Prognostic Value of Survival–Associated Splicing Factor SNRPA1 Overexpression And Its Potential Mechanism In Liver Cancer

Mengying Liu
Beijing University of Chemical Technology
Yueda Lu
Beijing University of Chemical Technology
Lei Wei
Beijing National Research Center for Information Science and Technology, Tsinghua University
Xiaowo Wang
Beijing National Research Center for Information Science and Technology, Tsinghua University
Yinying Lu
Fifth medical center of Chinese PLA general hospital
Xiaodong Jia
Fifth medical center of Chinese PLA general hospital
Shihui Wang
Beijing University of Chemical Technology
Wenlin An (wenlin.an@mail.buct.edu.cn)
Beijing University of Chemical Technology

Research Article

Keywords: SNRPA1, Liver cancer, Prognosis, Splicing factors, SCP2

Posted Date: June 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-509065/v1

Abstract

Background: Small nuclear ribonucleoprotein polypeptide A (SNRPA1) is a splicing factor (SF) responsible for the processing of pre−mRNA into mRNA. The purpose of this study was to explore the clinicopathological characteristics and prognostic significance of SNRPA1 mRNA expression in liver cancer and its potential mechanism as an SF.

Methods: A total of 418 RNA−Seq and clinical data were downloaded from The Cancer Genome Atlas (TCGA) database. Alternative splicing (AS) profiles were downloaded from TCGA SpliceSeq. Wilcoxon rank-sum tests were used to compare normal tissues with tumor tissues and analyze subgroups. We used the Kaplan−Meier analysis method to draw the survival curves. Univariate and multivariate Cox analyses were employed to estimate the prognostic value of SNRPA1. Gene set enrichment analysis (GSEA) was performed to identify the signaling pathways. Then we used univariate and Pearson's correlation tests to analyze the correlation between SFs and exon skip (ES) events. Wilcoxon rank-sum tests were applied to analyze the relationships between different spliceosome and cancers. We also analyzed the relationship between the expression of SCP2 (related to SNRPA1) and clinical symptoms. We then evaluated the expression levels of these genes with clinical samples and The Clinical Proteomic Tumor Analysis Consortium (CPTAC).

Results: The level of SNRPA1 mRNA expression in liver cancer was significantly up−regulated in tumor tissues compared with normal tissues (p = 1.411e−27) in liver cancer and was positively correlated with survival status (p = 0.035). In addition, we found that SNRPA1 mRNA levels can reflect the prognosis of liver cancer (hazard ratio [HR] = 1.08, 95% confidence interval [CI]: 1.02−1.14, p = 0.005). The enriched KEGG pathway by GESA revealed a splicesome as the main pathway of SNRPA1. Alternative splicing events of SNRPA1 was related to the expression of SCP2. SNRPA1 exon 6 skip and SCP2 exon 12 skip correlated with many cancer types. Finally, SNRPA1 and SCP2 percent−spliced−in (PSI) values positively correlated with survival status (p = 3.022e−04 and p = 2.932e−03).

Conclusions: Splicing Factor SNRPA1 is superior to its mRNA expression in predicting the prognosis of liver cancer and correlates with the SCP2 spliceosome, which is consistent with that of SNRPA1 in clinical samples.

Introduction

Liver cancer is the most common malignant tumor accounting for more than 780,000 deaths annually, ranking the second of cancer − related mortality worldwide (1). Similar to other malignant tumors, the pathogenesis of liver cancer is very complicated. The factors contributing to the occurrence of liver cancer include chronic hepatitis virus infection, alcohol, drugs, and genetic factors (2). So far, a surgical operation is still the first choice in treatment of liver cancer. It was reported that the recurrence − free survival (RFS) of patients within 5 years after liver cancer surgery is only 30.8 % − 42.8 % and 5 years overall survival rate (OS) is only 42.9 % − 60 % (3, 4). Even in the early stage of liver cancer, its 5 − year

cumulative recurrence rate was reported to be as high as 57.2%, with the 5 – year overall survival rate of only 76.4% (5). Therefore, it is a clinically important to identify a reliable biomarker that can be used for diagnosis and to predict the prognosis of liver cancer.

Multiple studies have shown that specific alternative splicing (AS) events such as cell proliferation, angiogenesis, tumor metastasis, and immune escape, are associated with the development and progression of cancer (6, 7). There are seven common patterns of AS events: Alternate Acceptor site (AA), Alternate Donor site (AD), Alternate Promoter (AP), Alternate Terminator (AT), Exon Skip (ES), Mutually Exclusive Exons (ME), and Retained Intron (RI) (8). More importantly, specific splicing factors (SFs) could regulate AS alternations (9, 10).

Splicing factor SF3B1 expression is associated with oncogenic splicing variants expression and decreased overall survival (11). Oncofetal splicing factor MBNL3 could promote tumorigenesis and indicates poor prognosis of hepatocellular carcinoma patients. The knockdown of MBNL3 almost completely abolishes hepatocellular carcinoma tumorigenesis (12). Therefore, understanding the roles of splicing factors and splicing events during tumorigenesis would open new avenues for targeted therapies.

Small nuclear ribonucleoprotein polypeptide A (SNRPA1) is a spliceosome component responsible for the processing of pre – mRNA into mRNA. It is a necessary factor for male reproductive ability and spliceosome defects to affect the differentiation of human spermatogonia (13). SNRPA1 regulates the expression of CDK1, PIK3R1, VEGFC, and MKI67 in colorectal cancer. It can be recruited to laser – induced DNA damage sites to prevent R – loop – induced DNA damage (14, 15). In another study, SNRPA1 and TCF7L2 were found to bind to the insertion allele of rs386772267, a genetic insertion which is associated with the increased risk of pancreatic cancer (16).

Although several studies have reported this gene (13–16), the prognostic value and potential mechanism of SNRPA1 as a splicing factor in liver cancer remain unclear. Being an SF related to liver cancer, understanding the function and regulatory genes of SNRPA1 will be of great importance. In this study we explored the SNRPA1 expression in liver cancer and the related signaling pathways through the Gene Set Enrichment Analysis (GSEA). We found that the main pathway of KEGG converged at a spliceosome. Pearson's correlation test indicated that SNRPA1 could modify the exon skip of SCP2. Further analysis revealed that SNRPA1 and SCP2 PSI values are significantly increased in many cancer types, suggesting that a particular type of splices is actively involved in cancer. The use of PSI values of SNRPA1 and SCP2 to predict survival status is significantly better than the use of mRNA expression of these two genes. In brief, our analysis reveals the importance of SNRPA1 in tumorigenesis as a splicing factor in liver cancer.

Methods

2.1 Data acquisition and preprocessing

The Level 3 expression data were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/repository). SNRPA1 expression pattern and its prognostic significance were validated from liver cancer tissues paired with normal liver tissues. We analyzed the AS profiles via TCGA SpliceSeq (http://bioinformatics.mdanderson.org/TCGASpliceSeq), a resource for investigating of cross−tumor and tumor−normal alterations in mRNA splicing patterns of RNA−Seq data (17). The percent−spliced−in (PSI) value of the database indicates the percentage of a transcript element over the total normalized reads for that event with values ranging from 0 to 1 (18). For validation, we downloaded the datasets from The NCI Clinical Proteomic Tumor Analysis Consortium (CPTAC) (https://proteomics.cancer.gov/programs/cptac), aiming at characterizing the protein inventory in tumors by leveraging the latest developments in mass spectrometry-based discovery proteomics (19).

2.2 Clinical sample acquisition

4 paired of tumor and adjacent non-tumor liver tissues were obtained from the Fifth medical center of Chinese PLA general hospital. All the patients underwent primary curative resection and received no prior anticancer treatments. Tissue samples were collected within 30 min after operation and snap-frozen in liquid nitrogen. All experiments were performed in accordance with relevant guidelines and regulations. The study was approved by Fifth medical center of Chinese PLA general hospital, and written informed consent was obtained from each patient.

2.3 Statistical analysis

2.3.1 SNRPA1 expression and clinical analysis

Statistical analyses were performed in R (version 3.6.3) software using R packages: limma, beeswarm, survival, and survminer (20-24). The Wilcoxon rank-sum tests were applied to compare normal tissues with tumor tissues and analyze the difference that included grade, stage, and T classification among subgroups. Then tumor tissues were divided into low and high expression groups by using SNRPA1 expression median level as a cut−off point. Overall Survival (follow−up time > 30 days) were compared between the high and low expression groups via Kaplan−Meier analysis. The independent prognostic value of SNRPA1 expression on liver cancer was assessed by univariate and multivariate Cox analyses. $p < 0.05$ was considered statistically significant.

2.3.2 Gene set enrichment analysis of SNRPA1

Gene Set Enrichment Analysis (GSEA) was performed to identify the signaling pathways related to the regulatory mechanism of SNRPA1 by using the GSEA v4.0.3 software (25). Kyoto Encyclopedia of Gene and Genomes (KEGG) gene sets (c2.cp.kegg.v7.1.symbols.gmt) and Gene Ontology (GO) gene sets (c5.all.v7.1.symbols.gmt) from the Molecular Signatures Database (MSigDB)

(http://www.broad.mit.edu/gsea/msigdb/index.jsp) were utilized to analyze pathways. The tumor tissues were divided into low and high expression groups using SNRPA1 expression median level as a cut−off point. Data sets with a $p < 0.05$ and a false discovery rate (FDR) < 0.25 were considered to be significantly enriched.

2.3.3 Analysis of the role of SNRPA1 as a splicing factor

The ‘upset’ function in the ‘UpSet’ R package was used to visualize the interactive AS events between the seven AS types, to clearly show quantitative results of multiple interactive sets. The prognostic relationship between AS events and OS (follow−up time > 30 days) was performed using the univariate Cox proportional hazards regression model.

2.3.4 Correlation analysis of AS events and SFs

A total of 404 splicing factors were retrieved from the SpliceAid2 database (26). Pearson correlation analysis was performed to explore the interaction and correlation between SFs and significant AS events ($p < 0.05$, OS−related ASs). The screening conditions were a correlation coefficient > 0.6 or < −0.6, and a $p < 0.001$. Finally, we visualized the regulatory networks between SFs and ES events using Cytoscape (version 3.7.2) (27).

2.3.5 Relationships between different splices and cancers

For further analysis of the relationships between different splices and cancers, we downloaded the genes of interest in different cancers on the TCGA SpliceSeq database. They included breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), and stomach adenocarcinoma (STAD). Then the Wilcoxon rank-sum tests were applied to compare normal with tumor tissues in different cancers.

2.3.6 PSI value of SNRPA1, SCP2 and clinical analysis

We used the PSI value of SNRPA1 and SCP2 to analyze the differences in grade, stage, and T classification, among subgroups via the Wilcoxon rank-sum tests. We then drew the survival curves according to PSI value (median level as a cut−off point) via Kaplan−Meier analysis (follow−up time > 30 days).

2.3.7 Clinical analysis of relationship between SCP2 expression and SNRPA1

Applying the methods used in Section 2.3.1 above, we performed the Wilcoxon rank-sum tests to compare the expression of SCP2 in normal and tumor tissues, and analyzed the differences in grade, stage, and T classification among the subgroups. Next, we performed Pearson's correlation test to establish the relationship between SCP2 expression and SNRPA1.

2.3.8 SNRPA1 and SCP2 expression in CPTAC database

To find out whether the genes identified from the TCGA database also are of prognostic significance in protein level, we downloaded and analyzed mass spectrometry-based proteomics data from CPTAC database. The Wilcoxon rank-sum tests were applied to compare normal tissues with tumor tissues. Overall Survival (follow−up time > 30 days) were compared between the high and low expression groups via Kaplan−Meier analysis. $p < 0.05$ was considered statistically significant.

2.4 Dection of protein expression levels of SNRPA1 and SCP2 in clinical samples

We analyzed the proteomics of 4 pairs of tumor tissues and its adjacent non-tumor liver tissues proteomics through 4D Label-free detection techniques: TIMS-TOF Pro mass spectrometry (MS/MS). The results were processed using Maxquant search engine (v1.6.6.0). Retrieval arguments is setting to Homo sapiens 9606 SP 20191115 (20380 sequences). Tandem mass spectra were searched against Human Uniprot database concatenated with reverse decoy database.

Results

3.1 Clinical characteristics

A set of data from 418 patients were downloaded from the TCGA database with corresponding patient demographic and clinical characteristics data including age, gender, histological grade, stage, T/N/M classification and survival status of liver cancer (Table 1). Paired tumor, adjacent non-tumor liver tissues from a cohort of 316 HBV-related HCC patients were downloaded from the current Clinical Proteomic Tumor Analysis Consortium (CPTAC) project (28).

Table 1. Clinical characteristics of the liver cancer patients.

Characteristics		*Number of patients (%)*
Age		
	<55	120(28.71)
	≥55	256(61.24)
	Not available	42(10.05)
Gender		
	Female	146(34.93)
	Male	254(60.77)
	Not available	18(4.31)
Histological grade		
	G1	55(13.16)
	G2	180(43.06)
	G3	124(29.67)
	G4	13(3.11)
	Not available	46(11)
Stage		
	I	194(46.41)
	II	98(23.44)
	III	90(21.53)
	IV	12(2.87)
	Not available	24(5.74)
T Classification		
	T1	204(48.8)
	T2	107(25.6)
	T3	90(21.53)
	T4	14(3.35)
	TX	1(0.24)
	Not available	2(0.48)
M Classification		

	M0	303(72.49)
	M1	8(1.91)
	MX	107(25.6)
N Classification		
	N0	290(69.38)
	N1	8(1.91)
	NX	119(28.47)
	Not available	1(0.24)
Survival status		
	Death	147(35.17)
	Survival	271(64.83)

Not available data and TX data were not used for subsequent analysis

3.2 SNRPA1 is highly expressed in liver cancer

SNRPA1 mRNA expression level was significantly up − regulated in tumor tissues compared with normal tissues (p = 1.411e − 27, Figure. 1A) in liver cancer. A paired comparison between normal and liver cancer tissue from the same patients also showed a significant up − regulation (p = 4.08e − 16, Figure. 1B). SNRPA1 expression level showed a positive correlation with survival status (p = 0.035, Figure. 1C). Furthermore, significant differences were observed in SNRPA1 expression based on histological grade and T classification (Fig. 1D − 1F).

3.3 SNRPA1 is an independent risk factor for evaluation of the survival rate in liver cancer patients

Univariate and multivariate Cox analyses indicated that the SNRPA1 mRNA expression (hazard ratio [HR] = 1.08, 95% confidence interval [CI]: 1.02–1.14, p = 0.005, Table 2) could be a useful biomarker for liver cancer prognosis.

Table 2. Univariate analysis and multivariate analyses of the correlation between SNRPA1 expression and clinical parameters.

	Univariate analysis			*Multivariate analysis*		
Parameters	HR	95%CI	P-value	HR	95%CI	P-value
age	1.01	1–1.03	0.177			
gender	0.82	0.56–1.21	0.317			
grade	1.12	0.87–1.45	0.382			
stage	1.67	1.36–2.06	0.000	1.19	0.51–2.8	0.680
T classification	1.65	1.36–2.01	0.000	1.41	0.63–3.18	0.402
SNRPA1	1.08	1.03–1.14	0.002	1.08	1.02–1.14	0.005

P-values in Bold indicate p < 0.05. HR: hazard ratio; CI: confidence interval.

3.4 SNRPA1 gene set enrichment analysis in liver cancer

To identify the potential mechanisms of SNRPA1 expression on liver cancer prognosis, we conducted the GSEA (GO and KEGG pathway enrichment analysis) between low and high SNRPA1 expression groups (Table 3). The GO and KEGG analyses results showed processes and pathways associated with AS. “RNA splicing”, “small nuclear ribonucleoprotein complex”, “RNA splicing via transesterification reactions”, “spliceosomal complex” and “mRNA processing” were enriched in GO analysis. “Spliceosome”, as well as some carcinogenesis and development associated pathways, like “DNA replication”, “base excision repair”, “RNA degradation” and “cell cycle” were enriched in KEGG analysis. These related results have been shown in Fig. 2. Our results suggested that SNRPA1 could be a useful biomarker and is related to other gene functions through alternative splicing.

Table 3
The top 10 enriched GO and KEGG pathways of high SNRPA1 expression groups

Name	NES	NOM p – val	FDR q – val
GO_POSITIVE_REGULATION_OF_DNA_BIOSYNTHETIC_PROCESS	2.151	0.000	0.124
GO_RNA_SPLICING	2.115	0.000	0.115
GO_SMALL_NUCLEAR_RIBONUCLEOPROTEIN_COMPLEX	2.106	0.000	0.090
GO_U1_SNRNP	2.105	0.000	0.069
GO_NEGATIVE_REGULATION_OF_NUCLEAR_DIVISION	2.088	0.000	0.075
GO_RNA_SPLICING_VIA_TRANSESTERIFICATION_REACTIONS	2.088	0.000	0.063
GO_SM_LIKE_PROTEIN_FAMILY_COMPLEX	2.081	0.000	0.065
GO_TELOMERASE_HOLOENZYME_COMPLEX	2.076	0.000	0.062
GO_SPLICEOSOMAL_COMPLEX	2.063	0.000	0.068
GO_MRNA_PROCESSING	2.053	0.000	0.072
KEGG_SPLICEOSOME	2.062	0.000	0.031
KEGG_DNA_REPLICATION	1.861	0.006	0.223
KEGG_HOMOLOGOUS_RECOMBINATION	1.847	0.004	0.170
KEGG_BASE_EXCISION_REPAIR	1.800	0.008	0.205
KEGG_RNA_DEGRADATION	1.790	0.000	0.179
KEGG_CELL_CYCLE	1.788	0.010	0.154
KEGG_PYRIMIDINE_METABOLISM	1.782	0.004	0.141
KEGG_OOCYTE_MEIOSIS	1.777	0.000	0.130
KEGG_PURINE_METABOLISM	1.706	0.000	0.214
KEGG_NUCLEOTIDE_EXCISION_REPAIR	1.700	0.018	0.201

NES: normalized enrichment score; NOM: nominal; FDR: false discovery rate; p-val: p value

3.5 Alternative splicing profiles of liver cancer in TCGA

By analyzing AS events of 418 cases of liver cancer patients from TCGA, we found 2666 AAs in 1937 genes, 2331 ADs in 1663 genes, 6325 APs in 2566 genes, 8087 ATs in 3532 genes, 12327 ESs in 5331 genes, 137 MEs in 135 genes, and 2263 RIs in 1561 genes. Detailed information about the specific AS types of genes was visualized in the Upset plot (Fig. 3A).

3.6 Analysis of OS – related AS events with univariate Cox

AS events data was used to perform univariate analyses for OS. The results of the univariate Cox proportional hazards regression are shown in **Supplementary Table S1**. Inclusively, 222 AAs, 226 ADs, 618 APs, 891 ATs, 1272 ESs, 16 MEs, and 221 RIs were significantly altered ($p < 0.05$). The Upset plot of significant OS − related AS types was shown in Fig. 3B. Interestingly, SNRPA1 (SNRPA1−32758−ES), belonging to ES events, was significantly ($p = 0.003$) associated with OS in the univariate Cox model (hazard ratio [HR] = 155.76, 95% confidence interval [CI]: 5.26–4612.99). The PSI value of every patient has been shown in **Supplementary Table S2**.

3.7 ES events-related genes interaction networks construction

With access to RNA − seq data and corresponding clinical information of liver cancer patients, we identified 404 candidate SFs whose expression levels were significantly associated with OS related ASs events. Among them, we found that SNRPA1 (COR = 0.606, p = 7.18E − 34), as well as LSM8, PCBP4, LSM2, SNRPB2, NOSIP, SNRPG, SNRPE, SNRPF, SF3A2, PRPF31, EIF2S2, SNRPA, DDX39A, SNRPD1, THOC5, LSM7, ISY1, SRRT, SNU13, RALY, CCDC12, and SNRPB were related to SCP2 (SCP2 − 3045 − ES) (Table 4). We constructed networks between the prognosis associated ES events and survival associated SFs to identify the underlying interactions (Fig. 3C). A total of 48 SFs and 21 AS events were constructed. Ultimately, 69 nodes and 72 edges were established in the PPI networks, which included 14 down − regulated AS events and 7 up − regulated AS events. The SFs and AS events that correlated positively (COR > 0.6) and negatively (COR < − 0.6) were shown by red and blue edges, respectively. From the results, SCP2 could be regulated by 23 SFs.

Table 4
The correlation analysis of AS events and the expression of SFs for SCP2

SF	*AS*	*COR*	*p – value*
LSM8	SCP2 – 3045 – ES	0.64622	1.09E – 39
PCBP4	SCP2 – 3045 – ES	0.606735	5.80E – 34
LSM2	SCP2 – 3045 – ES	0.671344	8.47E – 44
SNRPB2	SCP2 – 3045 – ES	0.621021	6.10E – 36
NOSIP	SCP2 – 3045 – ES	0.62959	3.54E – 37
SNRPG	SCP2 – 3045 – ES	0.633561	9.18E – 38
SNRPE	SCP2 – 3045 – ES	0.600225	4.29E – 33
SNRPF	SCP2 – 3045 – ES	0.702603	1.64E – 49
SF3A2	SCP2 – 3045 – ES	0.626232	1.09E – 36
PRPF31	SCP2 – 3045 – ES	0.630417	2.68E – 37
SNRPA1	SCP2 – 3045 – ES	0.606052	7.18E – 34
EIF2S2	SCP2 – 3045 – ES	0.655976	3.07E – 41
SNRPA	SCP2 – 3045 – ES	0.692363	1.47E – 47
DDX39A	SCP2 – 3045 – ES	0.641328	6.19E – 39
SNRPD1	SCP2 – 3045 – ES	0.716852	2.28E – 52
THOC5	SCP2 – 3045 – ES	0.604718	1.08E – 33
LSM7	SCP2 – 3045 – ES	0.711111	3.39E – 51
ISY1	SCP2 – 3045 – ES	0.748097	2.66E – 59
SRRT	SCP2 – 3045 – ES	0.633006	1.11E – 37
SNU13	SCP2 – 3045 – ES	0.724853	4.71E – 54
RALY	SCP2 – 3045 – ES	0.666096	6.60E – 43
CCDC12	SCP2 – 3045 – ES	0.619867	8.89E – 36
SNRPB	SCP2 – 3045 – ES	0.734261	4.13E – 56

COR: correlation coefficient

3.8 Relationships between different splices and cancer types

The different splices have been shown in Fig. 4A. When the normal and tumor tissues PSI values of liver cancer were compared, it was found SNRPA1 exon 6 skip (SNRPA1_exon_6) and SCP2 exon 12 skip (SCP2_exon_12) correlated with LICH (Fig. 4B). The tumor tissues were significantly different from normal tissues with Wilcoxon rank-sum tests ($p < 0.001$). Further analyses in BRCA, COAD, LUAD and STAD showed the PSI values of SNRPA1_exon_6 and SCP2_exon_12 were significantly increasing in tumor tissues ($p < 0.001$ for each cancer type, Fig. 4C – 4D).

3.9 PSI value of SNRPA1, SCP2 and clinical analysis

SNRPA1 and SCP2 PSI values showed the positive correlations with survival status ($p = 3.022e-04$ and $p = 2.932e-03$, respectively, Fig. 5A – B). Furthermore, the PSI values of SNRPA1 and SCP2 were significant different among subgroups in histological grade ($p = 3.87e-07$ and $p = 1.598e-08$), stage ($p > 0.005$ and $p = 0.007$), and T classification ($p = 0.003$ and $p = 0.002$) (Fig. 5C – H).

3.10 SCP2 is low expressed in liver cancer

On investigating the relationship between SNRPA1 and SCP2 at the gene expression level, we calculated the expression levels of SNRPA1 and SCP2 in tumor tissues and found that they were negatively correlated in clinical tumor samples ($COR = -0.417$, $p = 5.999e-21$; Fig. 6A). SCP2 mRNA expression level was down – regulated ($p = 1.343e-15$; Fig. 6B) in liver cancer, which corresponded to the histological grade and stage of liver cancer (Fig. 6D – F). However, the SCP2 expression level had no significant correlation with survival status ($p = 0.063$; Fig. 6C) using the median level as a cut – off point.

3.11 Validation the SNRPA1 and SCP2 expression at protein level

For the CPTAC analyzed results, SNRPA1 protein expression level was significantly up – regulated in tumor tissues compared with normal tissues ($p = 3.197e-47$, Figure. 7A) in liver cancer. SCP2 protein expression level was significantly down – regulated in tumor tissues compared with normal tissues ($p = 1.083e-23$, Figure. 7B). These two protein expression levels had no significant correlation with survival status ($p = 0.622$, $p = 0.352$; Fig. 7C-D) using the median level as a cut – off point. Our clinical samples results were consistent with the CPTAC database (Fig. 7E-F).

Discussion

Liver cancer is one of the most common cancer worldwide. Although significant improvement in diagnosis and treatment has been witnessed in clinical practice, the prognosis of liver cancer remains considerably unfavorable (29). In recent years, bioinformatics tools have been employed to screen for molecular markers. Identification of such new molecular markers is likely to improve the prognosis and survival rate of liver cancer patients. In our study, we found that high expression of SNRPA1 was associated with histological grade, T classification, and poor survival status in liver cancer. Univariate and multivariate Cox analyses indicated that the SNRPA1 mRNA level might be a useful biomarker for

predicting the prognosis of liver cancer. The enriched GO and KEGG pathway analyses performed using GESA method showed the SNRPA1 is a spliceosome.

Different splicing patterns in one gene produce diverse isoforms. For this reason, the regulation and mechanisms of AS are highly complex in cancer (30). Many AS genes are associated with human cancer. For instance, over-expressed CD44 isoforms (v6–v10) are associated with histological grade, metastasis, and poorer survival in cancer (31, 32). CCDC50 regulates EGFR to affect the phenotype of proliferation and apoptosis (30). AS genes, including the serine/threonine kinase Aurora kinase B (AURKB), the E3 ubiquitin ligase, the p53 − antagonistic protein MDM2, the Cadherin 17 (CDH17) and the p73 tumor suppressors, were shown to be possible oncogenic mechanisms for liver cancer (33).

SNRPA1 is a splicing factor which is responsible for the processing of pre − mRNA into mRNA to act on certain specific AS events. We combined clinical information and AS events in this study to analyze the AS events associated with survival in liver cancer. The univariate Cox analysis result showed that SNRPA1 of ES events could be an independent prognostic factor. Pearson's correlation test indicated that SNRPA1 could modify the exon skip of SCP2.

Sterol carrier protein 2 (SCP2) is well recognized as an intracellular cholesterol trafficking protein that targets cholesterol to cholesterol − rich membrane microstructural domains (34). The expression of SCP2 is related to the progression of gliomas. The down − regulation of SCP2 protein expression could suppress tumor cell proliferation through the autophagy pathway (35). Given that the SCP2 gene is related to SNRPA1, we analyzed clinical information with SCP2. When comparing the normal and tumor tissues PSI value of liver cancer, we found that SNRPA1 exon 6 skip (SNRPA1_exon_6) and SCP2 exon 12 skip (SCP2_exon_12) correlated with LICH, BRCA, COAD, LUAD and STAD. Further analysis revealed that PSI values could have a positive correlation with survival status and portrayed were significant differences among subgroups in histological grade, cancer stage as well as T classification.

Overall, we integrated clinical information with AS events to reveal the underlying mechanism of SNRPA1. Based on our results, SNRPA1 is a potential prognostic marker for liver cancer and may play a role in exon skip of SCP2. The PSI values of SNRPA1_exon_6 and SCP2_exon_12 were significantly increased in many cancer types, suggesting that a particular type of splices is actively involved in cancer. Furthermore, the use of PSI values of SNRPA1 and SCP2 to predict survival status ($p = 3.022e-04$ and $p = 2.932e-03$, respectively) is significantly better than the use of mRNA expression of these two genes ($p = 0.035$ and $p = 0.063$, respectively). The validation of CPTAC database and clincal samples also supported this point of view. Besides, the present study identified numerous SFs that may theoretically influence SCP2. However, experimental validation studies are required to confirm the splices associated with the progress of cancer and the exact pathogenic mechanism in liver cancer. For future research, a large number of paired liver cancer patient tissues will be tested. The RNA-seq from HCC cell lines depleted with SNRPA1 will be performed to analysis the alternative splicing events.

In conclusion, we have provided comprehensive landscape of splicing factor SNRPA1 in patients with liver cancer and identified OS-associated AS events. It might be potentially application value in clinical

practice. In addition, we demonstrated that the prognosis-related AS events could be applied to build predictive factors with high accuracy to stratify survival risk compared to mRNA expression level in liver cancer. Deep-mining analysis of AS patterns and SF might indeed show new oncological drivers and confer some potential insights into carcinogenesis mechanism.

Abbreviations

Small nuclear ribonucleoprotein polypeptide A	SNRPA1
splicing factor	SF
The Cancer Genome Atlas	TCGA
Gene set enrichment analysis	GSEA
exon skip	ES
The Clinical Proteomic Tumor Analysis Consortium	CPTAC
recurrence–free survival	RFS
overall survival rate	OS
alternative splicing	AS
percent–spliced–in	PSI
Kyoto Encyclopedia of Gene and Genomes	KEGG
Gene Ontology	GO
false discovery rate	FDR
breast invasive carcinoma	BRCA
colon adenocarcinoma	COAD
liver hepatocellular carcinoma	LIHC
lung adenocarcinoma	LUAD
stomach adenocarcinoma	STAD

Declarations

6.1 Ethics approval and consent to participate

All experiments were performed in accordance with relevant guidelines and regulations. The study for using human tissues was approved by the Ethics Review Committee of Fifth Medical Center of Chinese PLA General Hospital, and written informed consent was obtained from each patient.

6.2 Consent for publication

Consent for publication not applicable in this article.

6.3 Availability of data and materials

The datasets used or analysed during the current study are available from the TCGA database (https://portal.gdc.cancer.gov/repository) and TCGA SpliceSeq (http://bioinformatics.mdanderson.org/TCGASpliceSeq). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD025790. All datasets also could be obtained from the corresponding authors.

6.4 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6.5 Author Contributions

W.AN supervised and participated in the design of this study. M. LIU conceived, designed and carried out the study, and wrote the initial manuscript. M. LIU and Y. LU carried out literature search, data acquisition and analyses. Y. LU collected and provided important background information. L. WEI, X. WANG, Y. LU, X. JIA and S. WANG provided assistance for data analyses and statistical analyses. W.AN and L.WEI edited, corrected and proofread the full contents of the paper. All authors read and approved the final manuscript.

6.6 Funding

National Science and Technology Major Project during the 13th Five-Year Plan Period (2019ZX09721001-007), National Key R & D Program "Synthetic Biology" key special project, (2018YFA0903000); National Natural Science Foundation General Project, 81672001; National Science Foundation of China 81672467; Medical Big Data and AI R & D Project of General Hospital (2019MBD–025)

6.7 Acknowledgments

We acknowledge the funding by National Science and Technology Major Project during the 13th Five-Year Plan Period, National Key R & D Program "Synthetic Biology" key special project, National Natural Science

Foundation General Project, National Science Foundation of China, and Medical Big Data and AI R & D Project of General Hospital.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 68(6), 394-424 (2018) doi:10.3322/caac.21492
2. M. I. Shariff, I. J. Cox, A. I. Gomaa, S. A. Khan, W. Gedroyc and S. D. Taylor-Robinson: Hepatocellular carcinoma: current trends in worldwide epidemiology, risk factors, diagnosis and therapeutics. *Expert Rev Gastroenterol Hepatol*, 3(4), 353-67 (2009) doi:10.1586/egh.09.35
3. C. Lazzara, G. Navarra, S. Lazzara, A. Barbera, C. Saitta, G. Raimondo, et al.: Does the margin width influence recurrence rate in liver surgery for hepatocellular carcinoma smaller than 5 cm? *Eur Rev Med Pharmacol Sci*, 21(3), 523-529 (2017)
4. C. Li, T. F. Wen, L. N. Yan, W. S. Lu, B. Li, W. T. Wang, et al.: Liver resection versus liver resection plus TACE for patients with hepatocellular carcinoma beyond Milan criteria. *J Surg Res*, 209, 8-16 (2017) doi:10.1016/j.jss.2016.09.054
5. G. A. Margonis, K. Sasaki, N. Andreatos, Y. Nishioka, T. Sugawara, N. Amini, et al.: Prognostic impact of complications after resection of early stage hepatocellular carcinoma. *J Surg Oncol*, 115(7), 791-804 (2017) doi:10.1002/jso.24576
6. J. Munkley, K. Livermore, P. Rajan and D. J. Elliott: RNA splicing and splicing regulator changes in prostate cancer pathology. *Hum Genet*, 136(9), 1143-1154 (2017) doi:10.1007/s00439-017-1792-9
7. S. Oltean and D. O. Bates: Hallmarks of alternative splicing in cancer. *Oncogene*, 33(46), 5311-8 (2014) doi:10.1038/onc.2013.533
8. M. C. Ryan, J. Cleland, R. Kim, W. C. Wong and J. N. Weinstein: SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics*, 28(18), 2385-7 (2012) doi:10.1093/bioinformatics/bts452
9. C. M. Koh, M. Bezzi, D. H. Low, W. X. Ang, S. X. Teo, F. P. Gay, et al.: MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature*, 523(7558), 96-100 (2015) doi:10.1038/nature14351
10. A. Shilo, V. Ben Hur, P. Denichenko, I. Stein, E. Pikarsky, J. Rauch, et al.: Splicing factor hnRNP A2 activates the Ras-MAPK-ERK pathway by controlling A-Raf splicing in hepatocellular carcinoma development. *RNA*, 20(4), 505-15 (2014) doi:10.1261/rna.042259.113
11. J. L. López-Cánovas, M. Del Rio-Moreno, H. García-Fernandez, J. M. Jiménez-Vacas, M. T. Moreno-Montilla, M. E. Sánchez-Frias, et al.: Splicing factor SF3B1 is overexpressed and implicated in the aggressiveness and survival of hepatocellular carcinoma. *Cancer Lett*, 496, 72-83 (2021) doi:10.1016/j.canlet.2020.10.010

12. J. H. Yuan, X. N. Liu, T. T. Wang, W. Pan, Q. F. Tao, W. P. Zhou, et al.: The MBNL3 splicing factor promotes hepatocellular carcinoma by increasing PXN expression through the alternative splicing of lncRNA-PXN-AS1. *Nat Cell Biol*, 19(7), 820-832 (2017) doi:10.1038/ncb3538
13. H. Wu, L. Sun, Y. Wen, Y. Liu, J. Yu, F. Mao, et al.: Major spliceosome defects cause male infertility and are associated with nonobstructive azoospermia in humans. *Proc Natl Acad Sci U S A*, 113(15), 4134-9 (2016) doi:10.1073/pnas.1513682113
14. Q. Zeng, F. Lei, Y. Chang, Z. Gao, Y. Wang, Q. Gao, et al.: An oncogenic gene, SNRPA1, regulates PIK3R1, VEGFC, MKI67, CDK1 and other genes in colorectal cancer. *Biomed Pharmacother*, 117, 109076 (2019) doi:10.1016/j.biopha.2019.109076
15. M. Tanikawa, K. Sanjiv, T. Helleday, P. Herr and O. Mortusewicz: The spliceosome U2 snRNP factors promote genome stability through distinct mechanisms; transcription of repair factors and R-loop processing. *Oncogenesis*, 5(12), e280 (2016) doi:10.1038/oncsis.2016.70
16. J. W. Hoskins, A. Ibrahim, M. A. Emmanuel, S. M. Manmiller, Y. Wu, M. O'Neill, et al.: Functional characterization of a chr13q22.1 pancreatic cancer risk locus reveals long-range interaction and allele-specific effects on DIS3 expression. *Hum Mol Genet*, 25(21), 4726-4738 (2016) doi:10.1093/hmg/ddw300
17. M. Ryan, W. C. Wong, R. Brown, R. Akbani, X. Su, B. Broom, et al.: TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res*, 44(D1), D1018-22 (2016) doi:10.1093/nar/gkv1288
18. S. Schafer, K. Miao, C. C. Benson, M. Heinig, S. A. Cook and N. Hubner: Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr Protoc Hum Genet*, 87, 11 16 1-11 16 14 (2015) doi:10.1002/0471142905.hg1116s87
19. P. A. Rudnick, S. P. Markey, J. Roth, Y. Mirokhin, X. Yan, D. V. Tchekhovskoi, et al.: A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline. *J Proteome Res*, 15(3), 1023-32 (2016) doi:10.1021/acs.jproteome.5b01091
20. R. D. C. J. C. Team: R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*, 14, 12–21 (2009)
21. B. P. Matthew E Ritchie, Di Wu, Yifang Hu, Charity W Law, Wei Shi ,Gordon K Smyth: limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43, e47 (2015)
22. A. Eklund: beeswarm: The Bee Swarm Plot, an Alternative to Stripchart (2016)
23. T. Therneau: A Package for Survival Analysis in S (2015)
24. Alboukadel Kassambara, Marcin Kosinski and P. Biecek: survminer: Drawing Survival Curves using 'ggplot2' (2019)
25. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43), 15545-50 (2005) doi:10.1073/pnas.0506580102

26. F. Piva, M. Giulietti, L. Nocchi and G. Principato: SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics*, 25(9), 1211-3 (2009) doi:10.1093/bioinformatics/btp124
27. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, et al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11), 2498-504 (2003) doi:10.1101/gr.1239303
28. Q. Gao, H. Zhu, L. Dong, W. Shi, R. Chen, Z. Song, et al.: Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell*, 179(2), 561-577.e22 (2019) doi:10.1016/j.cell.2019.08.052
29. S. F. Altekruse, S. J. Henley, J. E. Cucinelli and K. A. McGlynn: Changing hepatocellular carcinoma incidence and liver cancer mortality rates in the United States. *Am J Gastroenterol*, 109(4), 542-53 (2014) doi:10.1038/ajg.2014.11
30. H. Wang, C. Z. Zhang, S. X. Lu, M. F. Zhang, L. L. Liu, R. Z. Luo, 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(1), 179-195 (2019) doi:10.1002/hep.30147
31. E. Karousou, S. Misra, S. Ghatak, K. Dobra, M. Gotte, D. Vigetti, et al.: Roles and targeting of the HAS/hyaluronan/CD44 molecular system in cancer. *Matrix Biol*, 59, 3-22 (2017) doi:10.1016/j.matbio.2016.10.001
32. G. L. Qiao, L. N. Song, Z. F. Deng, Y. Chen and L. J. Ma: Prognostic value of CD44v6 expression in breast cancer: a meta-analysis. *Onco Targets Ther*, 11, 5451-5457 (2018) doi:10.2147/ott.S156101
33. C. Berasain, S. Goni, J. Castillo, M. U. Latasa, J. Prieto and M. A. Avila: Impairment of pre-mRNA splicing in liver disease: mechanisms and consequences. *World J Gastroenterol*, 16(25), 3091-102 (2010) doi:10.3748/wjg.v16.i25.3091
34. N. I. Burgardt, A. R. Gianotti, R. G. Ferreyra and M. R. Ermacora: A structural appraisal of sterol carrier protein 2. *Biochim Biophys Acta Proteins Proteom*, 1865(5), 565-577 (2017) doi:10.1016/j.bbapap.2017.03.002
35. R. Liu, J. Li, T. Zhang, L. Zou, Y. Chen, K. Wang, et al.: Itraconazole suppresses the growth of glioblastoma through induction of autophagy: involvement of abnormal cholesterol trafficking. *Autophagy*, 10(7), 1241-55 (2014) doi:10.4161/auto.28912

Figures

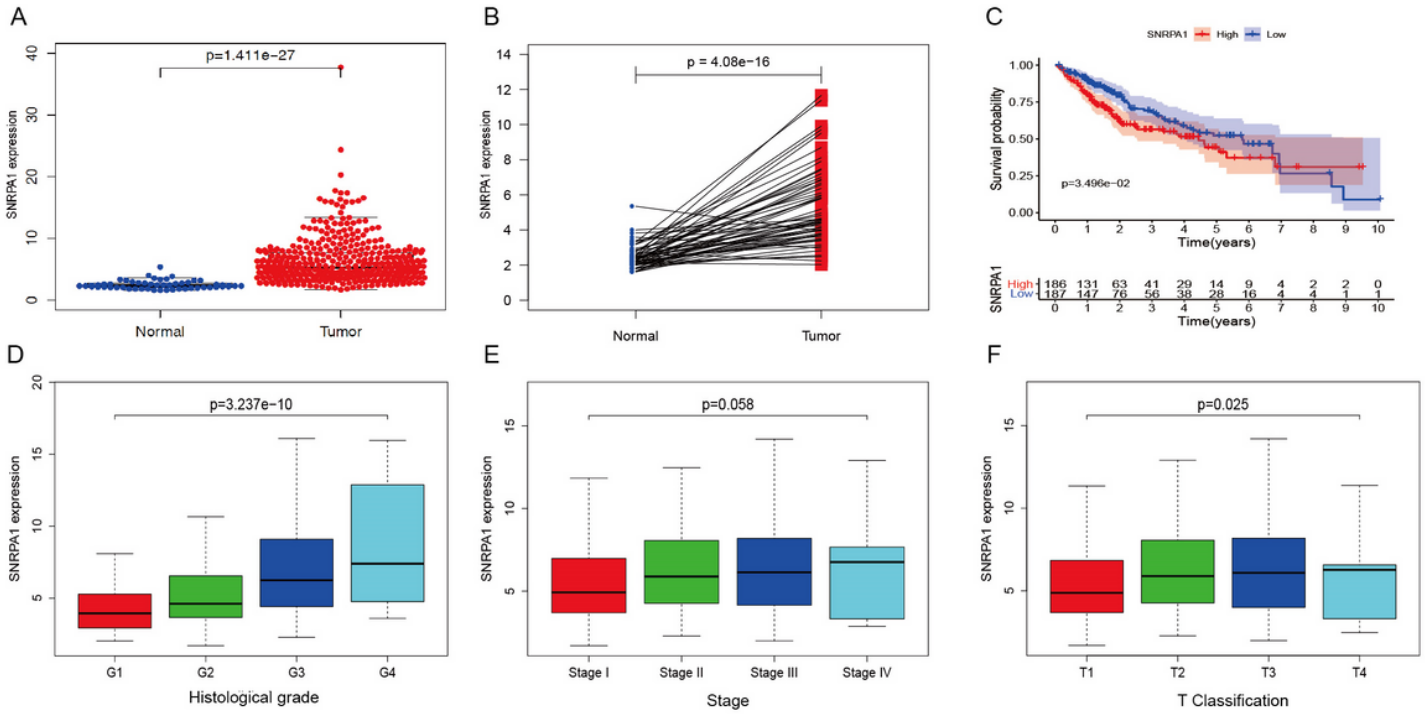

Figure 1

SNRPA1 expression in liver cancer. Comparison of SNRPA1 expression between (A) normal and cancer tissues and (B) paired samples. (C) Kaplan–Meier curves for OS in patients. Comparison of SNRPA1 expression according to clinical parameters: histologic grade (D), stage (E), and T classification (F).

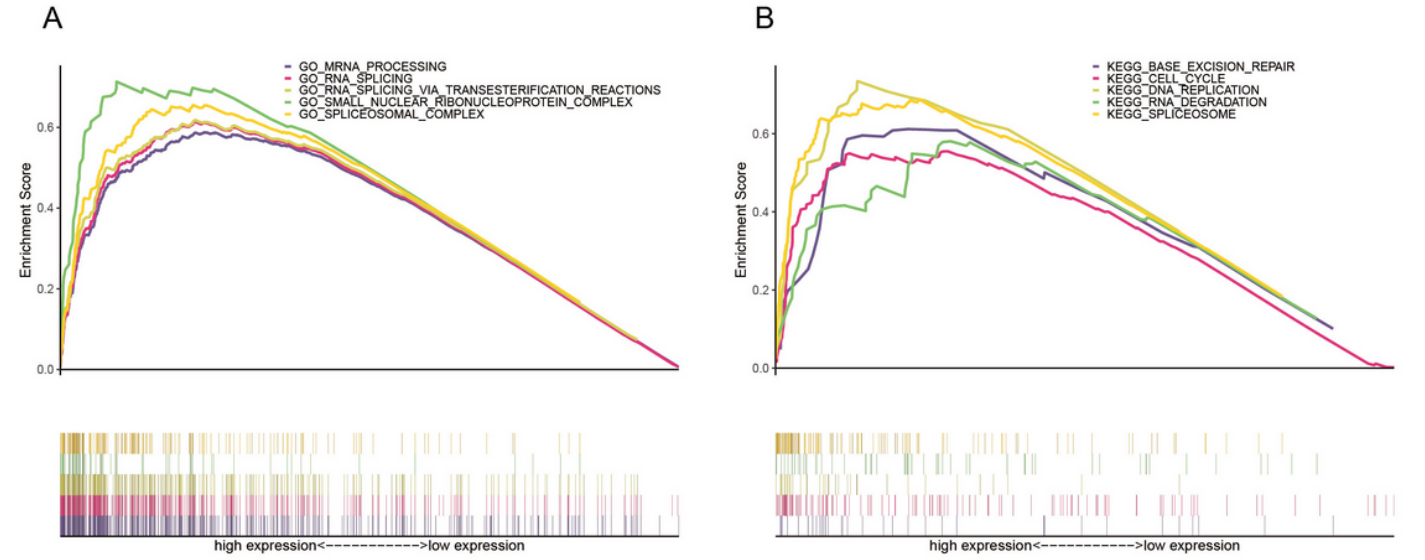

Figure 2

The main enriched GO and KEGG pathways of high SNRPA1 expression sets. The SNRPA1 expression groups of GO (A) and KEGG pathways (B) analyses

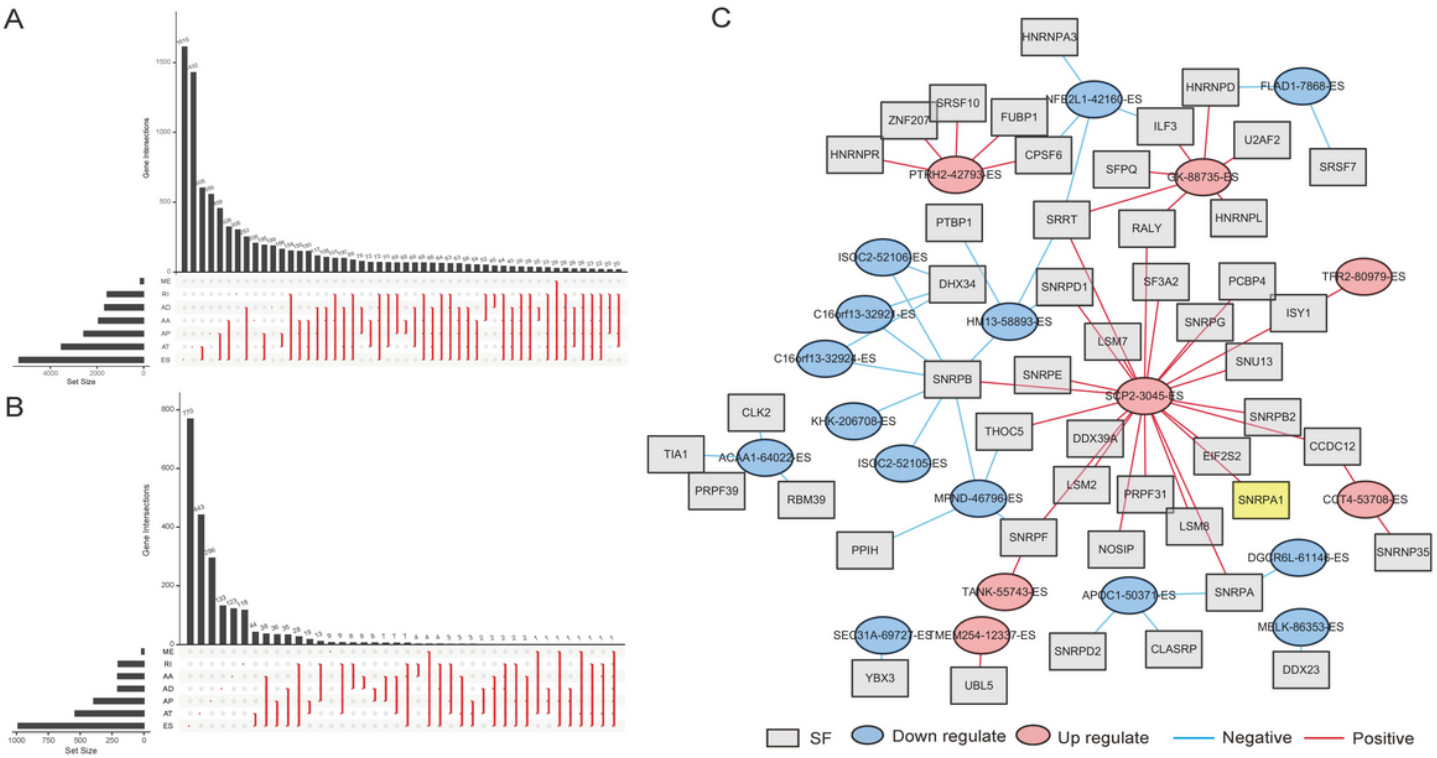

Figure 3

General characteristics of AS and OS–related AS events. (A) The UpSet plot for seven interactions types of AS events in liver cancer, one gene may have up to 6 types of AS. (B) UpSet plot for significant OS–related AS types. (C) The related genes interaction networks of ES events. AA: alternate acceptor site; AD: alternate donor site; AP: alternate promoter; AT: alternate terminator; ES: exon skip; ME: mutually exclusive exons; RI: retained intron.

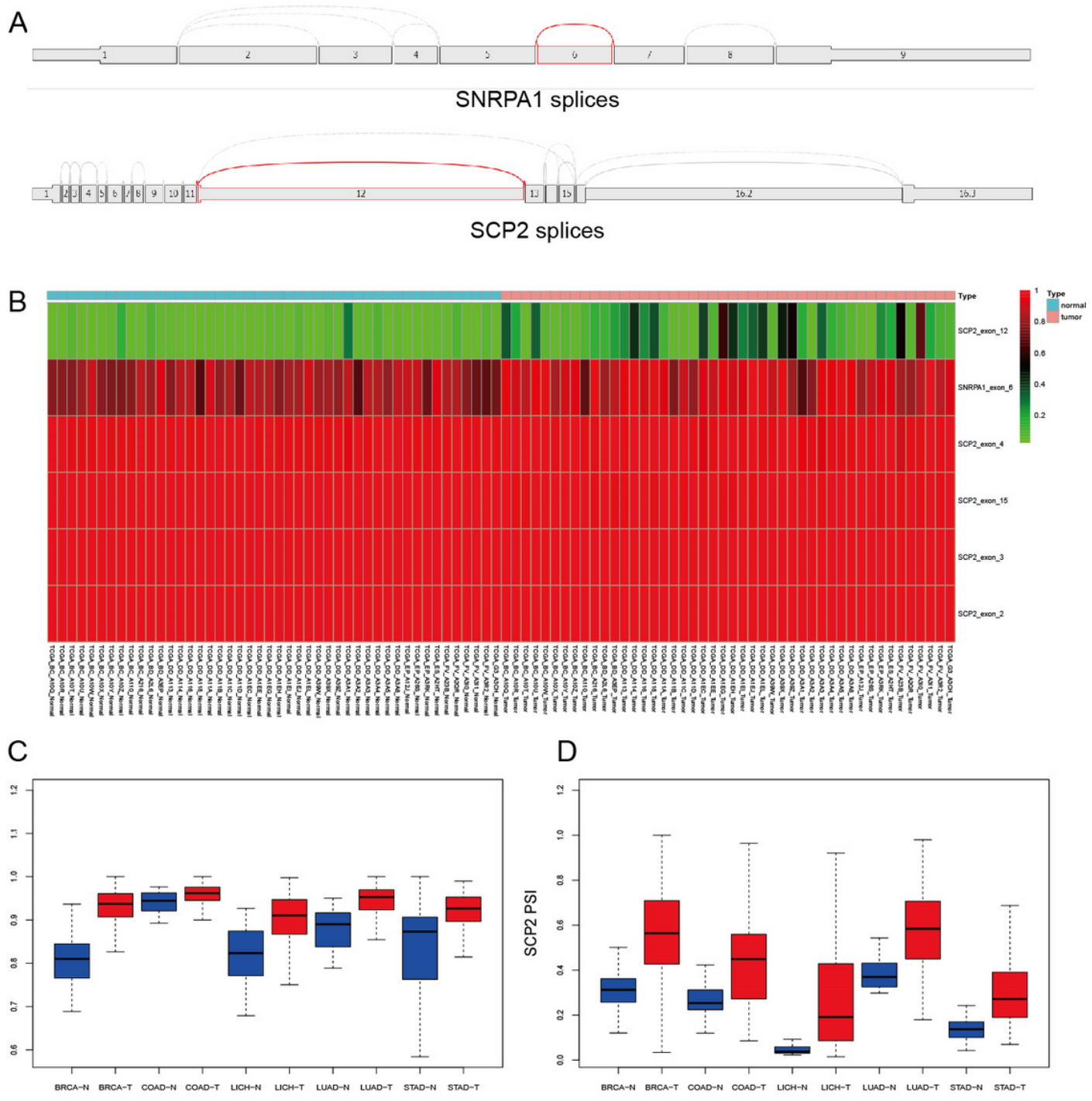

Figure 4

Relationships between different spliceosome and cancers. (A) The SNRPA1 and SCP2 splices. (B) Heatmap of the different spliceosomes in LICH (C) SNRPA1_exon_6 PSI value in different cancers. (D) SCP2_exon_12 PSI value in different cancers. BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; STAD: stomach adenocarcinoma; T: tumor tissues; N: normal tissues.

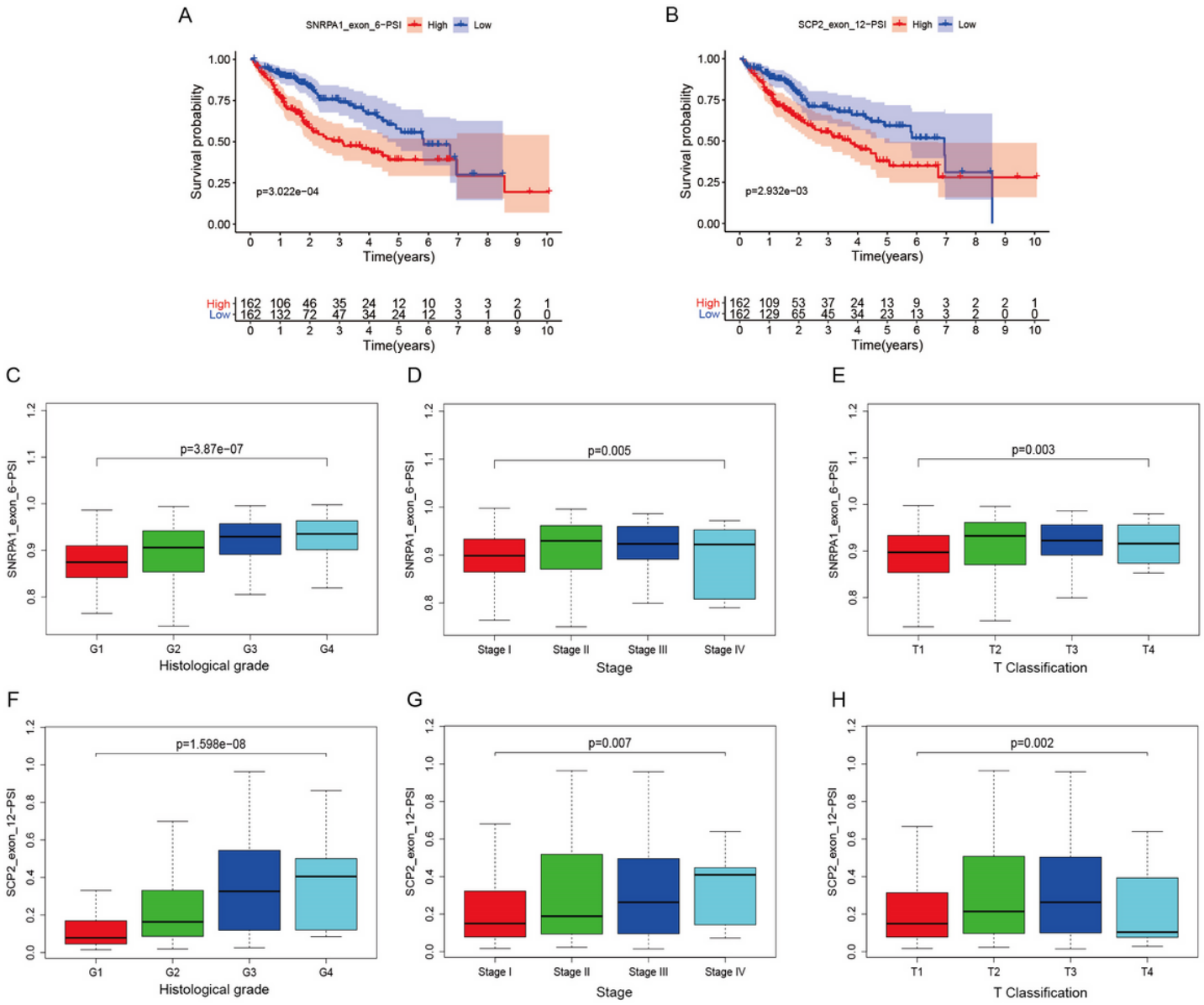

Figure 5

PSI value of SNRPA1 and SCP2 in liver cancer. The high and low PSI value groups of SNRPA1 and SCP2 corresponded with Kaplan–Meier curves for OS in patients (A–B). PSI values of SNRPA1 and SCP2 were compared with clinical parameters: histologic grade,A stage, and T classification (C–H).

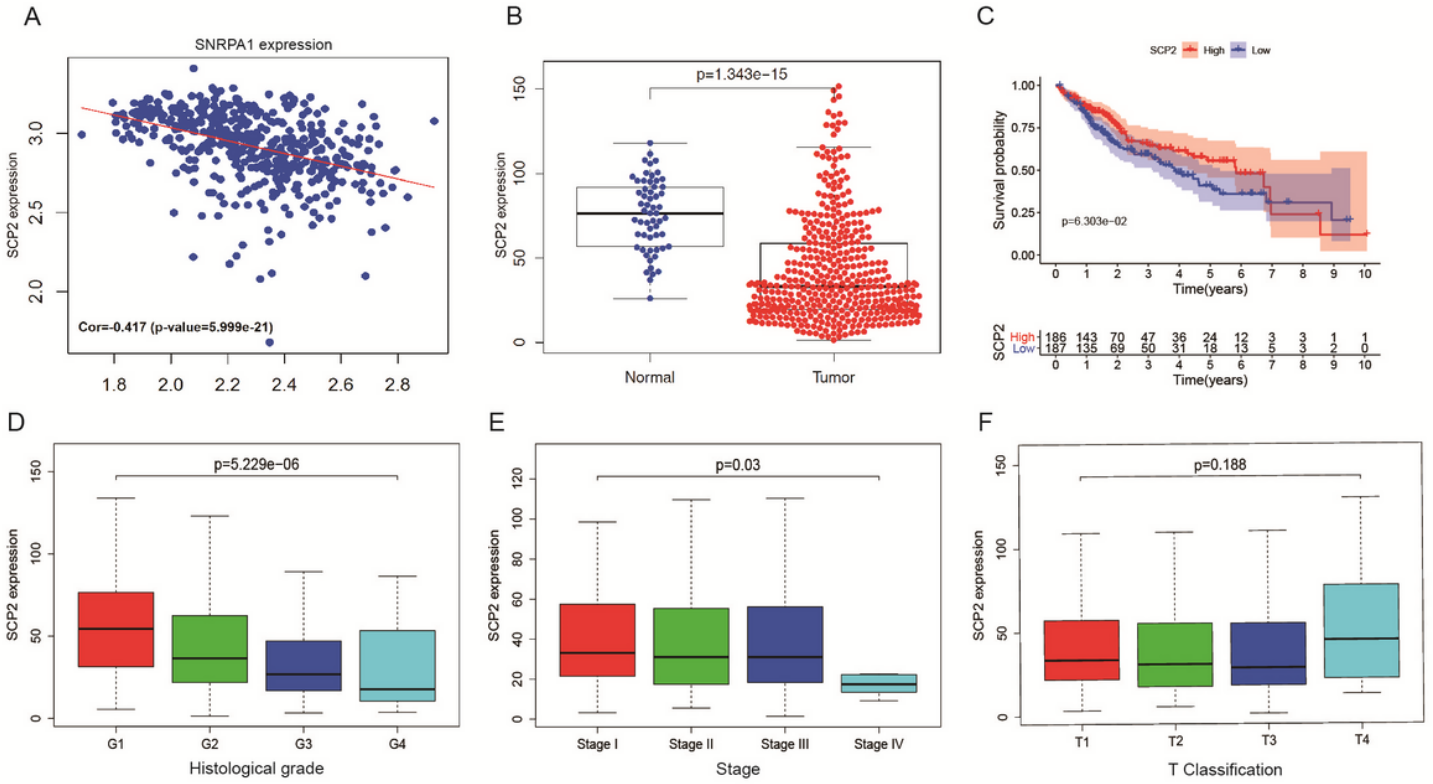

Figure 6

SCP2 expression in liver cancer. (A) The correlation coefficient of SCP2 and SNRPA1 in clinical samples. (B) Comparison of SCP2 expression between normal and cancer tissues. (C) Kaplan–Meier curves for OS in patients. Comparison of SCP2 expressions in different clinical parameters: histologic grade (D), stage (E), and T classification (F).

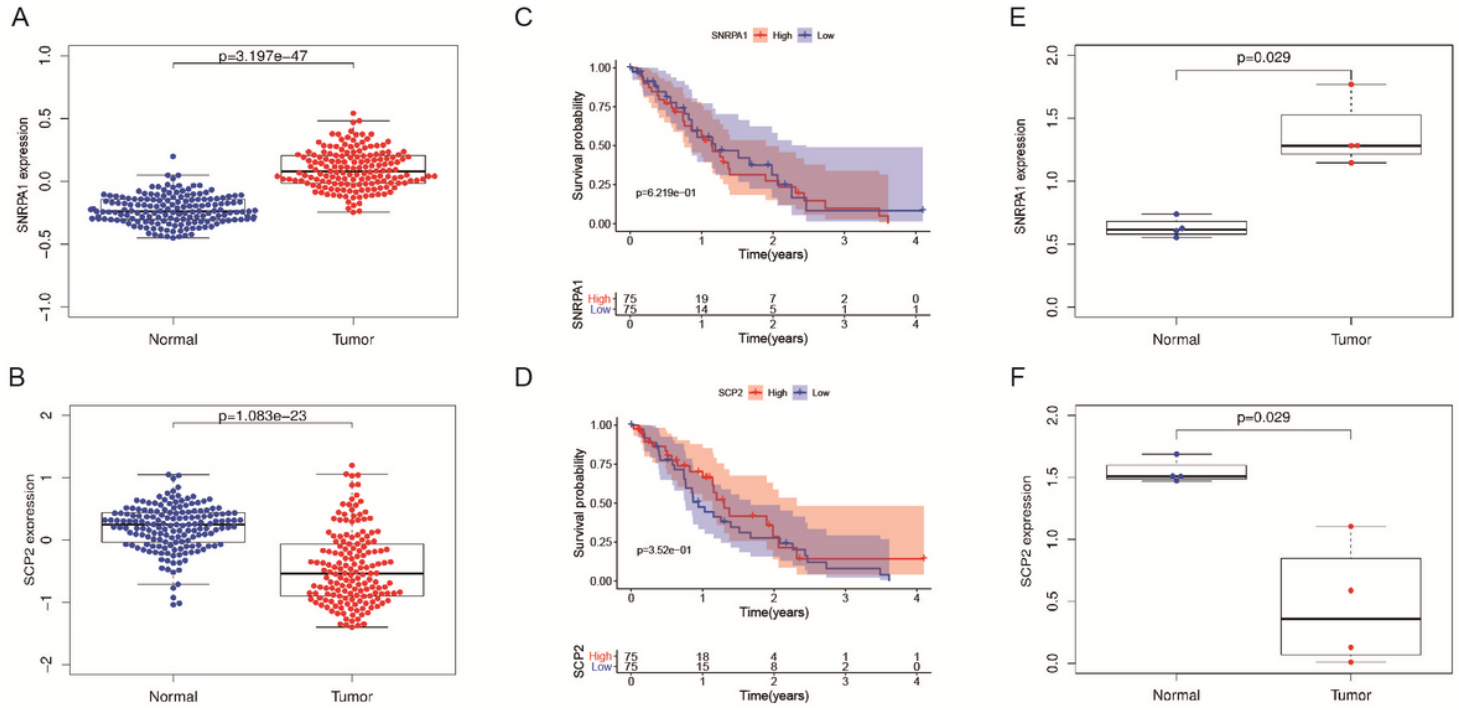

Figure 7

SNRPA1 and SCP2 expression in protein level. (A) and (B), Comparison of SNRPA1 or SCP2 protein expression between normal and cancer tissues in CPTAC database. (C) and (D), SNRPA1 and SCP2 Kaplan–Meier curves for OS in patients. (E) and (F), Comparison of SNRPA1 or SCP2 protein expression between normal and cancer tissues in clinical samples.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary1.xlsx
- Supplementary2.xlsx