



Genome-Wide Profiling of Prognostic Alternative Splicing Pattern in Pancreatic Cancer

Min Yu ^{1*†‡}, Weifeng Hong ^{2†}, Shiye Ruan ^{1,3†}, Renguo Guan ^{1,3†‡}, Lei Tu ¹, Bowen Huang ^{1,3}, Baohua Hou ¹, Zhixiang Jian ¹, Liheng Ma ² and Haosheng Jin ^{1*}

¹ Department of General Surgery, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, ² Department of Medical Imaging, The First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou, China, ³ The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China

OPEN ACCESS

Edited by:

Xiangqian Guo,
Henan University, China

Reviewed by:

Yan-feng Gao,
Zhengzhou University, China
Zhenyu Shi,
Henan University, China

*Correspondence:

Min Yu
yumin@gdph.org.cn
Haosheng Jin
thundercry@163.com

[†]These authors have contributed
equally to this work

#ORCID:

Min Yu
orcid.org/0000-0003-1875-740X
Renguo Guan
orcid.org/0000-0002-9487-7369

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 24 April 2019

Accepted: 31 July 2019

Published: 27 August 2019

Citation:

Yu M, Hong W, Ruan S, Guan R, Tu L,
Huang B, Hou B, Jian Z, Ma L and
Jin H (2019) Genome-Wide Profiling of
Prognostic Alternative Splicing Pattern
in Pancreatic Cancer.
Front. Oncol. 9:773.
doi: 10.3389/fonc.2019.00773

Alternative splicing (AS) has a critical role in tumor progression and prognosis. Our study aimed to investigate pancreatic cancer-specific AS events using RNA-seq data, gaining systematic insights into potential prognostic predictors. We downloaded 10,623 genes with 45,313 pancreatic cancer-specific AS events from the Cancer Genome Atlas (TCGA) and SpliceSeq database. Cox univariate analyses of overall survival suggested there was a remarkable association between 6,711 AS events and overall survival in pancreatic cancer patients ($P < 0.05$). The area under the curves (AUC) of the receiver operator characteristic curves (ROC) of risk score was 0.89 for final prognostic predictor. Results indicated that AS events of DAZAP1, RBM4, ESRP1, QKI, and SF1 were significantly associated with overall survival. The results of FunRich showed that transcription factors KLF7, GABPA, and SP1 were the most highly related to survival-associated AS genes. Furthermore, using DriverDBv2, we identified 13 driver genes associated with survival-associated AS events, including TP53 and CDC27. Thus, we concluded that the aberrant AS patterns in pancreatic cancer patients might serve as prognostic predictors.

Keywords: alternative splicing, TCGA, pancreatic cancer, prognosis, driver gene

INTRODUCTION

During the pre-mRNA splicing, introns are removed, and the exons are left to form the final mRNA products. In this process, exons which are left vary, and thus, one single gene may generate multiple mRNA isoforms by alternative splicing (AS). More than 95% of human genes undergo AS, and most of them vary in levels across different cells and tissues (1). Variations in AS may result in a spectrum of consequences from completely functional inactivation, to subtle or difficult-to-detect effects, or possibly to altering the location, stability or translation of a transcript, including oncogenes and tumor-suppressor genes. Alternative splicing has not only critical roles in normal development but also is indispensable in multiple pathological processes, including cancers (2–4). Previous studies have provided evidence that aberrant splicing patterns are closely related to tumor progression and prognosis (2). For example, alternative splicing in pre-mRNA of Epidermal Growth Factor Receptor (EGFR) produces several isoforms, some of which are constitutively active, leading to enhanced tumorigenicity, migration, and invasion (5, 6). EGFR, Insulin Receptor (INSR), and Vascular Endothelial Growth Factor Receptor (VEGFR), whose alternative splicing features varied, result in promoting tumor progression or reduced response to therapy (7). Recent evidence found that several tumor suppressor genes undergo aberrant AS in cancer, which leads to

either complete or partial loss of function, such as TP53 (8). Therefore, alternative splicing events might be ideal biomarkers for cancer diagnosis and prognosis and even be served as a potential target which might help scientists to discover new drugs.

The conventional molecular method for quantification of AS is a reverse transcription polymerase chain reaction (RT-PCR). There are several other techniques, including expressed sequence tags (ESTs) and splicing-sensitive microarrays, which were invented to identify the connections between genotypes and AS patterns in patients. However, these technologies have low throughput, high noise, or restrained to known splicing events. Powered by high-throughput RNA-seq, the amount of human transcriptome data has grown tremendously over the past decade, and large-scale studies in aberrant AS events at a more fine-grained level are now available. Recent advances in RNA-Seq and related bioinformatics methods allow researchers and clinicians to discover cancer-related AS and further investigate the molecular mechanism.

Pancreatic cancer is still known as one of the most malignant solid tumors whose 5-year survival rate has remained under 8% over the past 30 years. The disease is typically found at a late stage when the resection is impossible. Moreover, a response rate of only one-quarter or less can be expected, and resistance of current chemotherapy, such as gemcitabine, occurred in most of the pancreatic cancer patients. At present, the molecular mechanism of pancreatic cancer development and progression is still unclear. Researches have been undertaken to elucidate the mechanisms of this malignancy, including AS in specific gene transcription (9–11). However, few studies have tried to investigate the prognostic value of AS in pancreatic cancer. Therefore, the present study identified pancreatic cancer-specific AS events by analysis of RNA-seq data downloaded from The Cancer Genome Atlas (TCGA) program, gaining more information about their functions in cancer biology in detail.

MATERIALS AND METHODS

Alternative Splicing Events From TCGA RNA-Seq Data

TCGA (<https://tcga-data.nci.nih.gov/tcga/>) is a landmark cancer genomics program with a large amount of detailed information across various cancers in public database (12). The RNA-Seq data of pancreatic cancer cohorts (PAAD) was downloaded for further analysis. SpliceSeq (<http://bioinformatics.mdanderson.org/TCGASpliceSeq>) is a Java application which explores the mRNA alternative splicing patterns of TCGA data. The SpliceSeq tool was used to investigate the mRNA splicing pattern of PAAD samples from the TCGA database. SpliceSeq aligned reads to available transcripts of genes in the Ensembl database and built a unified splice graph. Then, the PAAD sample reads are aligned to the splice graph, and the feature of splicing for each transcript will be summarized. The Percent Spliced in (PSI) value is a parameter to assess the chance of each splicing event. There are several subtypes of splice events: Exon Skip (ES), Alternate Promoter (AP), Mutually Exclusive Exons (ME), Alternate Terminator

(AT), Retained Intron (RI), Alternate Donor site (AD), and Alternate Acceptor site (AA). The detailed information of each subtype of splicing event in PAAD was shown in **Figure 1A**.

Survival Analysis

Clinical information of the PAAD cohort with 178 patients was available in the TCGA database (12). Summary characteristics of these patients were shown in **Supplementary Table 1**. In order to build the model and further analysis, we used mean values to replace the null value in the dataset of the splicing events. For each AS event, the patients were divided into two groups according to the median value; then the Univariate Cox analyses were performed to identify survival-associated splicing AS events in pancreatic cancer ($P < 0.05$). The Multivariate Cox regression was performed to determine the prognostic value of splicing events ($P < 0.05$). Then, the most significant top 20 genes in each model were chosen for the forest plots. Above analyses were performed using R/Bioconductor (version 3.5.2) and SPSS (version 25.0).

Construction of the Model of Risk Scores

Predictive models were built with prognostic events from identical AS subtype, respectively, whereas the final model was constructed with the whole splicing events from PAAD. In order to evaluate accuracy of model of risk scores, we drew the K-M curve, and the cut-off value is $P < 0.01$. Receiver-operator characteristic (ROC) curves were drawn, and the values of the area under the curves (AUC) were used to compare the predictive power of each model. All analyses were performed using R/Bioconductor (version 3.5.2) and Graphpad Prism 8.0.

UpSet Plot and Gene Network Construction

Intersections between different types of AS were investigated by UpSet R (13). UpSet R is a novel R package which provides intersecting sets using matrix design, along with visualizations of several common sets, element, and attribute related tasks. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were performed and were significant when the P -value was < 0.05 in KEGG and 0.0001 in GO analysis. GO Enrichment plot were used to depicted gene interaction network, function annotation, and pathway enrichment of survival-associated AS genes. Therein, using Cytoscape (version 3.7.1), significant genes with the smallest P -value in univariate analysis were selected for the drawing of the PPI network.

Splicing Correlation Network Construction

The expression of splicing factor genes in mRNA splicing pathway was investigated by analysis of the level 3 mRNA-seq data in TCGA. Pearson correlation test was used to analyze the correlation between the mRNA expression of splicing factor gene and the PSI value of survival-associated alternative splicing events. Cytoscape (version 3.7.1) was used to construct the interaction network of the significant genes with the smallest P -value.

Analysis of Splicing-Factor, Transcription Factors, and Driver Gene

The association between survival-associated AS events and splicing factors was further investigated. Firstly, the log-rank test was used to identify survival-associated splicing factors. The list of 71 known splicing factors was extracted from the SpliceAid 2 (<https://bioinformatics.mdanderson.org>) database, which was released in February 2013 (14). The expression profiles of splicing factors were downloaded from the TCGA database and further converted into transcripts per million (TPM). Pearson correlation test was applied to assess the association between survival-associated AS and survival-associated splicing factors. FunRich (Functional Enrichment analysis tool for transcription factors) from ExoCarta (<http://www.exocarta.org/>), DriverDBv2 (A database for human cancer driver gene research) and David (<http://david.abcc.ncifcrf.gov/>) databases were used to perform the analysis. To find the correlation between gene mutation status and AS events, *t*-test was performed. Pearson correlation test was also performed to investigate the association between mRNA expression of driver genes and AS events. R software (version 3.5.2) was applied for bioinformatics analysis, and $P < 0.05$ was considered significant (Two-sided tests).

RESULTS

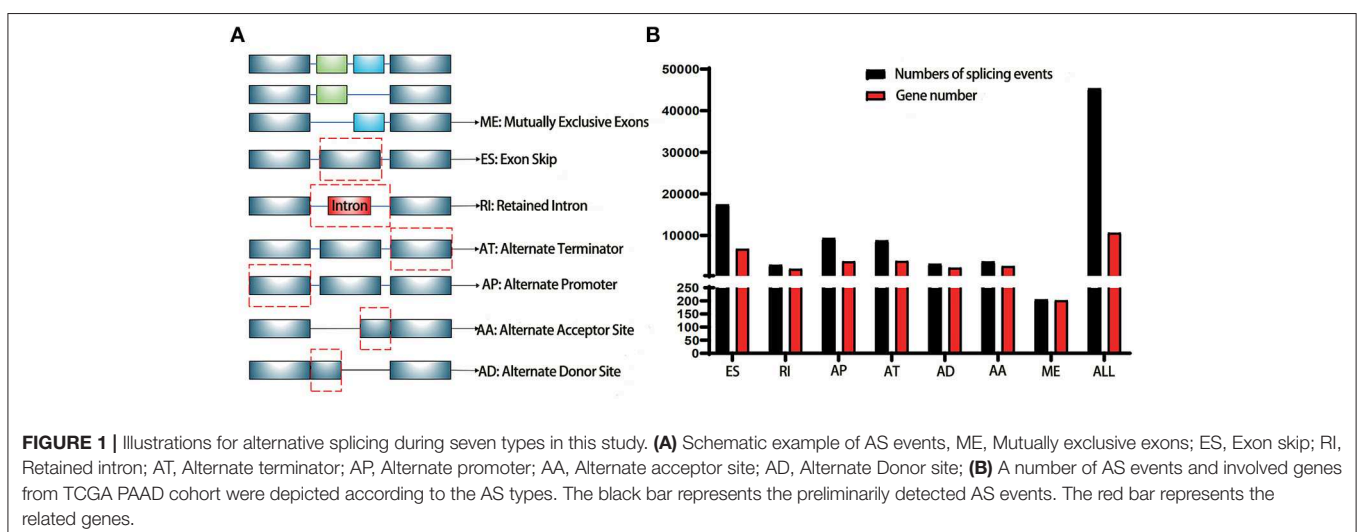
Number of mRNA Splicing Events in PADD Cohort From TCGA

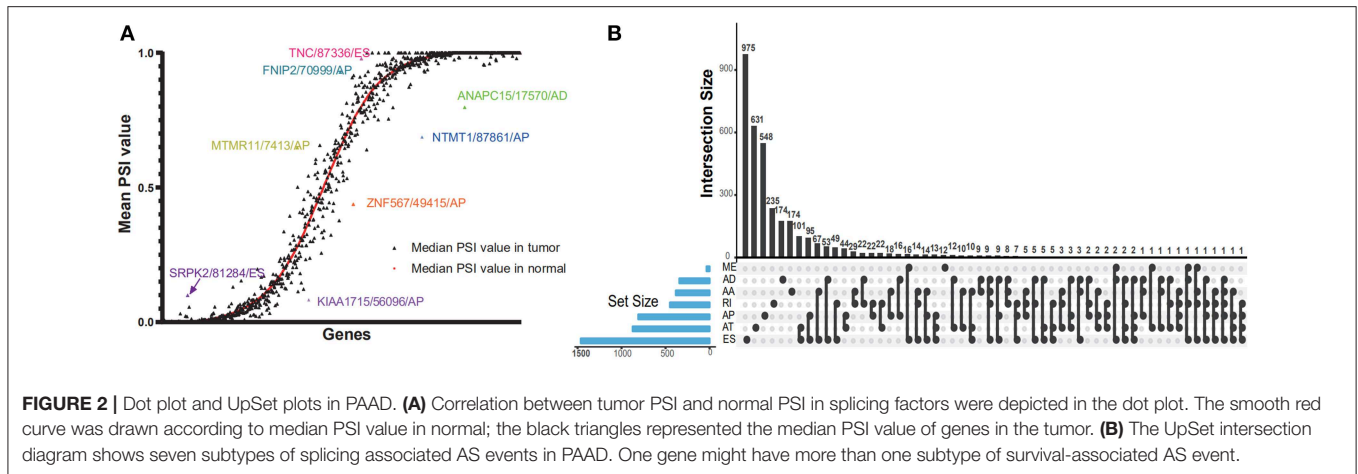
The PSI value of all the splicing events was calculated by SpliceSeq. To identify each AS event precisely, each AS event was named by gene name followed by the unique as_ID and AS types. For example, for the name S100A13/7733/AP, S100A13 is the gene name, 7733 is the as_ID in the dataset, and the AP is the AS subtype. As depicted in **Figure 1B**, a total of 10,623 genes with 45,313 AS events were detected in 178 pancreatic samples, including 17,402 ESs in 6,750 genes, 2,873 RIs in 1,922 genes, 9,325 APs in 3,724 genes, 8,733 ATs in 3,816 genes, 3,118 ADs in

2,210 genes, 3,657 AAs in 2,594 genes, and 205 MEs in 202 genes. Overall results showed that one gene might have an average of 4.2 AS events. Among those genes, 8,833 genes had more than one type of AS events. Gene collagen type 1 alpha 1 (COL1A1) had the maximum number of AS events ($n = 484$), followed by mitochondrial ribosomal protein L55 (MRPL55) ($n = 74$) and interleukin 32 (IL32) ($n = 68$). Among those splicing subtypes, ES was the main subtype of AS events, while ME was relatively rare in the tumor. Besides, only a small proportion of AS events (1,622 out of 45,313) were novel splice. The PADD cohort of TCGA also included four normal samples; the PSI median values of different genes were also summarized and further analyzed. Several genes splicing events, including KIAA1715/56096/AP, ZNF567/49415/AP, NTMT1/87861/AP, ANAPC15/17570/AD, SRPK2/81284/ES, MTMR11/7413/AP, FNIP2/70999/AP, and TNC/87336/ES, differed significantly between tumor and normal samples (**Figure 2A**). When compared to normal samples, cancer samples had reduced alternative splicing diversity (41,629 AS events in normal vs. 40,959 in cancers).

Survival-Associated AS Events in PADD Cohort

Cox univariate analyses of overall survival were applied to explore survival-associated AS events in PADD cohort. The results showed that 6,711 AS events strongly correlated with OS ($P < 0.05$), including 550 RIs from 449 genes, 421 AAs from 382 genes, 385 ADs from 342 genes, 1,499 APs from 809 genes, 1,649 ATs from 873 genes, 2,174 ESs from 1,463 genes, 33 MEs from 33 genes and 550 RIs from 449 genes. The UpSet plot was a novel method to display the intersecting sets, which may be more intuitive and superior to the Venn diagrams. As depicted in the plot, most of these genes had two or more AS subtypes associated with survival, but none of them possessed seven AS subtypes simultaneously (**Figure 2B**). The top 20 survival-associated AS events of the seven AS subtypes were presented in **Figure 3**. In top 300 genes from survival-associated AS events, some genes were top hub genes in the network, such





as VEGFA, CD44, pyruvate kinase gene (PKM), amyloid beta precursor protein (APP), ubiquitin-conjugating enzyme E2 L6 (UBE2L6) (Figure 4A). In pancreatic cancer, KEGG pathway analysis showed that “Metabolic pathway,” “Endocytosis,” and “Axon guidance” were most significantly enriched by these genes. GO analysis revealed that “Protein binding,” “poly(A) RNA binding,” and “RNA binding” in molecular function, “cytoplasm,” “cytosol,” and “extracellular exosome” in cellular component, “cell-cell adhesion,” “mRNA processing,” and “actin cytoskeleton organization” in biological process were the most significantly enriched (Figure 4B).

Prognostic Models for PADD Cohort

To evaluate the prognostic value of AS events in pancreatic cancer, the survival-associated AS events were selected to construct the prognostic risk score models in each subtype of AS events (Figure 5). As depicted in the results, all of the models showed significant value to predict the outcome of pancreatic cancer patients, including RI subtype ($P < 0.0001$), ES subtype ($P < 0.0001$), AP subtype ($P < 0.0001$), AT subtype ($P < 0.0001$), AA subtype ($P < 0.0001$), ME subtype ($P < 0.0001$), and AD subtype ($P < 0.0001$) (Figure 6A). The final prognostic model was built by a combination of prognostic AS events from different subtypes and showed significant prognostic value in distinguishing high-risk patients ($P < 0.0001$). Notably, the final prognostic model showed better performance than seven AS subtypes. The final prognostic predictor had the highest predicting efficiency analyzed by ROC (AUC = 0.89), followed by the AP model in subtypes (AUC = 0.88) (Figure 6B).

Network of Survival-Associated Splicing Factor, Transcription Factors, and Driver Gene

To identify survival-associated splicing factors, we performed a survival analysis about splicing factors based on PSI values. A total of 71 splicing factors from the SpliceAid2 database were chosen for survival analysis. Results showed that AS events of five splicing factors, including DAZ associated protein 1 (DAZAP1), RNA-binding motif 4 (RBM4), Epithelial Splicing Regulatory

Proteins 1 (ESRP1), Quaking (QKI), and steroidogenic factor 1 (SF1), significantly associated with overall survival. The level 3 RNA sequence data were downloaded from TCGA, and the correlations of splicing factors expression and survival were analyzed. As depicted in Figures 7A–E, the expression of ESRP1 ($P = 0.0025$) significantly associated with survival, but DAZAP1 ($P = 0.064$), QKI ($P = 0.45$) and SF1 ($P = 0.62$) and RBM4 ($P = 0.18$) were not. The association between PSI values of top significant AS events and survival-related splicing factors was still unknown. Thus, String tool was used to investigate the association and gain systematic insights into their interaction. Only genes that are significantly related to each other were included in the network. In the correlation network, there was a significant association between the expression of five survival-associated splicing factors and 95 survival-associated AS events. Among 95 survival-associated AS events, 56 AS events (green dots) predicted good survival, whereas 39 AS (red dots) events strongly associated with poor survival in pancreatic cancer (Figure 7F). Correlation between these five splicing factors and representative AS events was shown in dot plots, suggesting the potential association between them (Supplemental Figure 1).

A transcription factor enrichment prediction performed among the survival-associated AS events using the FunRich software. Results identified several transcription factors, including Krüppel-like factor 7 (KLF7), GA binding protein transcription factor subunit alpha (GABPA), trans-acting transcription factor 1 (SP1), that might be the most significant transcription factors associated with survival-associated AS events. Transcription factor SP1 was the most highly related to 53.4% of all the survival-associated AS genes, followed by KLF7 (36.5%) and GABPA (23.9%) (Figure 8A).

A list of driver genes was generated by at least five bioinformatics tools using the DriverDB, which is a database for the investigation of cancer driver gene and mutations. Results showed that 13 driver genes were identified, including tumor protein p53 (TP53), which were previously reported (15) (Figure 8B). In the mutation profile of driver genes, mutation of TP53, FSHD region gene 1 family member B (FRG1B), and cell division cycle 27 (CDC27) occurred in most of PAAD

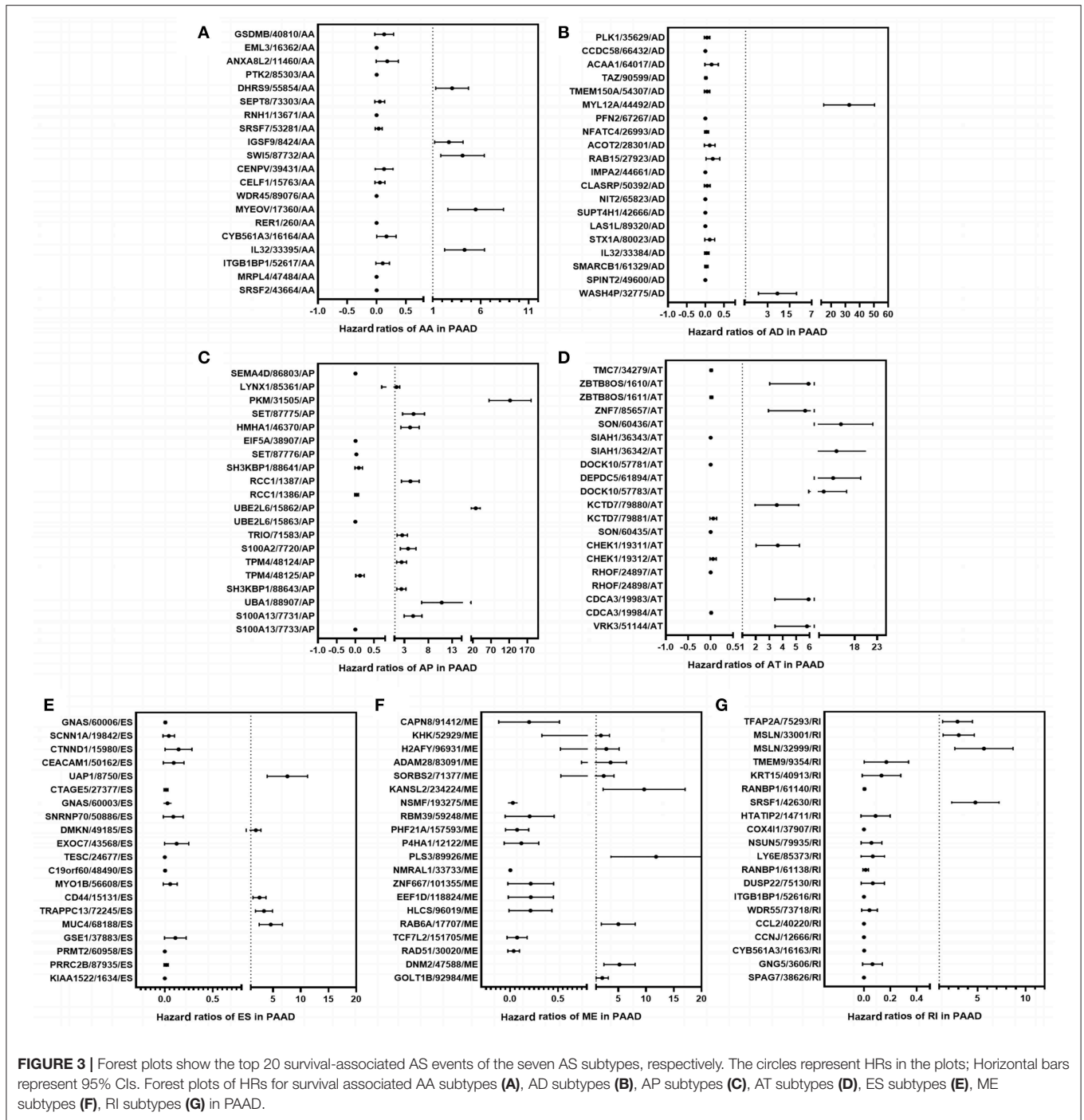
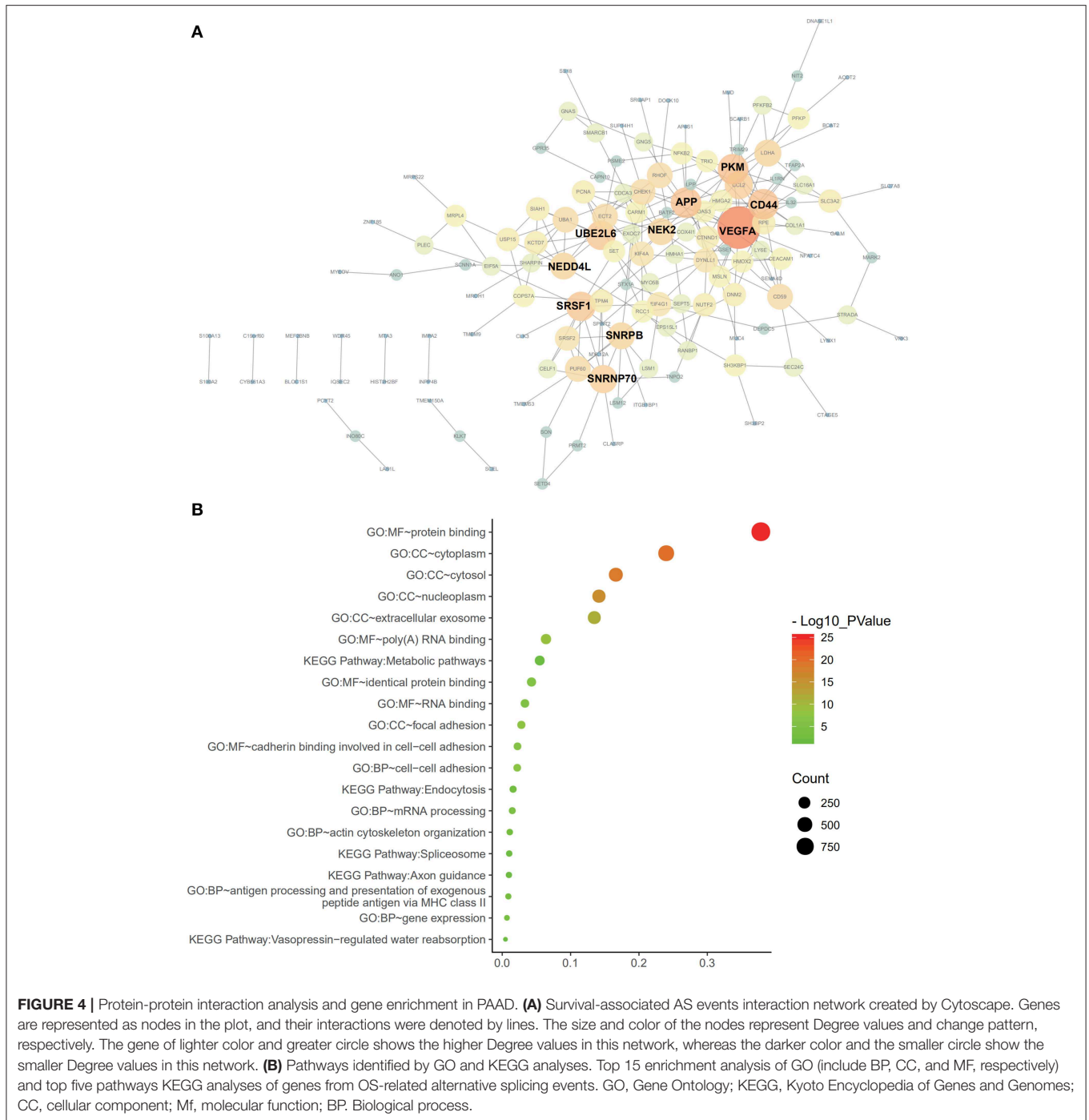


FIGURE 3 | Forest plots show the top 20 survival-associated AS events of the seven AS subtypes, respectively. The circles represent HRs in the plots; Horizontal bars represent 95% CIs. Forest plots of HRs for survival associated AA subtypes (A), AD subtypes (B), AP subtypes (C), AT subtypes (D), ES subtypes (E), ME subtypes (F), RI subtypes (G) in PAAD.

cohort from TCGA. As for the mutation class, truncating and missense were the two main types for driver genes, such as TP53, FRG1B, and CDC27 (Supplemental Figure 2). In addition, we investigated the correlations between mRNA expression of driver genes and the top 30 survival-associated AS events. Results indicated that mRNA expression of adaptor-related protein complex 3 subunit sigma 1 (AP3S1), integrin subunit beta 4 (ITGB4), and p21 (RAC1) activated kinase 1 (PAK1)

was significantly associated with most of the top 30 survival-associated AS events (Supplemental Figure 3). Samples were divided into several groups according to numbers of driver gene mutations, and results indicated that numbers of AS events for each sample were not significantly associated with numbers of driver gene mutations (Supplemental Figure 4). Furthermore, we explored the correlation of AS events and mutation profiles by the *t*-test and found that mutation status of TP53, splicing factor



3a subunit 1 (SF3A1), and CDC27 significantly correlated with most of the Top-100 survival-associated AS events (Figure 9).

DISCUSSION

Alternative splicing enables a single gene to generate multiple mRNAs. Moreover, these mRNAs can be translated into various proteins with diverse functions and structures. Emerging data have demonstrated that aberrant AS patterns were identified

in various cancers and engaged in multiple carcinogenic processes during cancer development and progression (16). The previous study demonstrated the AS events of tissue factors promoted neovascularization and monocyte recruitment via integrin ligation, thus contributing to activation of coagulation and tumor spread in pancreatic cancer (17). In pancreatic cancer, AS events of the PKM were differentially regulated and promoted the expression of the PKM2 isoform. Compared to PKM1, switching PKM2 AS events is beneficial to withstand

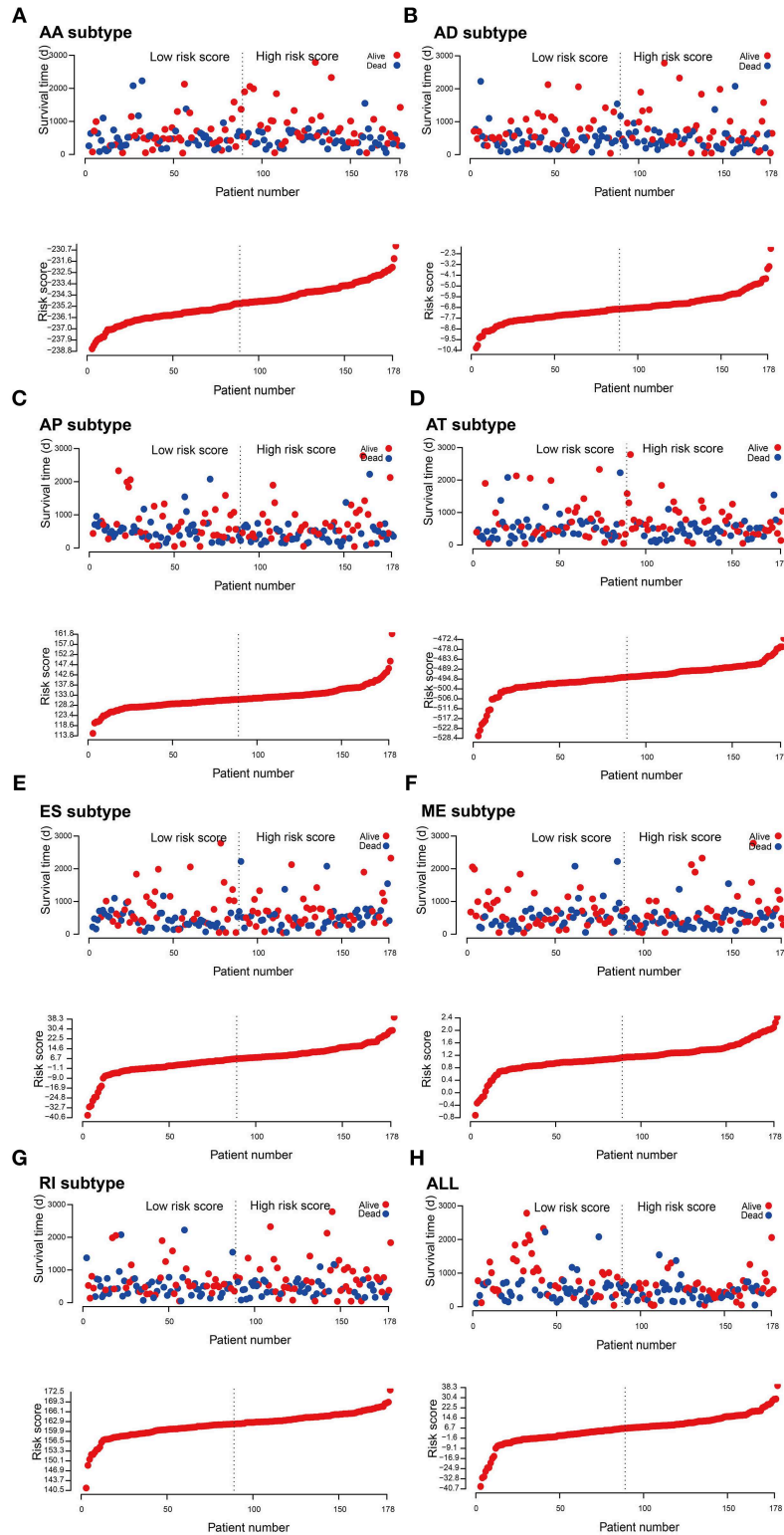


FIGURE 5 | Construction and analysis of risk score based on the survival-associated splicing events using multiple Cox regression analysis. PAAD patients were divided into low- and high-risk groups based on the median value of risk score. The top of each assembly drawing represents survival status and survival time of PAAD patients distributed by risk score, the bottom part is the risk score curve of patients with PAAD. Risk scores were constructed using (A) AA subtypes, (B) AD subtypes, (C) AP subtypes, (D) AT subtypes, (E) ES subtypes, (F) ME subtypes, (G) RI subtypes, and (H) ALL subtypes of survival-associated splicing events.

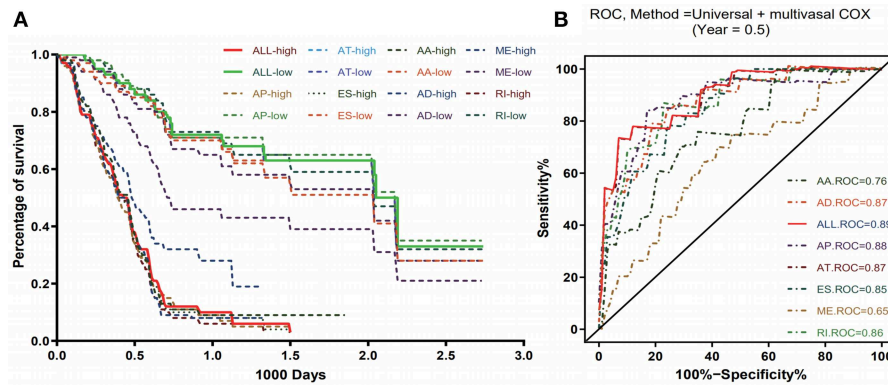


FIGURE 6 | Kaplan-Meier and ROC curves of prognostic predictors in PAAD cohort. **(A)** Kaplan-Meier plot depicting the survival difference between the high and low-risk group in these prognostic models. **(B)** ROC analysis for all prognostic models. The different color lines of ROC curves represent different subtypes of AS events.

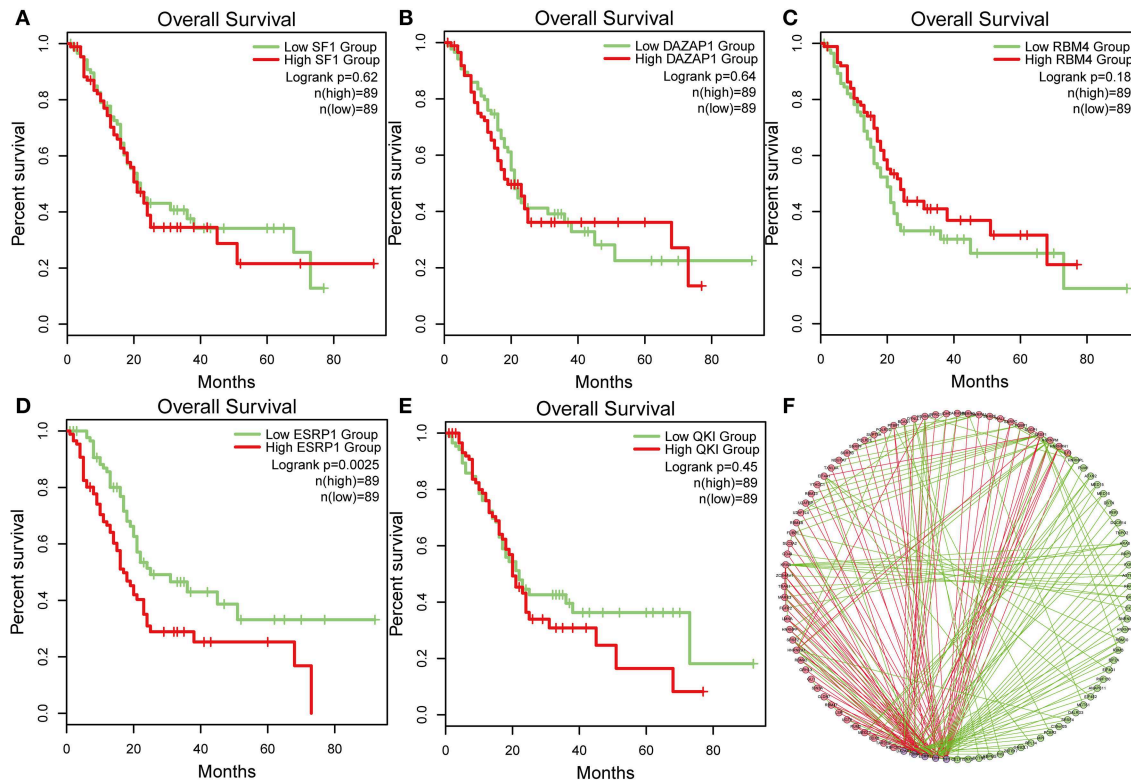
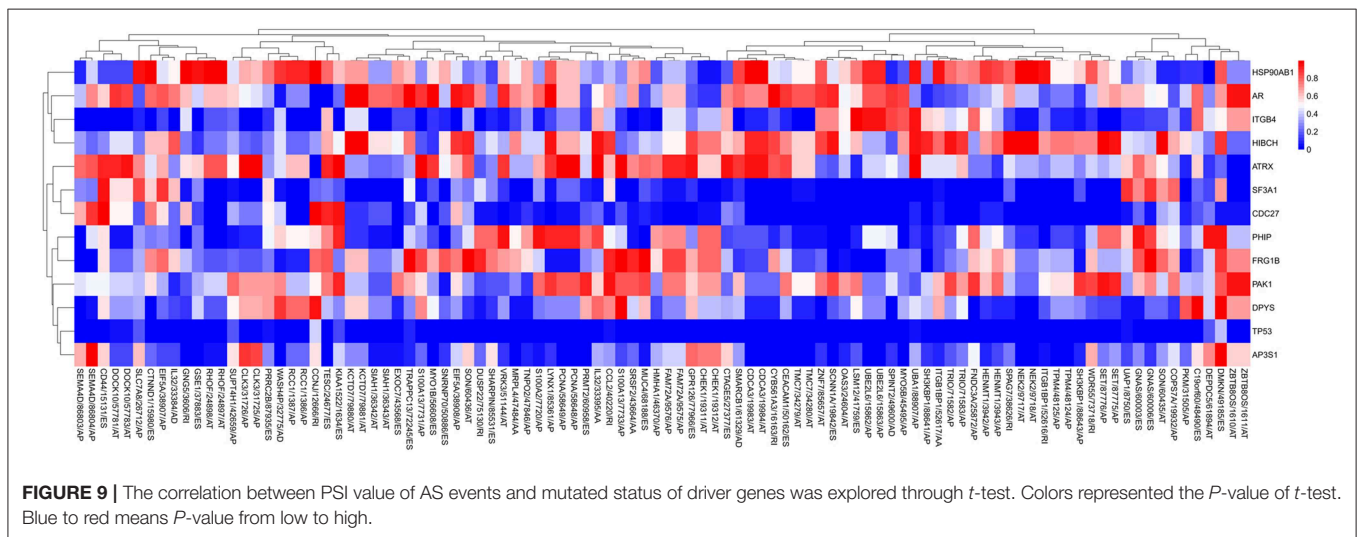
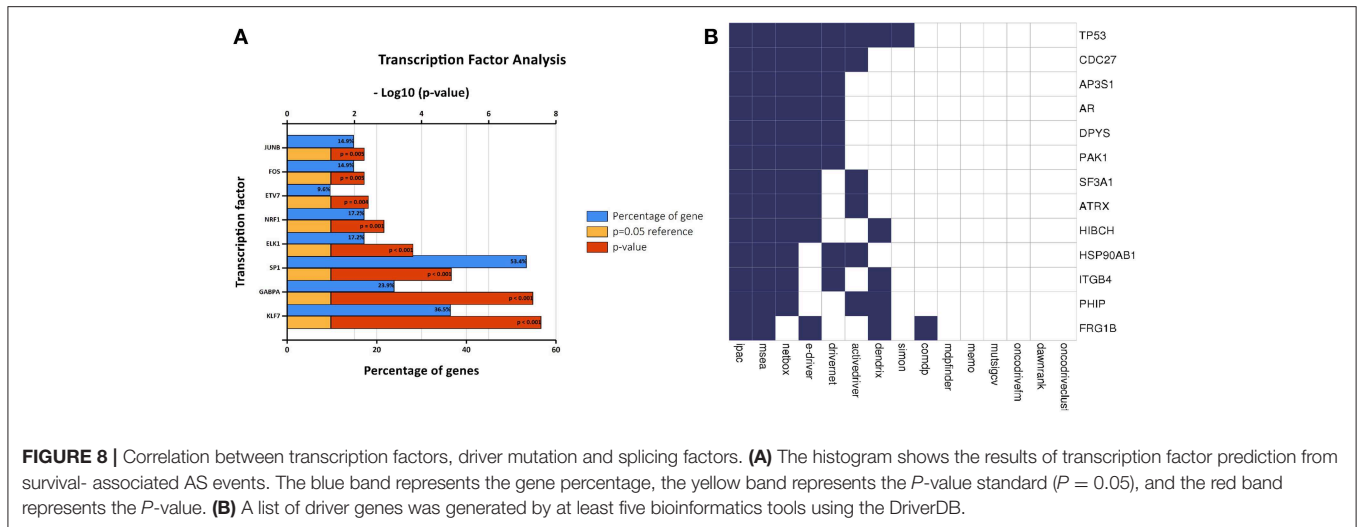


FIGURE 7 | Survival-associated splicing factors and splicing correlation network in PAAD. **(A–E)** The prognostic value of mRNA expression of five splicing factors expression, whose AS events was significantly associated with overall survival in PAAD. **(F)** Splicing correlation network in patients with PAAD constructed by Cytoscape. These five splicing factors (purple dots) were positively (red lines) or negatively (green lines) associated with AS events, which predicted good (green dots) or poor (red dots) outcomes in patients with PAAD.

gemcitabine and cisplatin-induced genotoxic stress, thus induced chemoresistance (18). Serine and arginine-rich splicing factor 1 (SRSF1) and heterogeneous nuclear ribonucleoprotein K (hnRNP K) were aberrantly upregulated in pancreatic cancer, leading to the increased expression of anti-apoptotic splice variants of Bcl-x and Mcl-1, significantly affected responses

to chemotherapy (19). Previous data concerning the function of AS events in pancreatic cancer mainly focused on one or several genes, and there was no study which had explored the prognostic value of AS comprehensively. Given the importance of AS events in cancer, we investigated AS events and gained a comprehensive insight into the prognostic value



of AS events in pancreatic cancer through the analysis of TCGA.

Among the genes with AS events, Gene COL1A1, which makes part of a large molecule called type I collagen, have the maximum number of AS events. Further analysis revealed some of COL1A1 AS events significantly correlated with survival. Our results were consistent with previous studies (20–22). Evidence showed that COL1A1 could activate β 1-integrin and the activation, along with the epithelial-mesenchymal transition, contributed to the development of PAAD (23). The previous study has also demonstrated that once PAAD cells met COL1A1, Snail expression conducted by the increasing of TGF- β 1 (Transforming Growth Factor- β 1) signaling would begin, which in turn accelerate the progress of PAAD invasion by the upregulated MT1-MMP (membrane type 1-MMP) expression (24). Evidence also showed that hypoxia augmented the transcription and deposition of COL1A1 by TGF- β pathway, and COL1A1 was identified

as a hypoxia marker in the non-small cell lung carcinoma (20). Abnormal COL1A1 lead to increasing radioresistance in cervical cancer and had its potential prognostic value in gastric cancer (21, 22). However, the implication of dysregulated splicing pattern of COL1A1 in cancer, including pancreatic cancer with abundance fibrosis, remains to be elucidated. When compared to normal samples, cancer samples had reduced alternative splicing diversity. A previous study reported that the splicing factor genes were upregulated in seven cancer types, including colorectal adenocarcinoma, breast cancer, and lung adenocarcinoma, while they were downregulated in four cancer types, including lymphoma and uterine cancer (2). In our study, we found that the total expression of the splicing factor genes in pancreatic cancer was downregulated. The results indicated that dysregulated expression of the splicing factor genes among cancer types was not in a fixed mode, which may partly result from tumor heterogeneity. Thus, systemic evaluation of the AS patterns in pancreatic cancer contributes

to the understanding of the underlying mechanism of tumor development and progression.

Survival analysis was conducted, and interaction analysis between these survival-associated genes was performed. Results indicated that VEGFA closely related to other genes and served as a hub gene in the network. Among the VEGFA AS events, patients with VEGFA/76330/ES had better survival, implying that loss of Exon8 may weaken or abolish the interaction of VEGFA with other proteins and then inhibit the growth of the tumor. However, VEGFA/76336/ES significantly associated poor survival in pancreatic cancer, which is inconsistent with previous data (25). Of note, VEGFA/76336/ES, whose splice occurred with removal of exon7.1 and exon7.2 loss, lack the neuropilin binding site at exon7. In breast cancer, the VEGF-A/Neuropilin 1 pathway promoted cancer stemness by activating Wnt/ β -Catenin axis, resulting in cancer stem cell phenotypes and chemoresistance (25). In acute myeloid leukemia, high expression of VEGFA was identified as an oncogenic factor, whose function may be reversed by SEMA3A competing for neuropilin (26). Theoretical speaking, removal of exon7, the binding site of neuropilin at VEGF sequence, abolish the interaction and inhibit tumor growth. However, VEGFA/76336/ES significantly associated unfavorable prognosis, which indicating its multifaceted roles in pancreatic cancer progression. It is hard to conclude that VEGFA/76336/ES promotes tumor growth due to a lack of experimental evidence. Nevertheless, our results indicated that neuropilin mediates cancer cell growth may rely on pathways independent of VEGFA. Additionally, blocking neuropilin may strengthen the role of anti-VEGF therapy in reducing the formation of new blood vessels. It is difficult to judge whether a gene is a cancer suppressor or a promoter since different AS events have varied, even opposite biological functions. Therefore, mRNA expression of a gene may be not adequate to determine the biological function, and the predominant AS events need to be taken into account.

Due to the characteristics of pancreatic cancer, including late diagnosis and poor outcome, several researchers had proposed some prognostic models based on mRNA, lncRNA, and microRNA (4, 27, 28). Nevertheless, seldom of these prognostic models come into widely used in clinical practices. Several studies published before finding that alternatively spliced variants contributed to cancer metastasis, cell cycle progression, and chemoresistance (18, 29, 30). As events have been previously identified as diagnostic, predictive, and prognostic biomarkers in pancreatic cancer (18, 31, 32). However, current knowledge about AS events was mostly derived from small samples studies or mainly focused on one single gene. Recently, a systemic analysis of AS events in pancreatic cancer was available due to high-throughput sequencing analysis and data from TCGA. Analysis of each subtype of splicing events was performed and found some of the AS events were of significant prognostic value in pancreatic cancer. Unlike other cancers, including colorectal cancer, lung cancer, the majority of AS events were closely associated with favorable prognosis in pancreatic cancer, especially in AD and RI subtypes. Prediction models were further built by each subtype, respectively or a combination of these seven subtypes. Among the models built by identical

subtype, AP events demonstrated the highest efficiency in the prediction of survival outcome than other six subtypes. Moreover, the final prediction model built by a combination of seven subtypes showed better performance than other prediction models, with an AUC of ROC reaching 0.89 in distinguishing poor survival outcome. Our current work is the first to provide a comprehensive and systemic analysis of AS events and risk score models based on survival-associated AS events in pancreatic cancer.

The network of survival-associated splicing factors was evaluated and found AS events of DAZAP1, RBM4, ESRP1, QKI, and SF1 were significantly associated with overall survival, but the only mRNA expression of ESRP1 correlated with overall survival. Therefore, investigation into the AS events is important to judge the function of gene products. Epithelial-mesenchymal transition (EMT) is defined as a process that epithelial cells with tight junctions acquire a mesenchymal phenotype (33). This means that epithelial cells become easily mobile after this transition, that is, EMT can regulate metastasis (34). ESRP1 is a critical regulator in the epithelial splicing program through targeting several genes, such as fibroblast growth factor receptor 2 (FGFR2) and CD44 (also called H-CAM) (35, 36). As the levels of the mRNA of ESRP1 is down-regulated, the CD44 variant isoform is replaced by the CD44 standard isoform which promotes EMT, increasing invasiveness in gallbladder cancer (37). Evidence showed that the role of inflammation-inducible Snail in the driving malignant transformation of both normal and at-risk human bronchial epithelial cells required the silencing of RNA splice regulator ESRP1 (38). However, the evidence about the function of ESRP1 in pancreatic cancer still lacks and further studies are required. Current evidence has pointed out that splicing factors can precisely bind to a splice-regulatory sequence located at the gene, thus control the process of splicing (39). According to the difference in the sequence and structure, these splicing factors can be divided into two families, including Ser/Arg rich proteins (SR proteins) and the heterogeneous nuclear ribonucleoproteins (hnRNPs). By binding to sequence silencers or enhancers of splicing, these two families possess the opposite function in the mRNA splicing. However, the potential regulatory network of splicing factors during the splicing process remains unclear and clarifying the function of ESRP1 is critical in the interpretation of the molecular mechanism of pancreatic cancer. More attention should thus be paid to the study of AS events in pancreatic cancer.

The transcription process can impact AS events by a variety of mechanisms. Transcription factors can regulate the recruitment of splicing components, and modulate Pol II elongation rate, which regulates the kinetics of exposure of competing for splice sites (40). We evaluated the association between survival-associated AS events and transcription factors. Transcription factors KLF7, GABPA, and SP1, were the most highly related to survival-associated AS genes, which implied that one transcription factor might participate in splicing control of several genes. Krüppel-like factors (KLFs) was involved with many cellular activities, such as proliferation and metabolism (41–43). Moreover, a previous study reported

that KLF7 transcriptionally activated argininosuccinate lyase, which resulted in polyamines production and the oncogenesis of glioma (44). KLF7 can also contribute to the migration and epithelial-mesenchymal transition of oral squamous cell carcinoma (45). However, the mechanism of how transcription factors engaged in the process of splicing is still unknown. It is reasonable that one single transcription factor may regulate several genes not only by direct binding to the promoter of targeted genes but also by indirect impact on splicing process.

Recent evidence showed that several genetic mutations, including K-Ras, TP53, SMAD family member 4 (SMAD4), and cyclin dependent kinase inhibitor 2A/P16 (CDKN2A/P16), drove the oncogenesis of pancreatic cancer (46). Except for these four driver genes, more and more genes are identified as the critical genes in the process of pancreatic cancer, including ret proto-oncogene (RET), AT-rich interaction domain 1A (ARID1A), and ATM (47). Driver genes have been identified as the building blocks in pancreatic cancer, and emerging data suggested that driver gene K-Ras involved in the process of splicing control, such as mucin 6 (MUC6), hepatocyte growth factor (HGF), VEGFR-2, and VEGFB (48). The abnormal expression of splicing factors of SR and hnRNP families results in dysfunction of targeting apoptotic genes, including p53 (19). However, rare studies had been conducted in the exploration of the association between driver genes and AS events. Potential driver genes were identified by the bioinformatic tool in the present study. Further analysis revealed that splicing events of each gene did not increase with accumulating gene mutations. Though the expression of TP53 and SF3A1 correlated with rare survival-associated AS events, mutation status of these two driver genes significantly correlated with many of the top 100 survival-associated AS events. SF3A1, which belong to candidate U2-dependent spliceosome genes family, was identified as driver genes by five prediction tools. Previous studies indicated that two SNPs (rs5994293 and rs9608886) of SF3A1, locating to the region of 22q12.2, were strongly correlated with pancreatic cancer (49). However, the mechanism of how driver genes, including SF3A1, lead to increasing AS events is still unclear. Our study findings enriched our knowledge about the mutation status of driver genes and regulation of splicing, gaining systemic insight into the molecular mechanism underlying PAAD.

Several limitations should be considered when interpreting the results. First, the included number of the PAAD samples was relatively small, and only four normal samples were available for PSI analysis. Second, the prognostic value of survival-associated AS events lack the external independent validation cohort. Third, the present study only investigated the data from high-throughput genomic sequence; experimental validation should be performed in the future.

In conclusion, our comprehensive investigation first focused on the aberrant AS patterns in pancreatic cancer and may contribute to the improvement of pancreatic cancer

management and broaden to the novel field of prognosis and targeted molecular implications.

DATA AVAILABILITY

The original data of the present study can be found at TCGA (<https://tcga-data.nci.nih.gov/tcga/>) and the SpliceAid 2 (<https://bioinformatics.mdanderson.org>).

AUTHOR CONTRIBUTIONS

MY contributed to conception and design, and acquisition, analysis, and interpretation of data. WH contributed to the acquisition of data of acquisition and data analysis. SR contributed to the acquisition of data, analysis, and interpretation of data. RG has been involved in drafting the manuscript and revising it critically. LT contributed significantly to drafting the manuscript. BHU contributed to acquisition of data. BHO contributed to revising the manuscript. ZJ contributed to interpretation of data. LM contributed to data interpretation. HJ conducted the study. All the authors participated in the discussion and editing of the manuscript.

FUNDING

This grant of the study was from the National Natural Science Foundation of China (Grant No. 81701560), National Science Foundation of Guangdong Province, People's Republic of China (Grant No. 2017A030313530) and Guangzhou Science and Technology Plan of Scientific Research Projects, People's Republic of China (Grant No. 201904010021). These fundings made a significant contribution to study design, data interpretation, and writing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2019.00773/full#supplementary-material>

Supplemental Figure 1 | Correlation between these five survival-associated splicing factors and representative AS events was shown in dot plots.

Supplemental Figure 2 | The mutation profile of 13 driver genes. The red band represents truncating, the purple band represents missense, and the green band represents inframe.

Supplemental Figure 3 | The heatmap of the correlations between the mRNA expression of driver genes and PSI values of top 30 survival-associated AS events. Colors represented the correlation coefficient r .

Supplemental Figure 4 | Samples from PAAD cohort were divided into several groups according to numbers of driver gene mutations from 0 to 12 in X-axis. No sample has thirteen gene mutations concurrently. The Y-axis represents the numbers of AS events of each sample.

Supplementary Table 1 | Baseline characteristics according to TCGA Clinical data.

REFERENCES

- Wang ET, Rickard S, Shujun L, Irina K, Lu Z, Christine M, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. (2008) 456:470–6. doi: 10.1038/nature07509
- Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene*. (2015) 35:2413–27. doi: 10.1038/onc.2015.318
- Aversa R, Sorrentino A, Esposito R, Ambrosio MR, Amato A, Zambelli A, et al. Alternative splicing in adhesion- and motility-related genes in breast cancer. *Int J Mol Sci*. (2016) 17:121. doi: 10.3390/ijms17010121
- Shi G, Zhang J, Lu Z, Liu D, Yang W, Wu P, et al. A novel messenger RNA signature as a prognostic biomarker for predicting relapse in pancreatic ductal adenocarcinoma. *Oncotarget*. (2017) 8:110849–60. doi: 10.18632/oncotarget.22861
- Angélique G, Karine D, Hélène RC, Isabelle P, Laura M, Sandrine R, et al. Adult diffuse gliomas produce mRNA transcripts encoding EGFR isoforms lacking a tyrosine kinase domain. *Int J Oncol*. (2012) 40:1142–52. doi: 10.3892/ijo.2011.1287
- Padfield E, Ellis HP, Kurian KM. Current therapeutic advances targeting EGFR and EGFRvIII in glioblastoma. *Front Oncol*. (2015) 5:5. doi: 10.3389/fonc.2015.00005
- Abou Faycal C, Hatat A, Gazzeri S, Eymin B. Splice variants of the RTK family: their role in tumour progression and response to targeted therapy. *J Mol Sci*. (2017) 18:383. doi: 10.3390/ijms18020383
- Surget S, Khoury MP, Bourdon J. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *OncoTargets Therap*. (2013) 7:57–68. doi: 10.2147/OTT.S53876
- Chakedis J, French R, Babicky M, JnqfAQuish D, Howard H, Mose E, et al. A novel protein isoform of the RON tyrosine kinase receptor transforms human pancreatic duct epithelial cells. *Oncogene*. (2016) 35:3249–59. doi: 10.1038/onc.2015.384
- Zheng KL, He TL, Ji WP, Jiang H, Shen Y, Li G, et al. Alternative splicing of NUMB, APP and VEGFA as the features of pancreatic ductal carcinoma. *Int J Clin Exp Pathol*. (2015) 8:6181–91.
- Jiang P, Li Z, Tian F, Li X, Yang J. Fyn/heterogeneous nuclear ribonucleoprotein E1 signaling regulates pancreatic cancer metastasis by affecting the alternative splicing of integrin $\beta 1$. *Int J Oncol*. (2017) 51:169–83. doi: 10.3892/ijo.2017.4018
- Katarzyna T, Patrycja C, Maciej W. The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol*. (2015) 19:68–77. doi: 10.5114/wo.2014.47136
- Alexander L, Nils G, Hendrik S, Romain V, Hanspeter P. UpSet: visualization of intersecting sets. *Vis Comput Grap IEEE Trans*. (2014) 20:1983–92. doi: 10.1109/TVCG.2014.2346248
- Francesco P, Matteo G, Alessandra Ballone B, Giovanni P. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum Mutat*. (2015) 33:81–5. doi: 10.1002/humu.21609
- Minoru O, Keiichi O, Shinobu M, Reiji H, Takashi M, Yasuyuki S, et al. Immunohistochemically detected expression of 3 major genes (CDKN2A/p16, TP53, and SMAD4/DPC4) strongly predicts survival in patients with resectable pancreatic cancer. *Ann Surg*. (2013) 258:336–46. doi: 10.1097/SLA.0b013e3182827a65
- Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet*. (2016) 17:257. doi: 10.1038/nrg.2016.10
- Dusten U, Kevin T, Ramprasad S, Begüm K, Xiaoyang Q, Zhengtao C, et al. Alternatively spliced tissue factor contributes to tumor spread and activation of coagulation in pancreatic ductal adenocarcinoma. *Int J Cancer*. (2014) 134:9. doi: 10.1002/ijc.28327
- Calabretta S, Bielli P, Passacantilli I, Pilozi E, Fendrich V, Capurso G, et al. Modulation of PKM alternative splicing by PTBP1 promotes gemcitabine resistance in pancreatic cancer cells. *Oncogene*. (2016) 35:2031–9. doi: 10.1038/onc.2015.270
- Kedzierska H, Piekietko-Witkowska A. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. *Cancer Lett*. (2017) 396:53. doi: 10.1016/j.canlet.2017.03.013
- Oleksiewicz U, Liloglou T, Tasopoulou K, Daskoulidou N, Gosney JR, Field JK, et al. COL1A1, PRPF40A, and UCP2 correlate with hypoxia markers in non-small cell lung cancer. *J Cancer Res Clin*. (2017) 143:1133–41. doi: 10.1007/s00432-017-2381-y
- Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J Surg Oncol*. (2016) 14:297. doi: 10.1186/s12957-016-1056-5
- Liu S, Liao G, Li G. Regulatory effects of COL1A1 on apoptosis induced by radiation in cervical cancer cells. *Cancer Cell Int*. (2017) 17:73. doi: 10.1186/s12935-017-0443-5
- Duan W, Ma J, Ma Q, Xu Q, Lei J, Han L, et al. The activation of $\beta 1$ -integrin by type I collagen coupling with the hedgehog pathway promotes the epithelial-mesenchymal transition in pancreatic cancer. *Curr Cancer Drug Targets*. (2014) 14:446–57. doi: 10.1158/1538-7445.PANCA2014-A39
- Shields MA, Dangi-Garimella S, Krantz SB, Bentrem DJ, Munshi HG. Pancreatic cancer cells respond to type I collagen by inducing snail expression to promote membrane type 1 matrix metalloproteinase-dependent collagen invasion. *J Biol Chem*. (2011) 286:10495–504. doi: 10.1074/jbc.M110.195628
- Zhang L, Wang H, Li C, Zhao Y, Wu L, Du X, et al. VEGF-A/Neuropilin 1 pathway confers cancer stemness via activating Wnt/ β -catenin axis in breast cancer cells. *Cell Physiol Biochem*. (2017) 44:1251–62. doi: 10.1159/000485455
- Palodetto B, Duarte ADSS, Lopes MR, Corrocher FA, Roversi FM, Niemann FS, et al. SEMA3A partially reverses VEGF effects through binding to neuropilin-1. *Stem Cell Res*. (2017) 22:70. doi: 10.1016/j.scr.2017.05.012
- Zhigang Z, Bing P, Shaocheng L, Zhiwei J, Qian W, Ren L, et al. Integrating MicroRNA expression profiling studies to systematically evaluate the diagnostic value of MicroRNAs in pancreatic cancer and validate their prognostic significance with the cancer genome atlas data. *Cell Physiol Biochem*. (2018) 49:678–95. doi: 10.1159/000493033
- Wu B, Wang K, Fei J, Bao Y, Wang X, Song Z, et al. Novel three - lncRNA signature predicts survival in patients with pancreatic cancer. *Oncol Rep*. (2018) 40:3427–37. doi: 10.3892/or.2018.6761
- Xu Q, Gao J, Li Z. Identification of a novel alternative splicing transcript variant of the suppressor of fused: relationship with lymph node metastasis in pancreatic ductal adenocarcinoma. *Int J Oncol*. 49:2611–9. doi: 10.3892/ijo.2016.3753
- Daoyan W, Liwei W, Masashi K, Zhiliang J, Xiangdong L, Qiang L, et al. KLF4 α up-regulation promotes cell cycle progression and reduces survival time of patients with pancreatic cancer. *Gastroenterology*. (2010) 139:2135–45. doi: 10.1053/j.gastro.2010.08.022
- Ueda J, Matsuda Y, Yamahatsu K, Uchida E, Naito Z, Korc M, et al. Epithelial splicing regulatory Protein 1 is a favorable prognostic factor in pancreatic cancer that attenuates pancreatic metastases. *Oncogene*. (2013) 33:4485–95. doi: 10.1038/onc.2013.392
- Arafat H, Lazar M, Salem K, Chipitsyna G, Gong Q, Pan TC, et al. Tumor-specific expression and alternative splicing of the COL6A3 gene in pancreatic cancer. *J Surg Res*. (2011) 165:176. doi: 10.1016/j.jss.2010.11.782
- Gottgens EL, Span PN, Zegers MMP. Roles and regulation of epithelial splicing regulatory proteins 1 and 2 in epithelial-mesenchymal transition. *Int Rev Cel Mol Bio*. (2016) 327:163–94. doi: 10.1016/bs.ircmb.2016.06.003
- Tsai JH, Jing Y. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev*. (2013) 27:2192–206. doi: 10.1101/gad.225334.113
- Warzecha CC, Shen S, Xing Y, Carstens RP. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol*. (2009) 6:546–62. doi: 10.4161/rna.6.5.9606
- Francesca DM, Pierluigi I, Aaron B, Marcella M, Irene T, Letizia P, et al. Splicing program of human MENA produces a previously undescribed isoform associated with invasive, mesenchymal-like breast tumors. *Proc Natl Acad Sci USA*. (2012) 109:19280–5. doi: 10.1073/pnas.1214394109
- Miwa T, Nagata T, Kojima H, Sekine S, Okumura T. Isoform switch of CD44 induces different chemotactic and tumorigenic ability in gallbladder cancer. *Int J Oncol*. (2017) 51:771–80. doi: 10.3892/ijo.2017.4063
- Walser TC, Jing Z, Tran LM, Lin YQ, Yakobian N, Wang G, et al. Silencing the snail-dependent RNA splice regulator ESRP1 drives malignant transformation of human pulmonary epithelial cells. *Cancer Res*. (2018) 78:1986–99. doi: 10.1158/0008-5472.CAN-17-0315

39. Zong Z, Li H, Yi C, Ying H, Zhu Z, Wang H. Genome-wide profiling of prognostic alternative splicing signature in colorectal cancer. *Front Oncol.* (2018) 8:537. doi: 10.3389/fonc.2018.00537
40. Han H, Braunschweig U, Gonatopoulos-Pournatzis T, Weatheritt RJ, Hirsch CL, Ha KCH, et al. Multilayered control of alternative splicing regulatory networks by transcription factors. *Mol Cell.* (2017) 65:539–53. doi: 10.1016/j.molcel.2017.01.011
41. Sue N, Jack BH, Eaton SA, Pearson RC, Funnell AP, Turner J, et al. Targeted disruption of the basic Kruppel-like factor gene (Klf3) reveals a role in adipogenesis. *Mol Cell Biol.* (2008) 28:3967–78. doi: 10.1128/MCB.01942-07
42. Dong J, Chen C. Essential role of KLF5 transcription factor in cell proliferation and differentiation and its implications for human diseases. *Cell Mol Life Sci.* (2009) 66:2691. doi: 10.1007/s00018-009-0045-z
43. Fan L, Hsieh PN, Sweet DR, Jain MK. Krüppel-like factor 15: regulator of BCAA metabolism and circadian protein rhythmicity. *Pharmacol Res.* (2018) 130:S1043661817314986. doi: 10.1016/j.phrs.2017.12.018
44. Guan F, Kang Z, Zhang J, Xue N, Yin H, Wang L, et al. KLF7 promotes polyamine biosynthesis and glioma development through transcriptionally activating ASL. *Biochem Bioph Res Commun.* (2019) 514:51–7. doi: 10.1016/j.bbrc.2019.04.120
45. Ding X, Wang X, Gong Y, Ruan H, Sun Y, Yu Y. KLF7 overexpression in human oral squamous cell carcinoma promotes migration and epithelial-mesenchymal transition. *Oncol Lett.* (2017) 13:2281–9. doi: 10.3892/ol.2017.5734
46. Shin SH, Cheol Kim S, Hong S, Kim YH, Song KB, Park K, et al. Genetic alterations of K-ras, p53, c-erbB-2, and DPC4 in pancreatic ductal adenocarcinoma and their correlation with patient survival. *Pancreas.* (2013) 42:216–22. doi: 10.1097/MPA.0b013e31825b6ab0
47. Makohonmoore AP, Zhang M, Reiter JG, Bozic I, Allen B, Kundu D, et al. Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat Genet.* (2017) 49:358–66. doi: 10.1038/ng.3764
48. Bittoni A, Piva F, Santoni M, Andrikou K, Conti A, Loretelli C, et al. KRAS mutation status is associated with specific pattern of genes expression in pancreatic adenocarcinoma. *Future Oncol.* (2015) 11:1905–17. doi: 10.2217/fon.15.98
49. Tian J, Liu Y, Zhu B, Tian Y, Zhong R, Chen W, et al. SF3A1 and pancreatic cancer: new evidence for the association of the spliceosome and cancer. *Oncotarget.* (2015) 6:37750–7. doi: 10.18632/oncotarget.5647

Conflict of Interest Statement: 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.

Copyright © 2019 Yu, Hong, Ruan, Guan, Tu, Huang, Hou, Jian, Ma and Jin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.