst1 exhibits growth promoting function in human hepatocellular carcinoma cells. *FEBS Lett* 587: 496-503, 2013.
43. Kawanaka T, Emma I, Christine H and Fei-Fei L: Prognostic significance of UROD expression on cervical cancer with radiation therapy. *Int J Radiat Oncol Biol Phys*: Oct 1, 2013 (Epub ahead of print).
44. Liu Z and Lv C: RNA binding protein PUM2 promotes hepatocellular carcinoma proliferation and apoptosis via binding to the 3'UTR of BTG3. *Oncol Lett* 24: 346, 2022.
45. Gong Y, Liu Z, Yuan Y, Yang Z, Zhang J, Lu Q, Wang W, Fang C, Lin H and Liu S: PUMILIO proteins promote colorectal cancer growth via suppressing p21. *Nat Commun* 13: 1627, 2022.
46. Zhang H, Liang J, Lu T, Li M, Shan G, Bi G, Zhao M, Jin X, Wang Q, Chen Z and Zhan C: AGRN promotes lung adenocarcinoma progression by activating Notch signaling pathway and acts as a therapeutic target. *Pharmacol Res* 194: 106819, 2023.
47. Topno R, Singh I, Kumar M and Agarwal P: Integrated bioinformatic analysis identifies UBE2Q1 as a potential prognostic marker for high grade serous ovarian cancer. *BMC Cancer* 21: 220, 2021.
48. Melendez-Torres J and Singal AG: Early detection of hepatocellular carcinoma: Roadmap for improvement. *Expert Rev Anticancer Ther* 22: 621-632, 2022.
49. Pons F, Varela M and Llovet JM: Staging systems in hepatocellular carcinoma. *HPB (Oxford)* 7: 35-41, 2005.
50. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR and Winchester DP: The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more 'personalized' approach to cancer staging. *CA Cancer J Clin* 67: 93-99, 2017.

51. Horlings HM, Shah SP and Huntsman DG: Using somatic mutations to guide treatment decisions: Context matters. *JAMA Oncol* 1: 275-276, 2015.
52. Tang G, Liu X, Cho M, Li Y, Tran DH and Wang X: Pan-cancer discovery of somatic mutations from RNA sequencing data. *Commun Biol* 7: 619, 2024.
53. Lin M, Whitmire S, Chen J, Farrel A, Shi X and Guo JT: Effect of short indels on protein structure and function in human genomes. *Sci Rep* 7: 9313, 2017.
54. Yang HD and Nam SW: Pathogenic diversity of RNA variants and RNA variation-associated factors in cancer development. *Exp Mol Med* 52: 582-593, 2020.
55. Chen J and Guo JT: Comparative assessments of indel annotations in healthy and cancer genomes with next-generation sequencing data. *BMC Med Genomics* 13: 170, 2020.
56. Chen XX, Zhang BH, Lu YC, Li ZQ, Chen CY, Yang YC, Chen YJ and Ma D: A novel 16-gene alternative mRNA splicing signature predicts tumor relapse and indicates immune activity in stage I-III hepatocellular carcinoma. *Front Pharmacol* 13: 939912, 2022.
57. Cai Y, Xia J, Wang N and Zhou H: Identification of prognostic alternative splicing signatures in hepatitis B or/and C viruses related hepatocellular carcinoma. *Genomics* 112: 3396-3406, 2020.
58. Wu F, Chen Q, Liu C, Duan X, Hu J, Liu J, Cao H, Li W and Li H: Profiles of prognostic alternative splicing signature in hepatocellular carcinoma. *Cancer Med* 9: 2171-2180, 2020.
59. Wang C, Davila JI, Baheti S, Bhagwate AV, Wang X, Kocher JP, Slager SL, Feldman AL, Novak AJ, Cerhan JR, *et al*: RVboost: RNA-seq variants prioritization using a boosting method. *Bioinformatics* 30: 3414-3416, 2014.
60. Prodduturi N, Bhagwate A, Kocher JA and Sun Z: Indel sensitive and comprehensive variant/mutation detection from RNA sequencing data for precision medicine. *BMC Med Genomics* 11 (Suppl 3): S67, 2018.
61. Zhang FH, Yin RX, Yao LM, Yan RQ, Lu L and Su Y: Correlation between the *APOB* rs1042034 SNP and blood lipid characteristics of 2 ethnic groups in China. *Clin Appl Thromb Hemost* 25: 1076029619892088, 2019.
62. Raffield LM, Louie T, Sofer T, Jain D, Ipp E, Taylor KD, Papanicolaou GJ, Avilés-Santa L, Lange LA, Laurie CC, *et al*: Genome-wide association study of iron traits and relation to diabetes in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL): Potential genomic intersection of iron and glucose regulation? *Hum Mol Genet* 26: 1966-1978, 2017.
63. Gittleman HR, Merkulova A, Alhalabi O, Stavrou EX, Veigl ML, Barnholtz-Sloan JS and Schmaier AH: A cross-sectional study of *KLKB1* and *PRCP* polymorphisms in patient samples with cardiovascular disease. *Front Med (Lausanne)* 3: 17, 2016.
64. Kwon MJ, Seo J, Kim YJ, Kwon MJ, Choi JY, Kim TE, Lee DH, Park S, Shin YK, Han J and Choi YL: Prognostic significance of *CD151* overexpression in non-small cell lung cancer. *Lung Cancer* 81: 109-116, 2013.
65. Zhou B, Li Y, Deng Q, Wang H, Wang Y, Cai B and Han ZG: *SRPK1* contributes to malignancy of hepatocellular carcinoma through a possible mechanism involving *PI3K/Akt*. *Mol Cell Biochem* 379: 191-199, 2013.
66. Soni K, Jagtap PKA, Martínez-Lumbreras S, Bonnal S, Geerlof A, Stehle R, Simon B, Valcárcel J and Sattler M: Structural basis for specific RNA recognition by the alternative splicing factor *RBM5*. *Nat Commun* 14: 4233, 2023.
67. Dvinge H and Bradley RK: Widespread intron retention diversifies most cancer transcriptomes. *Genome Med* 7: 45, 2015.
68. Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D, Fugmann T, Wong KK and Van Allen EM: Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* 36: 1056-1058, 2018.
69. Shen SM, Ji Y, Zhang C, Dong SS, Yang S, Xiong Z, Ge MK, Yu Y, Xia L, Guo M, *et al*: Nuclear *PTEN* safeguards pre-mRNA splicing to link Golgi apparatus for its tumor suppressive role. *Nat Commun* 9: 2392, 2018.
70. Shilo A, Ben Hur V, Denichenko P, Stein I, Pikarsky E, Rauch J, Kolch W, Zender L and Karni R: Splicing factor hnRNP A2 activates the Ras-MAPK-ERK pathway by controlling A-Raf splicing in hepatocellular carcinoma development. *RNA* 20: 505-515, 2014.
71. Wang H, Zhang CZ, Lu SX, Zhang MF, Liu LL, Luo RZ, Yang X, Wang CH, Chen SL, He YF, *et al*: A coiled-coil domain containing 50 splice variant is modulated by serine/arginine-rich splicing factor 3 and promotes hepatocellular carcinoma in mice by the Ras signaling pathway. *Hepatology* 69: 179-195, 2019.
72. Lin KT, Ma WK, Scharner J, Liu YR and Krainer AR: A human-specific switch of alternatively spliced *AFMID* isoforms contributes to TP53 mutations and tumor recurrence in hepatocellular carcinoma. *Genome Res* 28: 275-284, 2018.
73. Lin KT, Shann YJ, Chau GY, Hsu CN and Huang CY: Identification of latent biomarkers in hepatocellular carcinoma by ultra-deep whole-transcriptome sequencing. *Oncogene* 33: 4786-4794, 2014.
74. Tremblay MP, Armero VE, Allaire A, Boudreault S, Martenon-Brodeur C, Durand M, Thibault P, Tremblay-Létourneau M, Perreault JP, Scott MS and Bisaillon M: Global profiling of alternative RNA splicing events provides insights into molecular differences between various types of hepatocellular carcinoma. *BMC Genomics* 17: 683, 2016.
75. Prall OWJ, Nastevski V, Xu H, McEvoy CRE, Vissers JHA, Byrne DJ, Takano E, Yerneni S, Ellis S, Green T, *et al*: *RAF1* rearrangements are common in pancreatic acinar cell carcinoma. *Mod Pathol* 33: 1811-1821, 2020.
76. Song Y, An O, Ren X, Chan THM, Tay DJT, Tang SJ, Han J, Hong H, Ng VHE, Ke X, *et al*: RNA editing mediates the functional switch of *COPA* in a novel mechanism of hepatocarcinogenesis. *J Hepatol* 74: 135-147, 2021.
77. Choi S, Cho N, Kim EM and Kim KK: The role of alternative pre-mRNA splicing in cancer progression. *Cancer Cell Int* 23: 249, 2023.
78. Bai J, Li P, Jiao H, Jin Y, Wang H, Jiang Q, Song F, Peng X and Qu Y: A novel intronic variant in the *ASAH1* gene enhances aberrant splicing, causing spinal muscular atrophy with progressive myoclonic epilepsy. *Ital J Pediatr* 51: 214, 2025.
79. Pan Q, Shai O, Lee L, Frey BJ and Blencowe BJ: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40: 1413-1415, 2008.
80. Dvinge H, Kim E, Abdel-Wahab O and Bradley RK: RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* 16: 413-430, 2016.
81. Urbanski LM, Leclair N and Anczukow O: Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip Rev RNA* 9: e1476, 2018.
82. Dong X and Chen R: Understanding aberrant RNA splicing to facilitate cancer diagnosis and therapy. *Oncogene* 39: 2231-2242, 2020.
83. Zhu H, Li Z, Xie K, Kassim SH, Cao C, Huang K, Lu Z, Ma C, Li Y, Jiang K and Yin L: Liquid Biopsy in Early Screening of Cancers: Emerging Technologies and New Prospects. *Biomedicines* 14: 158, 2026.
84. Yang JD, Addissie BD, Mara KC, Harmsen WS, Dai J, Zhang N, Wongjarupong N, Ali HM, Ali HA, Hassan FA, *et al*: GALAD score for hepatocellular carcinoma detection in comparison with liver ultrasound and proposal of GALADUS score. *Cancer Epidemiol Biomarkers Prev* 28: 531-538, 2019.



Copyright © 2026 Zezulinski et al. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.