

Prognostic significance of survival-associated alternative splicing events in gastric cancer

Shichao Zhang^{1,*}, Zuquan Hu^{1,2,*}, Yingwu Lan¹, Jinhua Long³, Yun Wang¹, Xiaowen Chen¹, Xiaofeng Xu⁴, Zhu Zeng^{1,2}, Yan Ouyang¹

¹Immune Cells and Antibody Engineering Research Center of Guizhou Province, Key Laboratory of Biology and Medical Engineering, School of Biology and Engineering/School of Basic Medical Sciences, Guizhou Medical University, Guiyang 550025, Guizhou, P.R. China

²Key Laboratory of Environmental Pollution Monitoring and Disease Control, Ministry of Education of China, Guiyang 550025, Guizhou, P.R. China

³Affiliated Tumor Hospital, Guizhou Medical University, Guiyang 550025, Guizhou, P.R. China

⁴The Clinical Laboratory of Beijing Tongren Hospital, Capital Medical University, Beijing 100730, P.R. China

*Equal contribution

Correspondence to: Yan Ouyang, Zhu Zeng, Xiaofeng Xu; **email:** ouyangyan@gmc.edu.cn, zengzhu@gmc.edu.cn, xuxiaofeng1978@sohu.com

Keywords: gastric cancer, alternative splicing events, prognostic signature, survival, splicing factors

Received: March 4, 2020

Accepted: August 17, 2020

Published: November 7, 2020

Copyright: © 2020 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Alternative splicing events are a major source of transcript and protein diversity in eukaryotes. Aberrant alternative splicing events have been increasingly reported in various cancers, including gastric cancer. To further explore the prognostic significance of alternative splicing events in gastric cancer patients, a comprehensive and systematic investigation was conducted by integrating alternative splicing event data and clinical information. Univariate Cox regression analysis identified 1383 alternative splicing events to be significantly associated with the overall survival of gastric cancer patients. Then, least absolute shrinkage and selection operator (LASSO) and multivariate Cox analyses were performed for the development of prognostic signatures. The final prognostic signature based on all seven types of alternative splicing events can act as an independent prognostic indicator after multivariate adjustment of several clinical parameters. Furthermore, the correlation and function analysis identified *CEL2*, *BAG2*, *RBFOX2*, *PTBP2* and *QKI* as hub splicing factors, and the focal adhesion signaling pathway was most significantly correlated with survival-associated alternative splicing events. The results of this study may establish a foundation for further research investigating the underlying mechanism of alternative splicing events in the progression of gastric cancer.

INTRODUCTION

Alternative splicing (AS) can edit a single pre-mRNA molecule and produce diverse mature mRNAs in eukaryotic organisms. These transcript variants can subsequently generate proteins with different structures and biological functions. Therefore, AS is an important mechanism for posttranscriptional

regulation of gene expression and plays a vital role in the diversification of both the transcriptome and the encoded proteome [1]. Generally, there are seven main patterns of AS events, such as exon skip (ES), retained intron (RI), alternate donor site (AD), alternate acceptor site (AA), alternate promoter (AP), alternate terminator (AT) and mutually exclusive exons (ME) [2, 3]. Recent high-throughput

sequencing studies indicate that >95% of genes undergo AS and generate at least two alternative pre-mRNA isoforms [1, 4]. Aberrant AS events may lead to multiple pathological processes, especially cancer initiation, progression, metastasis and resistance to therapy [5–8]. AS events could be developed as diagnostic or prognostic biomarkers, as well as for exploiting therapeutic targets in cancer patients [2].

Gastric cancer (GC) is one of the most common malignant tumors that originate from the gastric mucosal epithelium. It has been reported that GC has the second highest incidence among various cancers in China and ranks as the third leading cause of cancer-related death worldwide [9, 10]. After the importance of AS events in Epstein-Barr virus-associated GC was investigated [11], a simple prognosis analysis was conducted to assess AS events in stomach adenocarcinoma [12]. However, there is still a lack of clinical references regarding the prognostic value of AS, and the regulatory mechanism governing survival-associated AS events warrants further study. Thus, an in-depth and systematic investigation of survival-associated AS events in GC patients should be conducted to build an independent prognostic signature by integrating seven types of AS events, which may provide suggestions for exploiting personalized treatment strategies and therapeutic targets.

In this study, we illustrated the effects of different AS patterns in a GC cohort using the genome-wide transcriptome approach. RNA-seq data in The Cancer Genome Atlas (TCGA) were used to analyze the incidence of seven AS patterns and explore splicing variant function and survival-associated AS events in GC patients. The potential regulatory mechanisms for survival-related AS events were revealed. More importantly, a final prognostic signature was successfully constructed by combining seven types of AS events, which was demonstrated to be an independent prognostic indicator after multivariate adjustment of clinical parameters.

RESULTS

Integrated AS events in the GC cohort

The SpliceSeq package provides a comprehensive, detailed evaluation of seven types of AS events, including AA, AD, AP, AT, ES, ME, and RI. A total of 48141 AS events of 10610 genes were identified in 415 GC patients, showing that a single gene might have more than one type of mRNA splicing event. A single gene may contain up to six types of splicing events. ES was the most frequent splice signature among the seven

AS types followed by AT and AP. Specifically, we detected 19121 ESs in 6972 genes, 8390 ATs in 3666 genes, 10004 APs in 4025 genes, 4006 AAs in 2799 genes, 3450 ADs in 2401 genes, 2944 RIs in 1956 genes, and 226 MEs in 219 genes (Figure 1).

Survival-associated AS events in the GC cohort

In the survival analysis, 47 patients with an overall survival time of less than 30 days were excluded. The remaining 368 patients were then matched with their corresponding entries in the SpliceSeq database, and 31 cases were excluded once again due to >20% missing AS events. Thus, 337 patients were included in this study and their overall survival status was listed in Supplementary Table 1. To search the prognostic value of AS events in GC patients, a univariate Cox regression analysis was carried out to estimate the influence of each AS event on the overall survival of patients. A total of 1383 AS events were found to be significantly associated with the overall survival of GC patients, including 517 ES events, 354 AP events, 225 AT events, 98 AA events, 104 AD events, 72 RI events, and 13 ME events. The top 20 most significant survival-associated genes in the seven AS events are presented in Figure 2. Notably, most of these AS events were correlated with prognosis, and one gene might have two or more survival-associated splicing events in GC patients. Thus, a subset of overlapping AS events among the seven AS types in GC patients was further analyzed. As illustrated by the UpSet plot diagram in Figure 3A, one gene might undergo two or three types of AS events that were significantly associated with patient survival.

To further explore the functional relationship among these survival-associated AS events, Reactome was used to plot the protein interaction networks. Figure 3B shows the hub genes with survival-associated AS events, such as *STAT3*, *SRSF7*, *KAT5*, *SRSF3*, *SF1*, *LEF1*, *APC*, *RAD51B*, *SNRNP70*, *COL1A1* and *DLG4*.

Prognostic predictors for GC patients

To develop prognostic predictors for GC patient survival, the top significant survival-associated AS events in the seven types were selected as candidates. Least absolute shrinkage and selection operator (LASSO) Cox analysis was carried out to develop seven prognostic signatures based on AA, AD, AP, AT, ES, ME and RI events (Figure 4). As shown in Figure 5, all seven prognostic models built on each type of splicing pattern showed strong potential to predict the outcome of GC patients. Simultaneously, each prognostic model

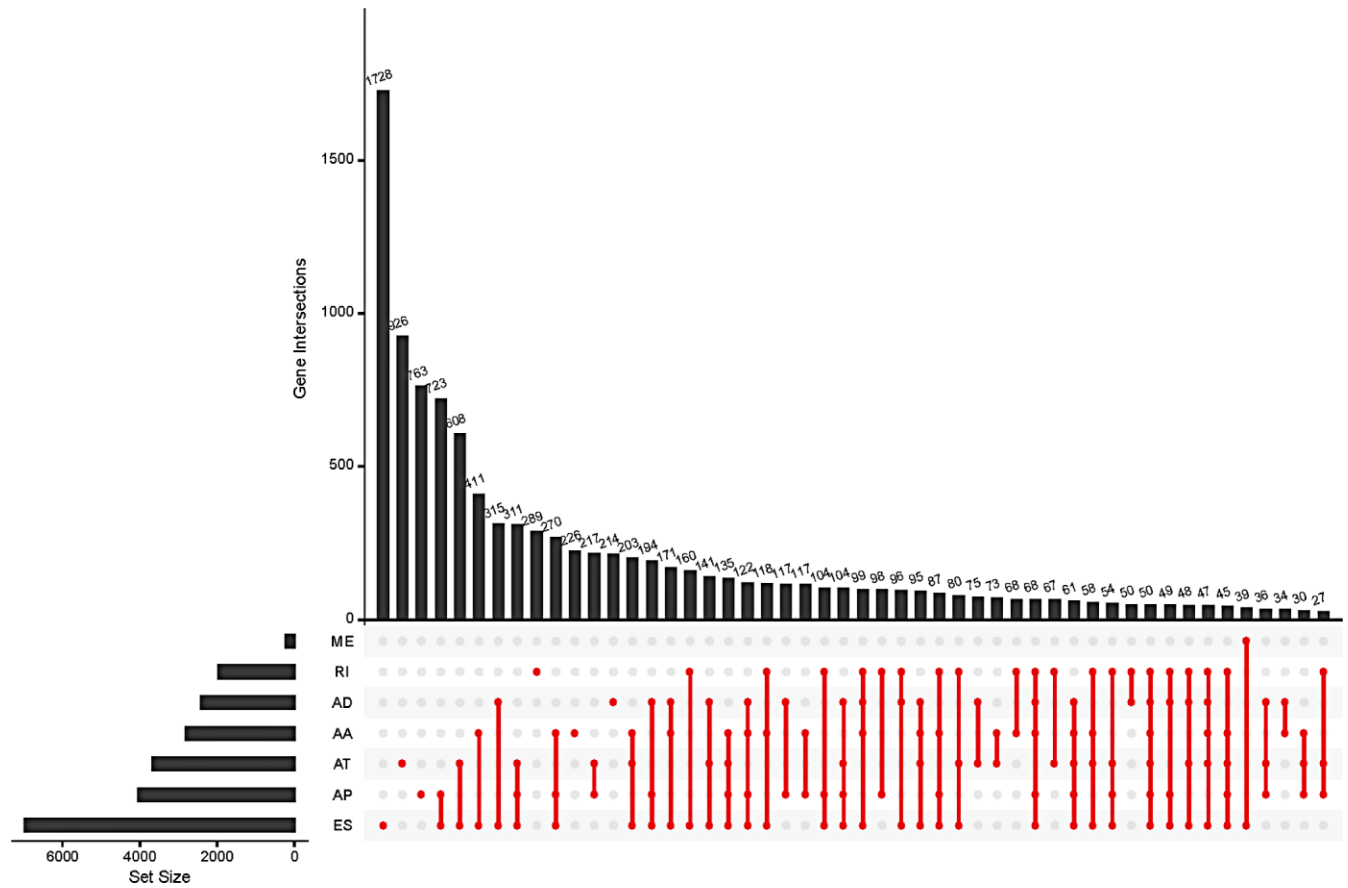


Figure 1. Summary of total AS occurrence samples in the GC cohort.

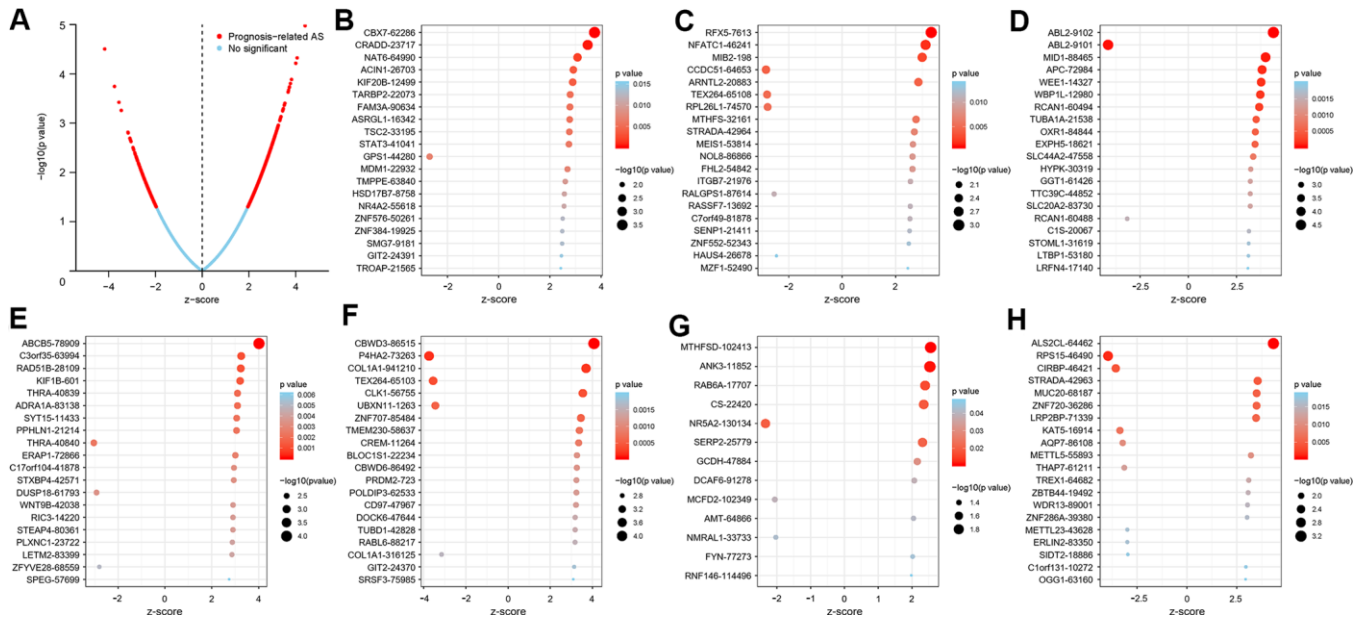


Figure 2. Top 20 most significant AS events in the GC cohort. (A) The blue dots represent no significant AS events, whereas the red dots represent prognosis-related AS events. The top AS events correlated with survival outcome based on AA (B), AD (C), AP (D), AT (E), ES (F), ME (G), and RI (H) events.

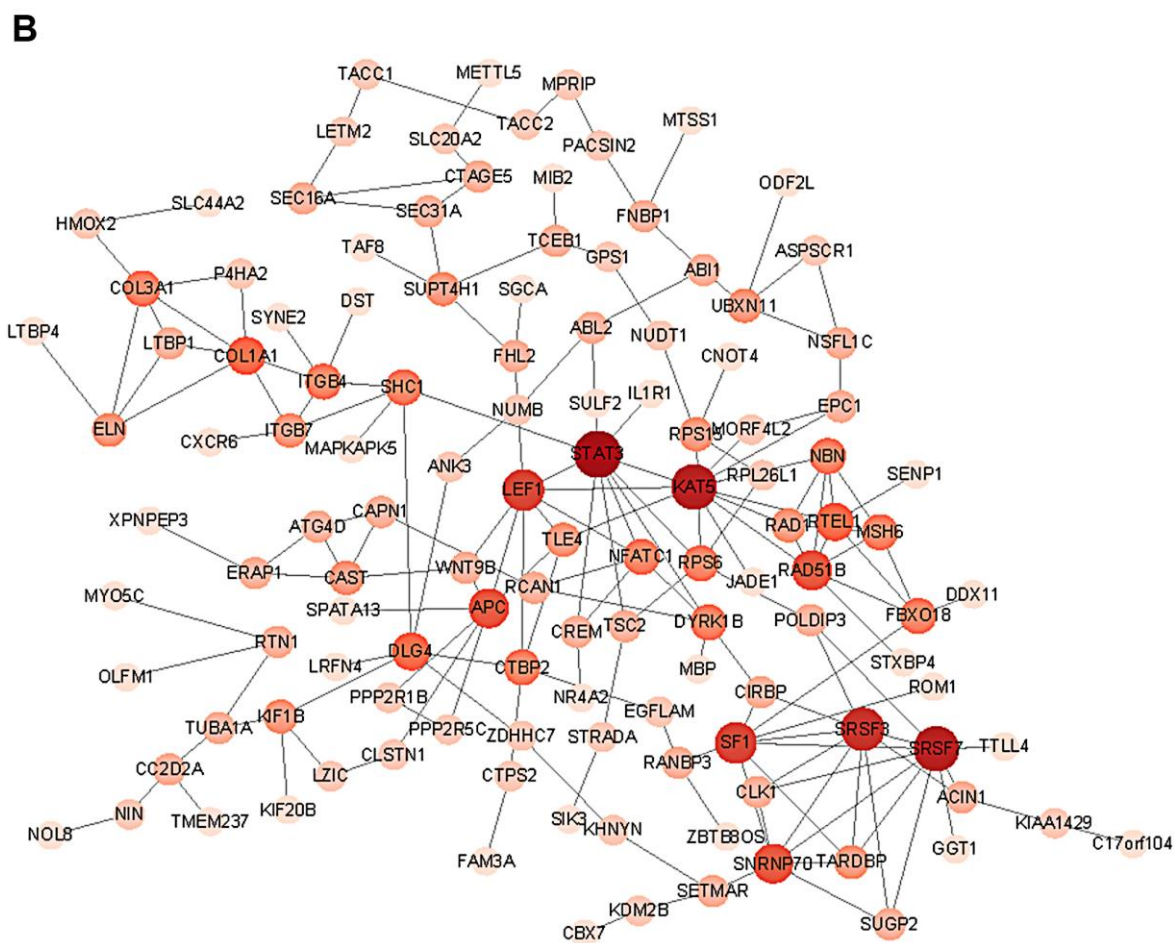
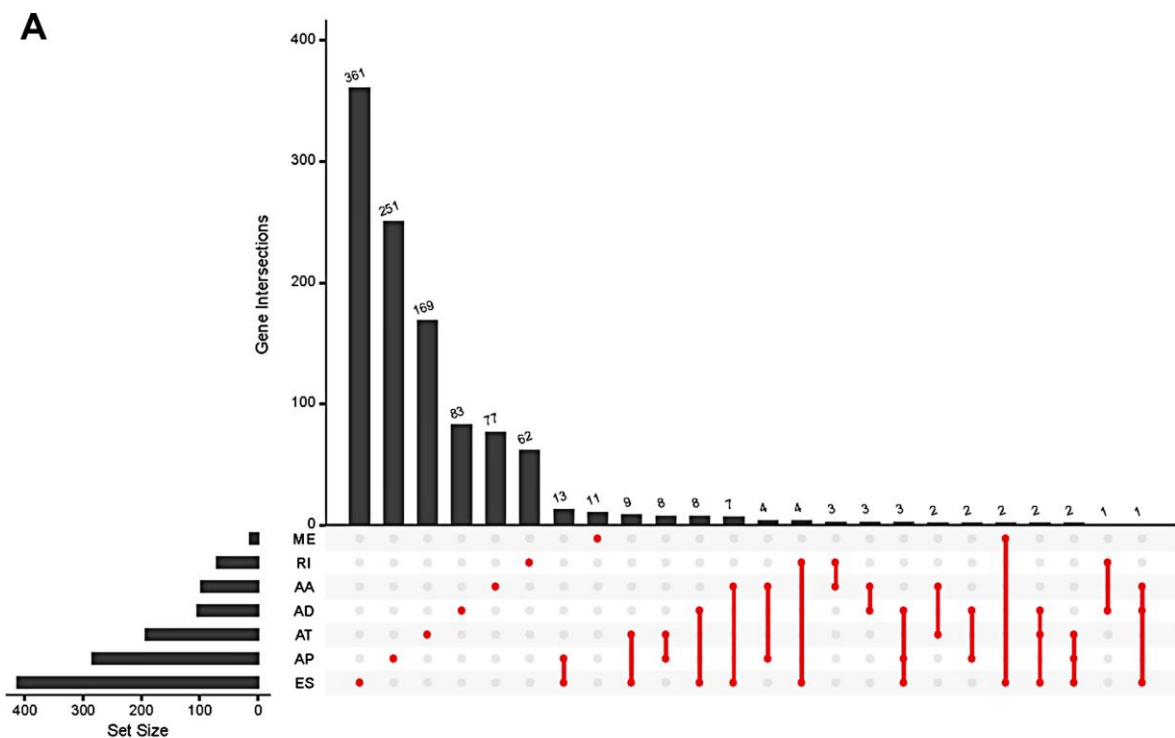


Figure 3. Summary (A) and protein network (B) of survival-associated AS events in the GC cohort.

had significant discrepancies for predicting the survival probability, and the AA-based model showed the most promising outcome prediction among the seven prognostic models. The area under the curve (AUC) of the receiver operating characteristic (ROC) for the AA model was 0.939 followed by the ES, AT, RI, AP, AD and ME models with AUCs of 0.860, 0.808, 0.806,

0.802, 0.789 and 0.698, respectively. Furthermore, these prognostic AS events in seven different types were combined to build the final prognostic signature. Notably, the final prognostic predictor indeed showed better performance in predicting the outcome of GC patients with an AUC of 0.948 (Figure 5P). The final prognostic model provided an indicator to predict the

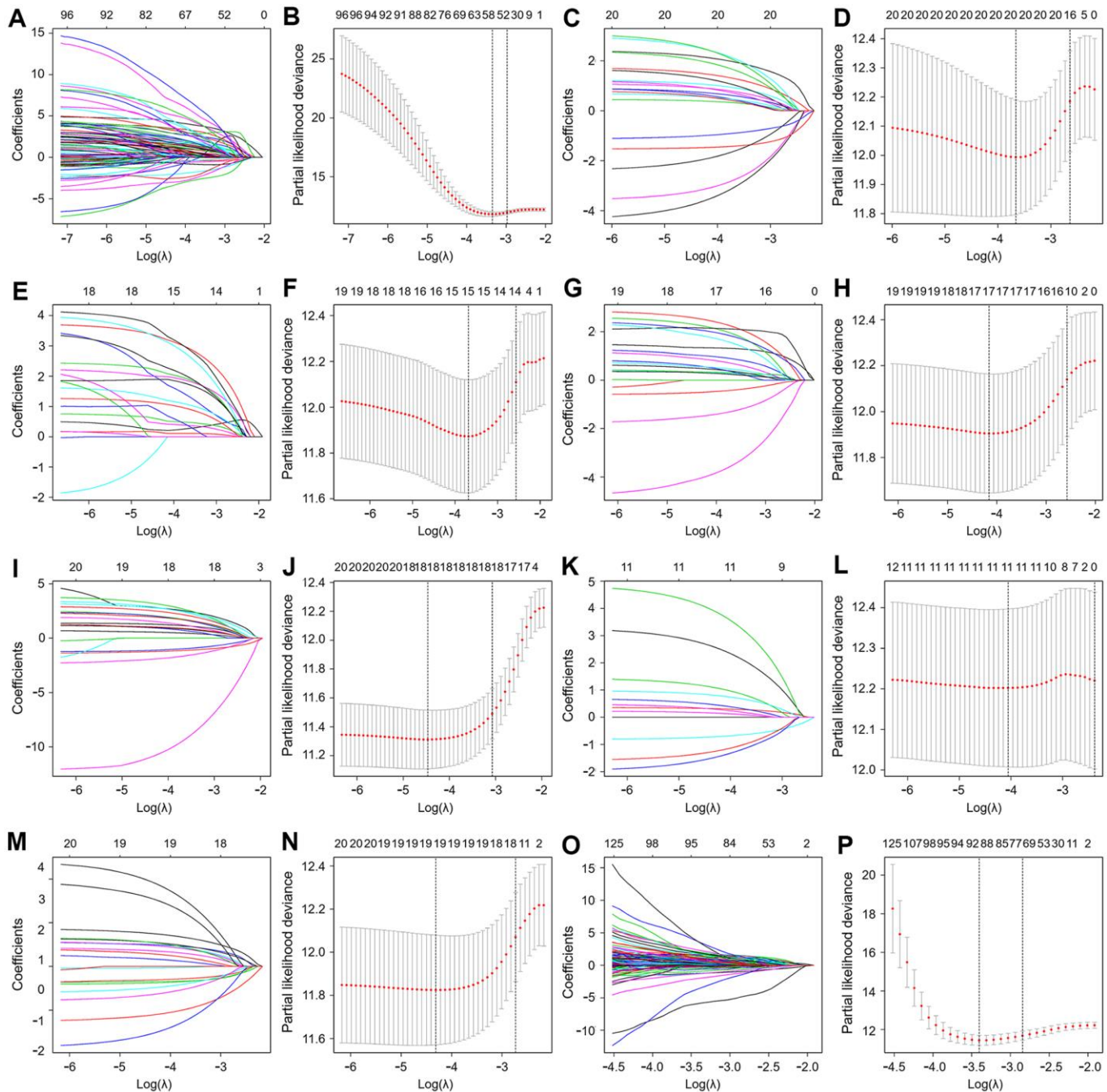


Figure 4. Least absolute shrinkage and selection operator (LASSO) coefficient profiles of the AS events. The coefficient profiles of AA (A), AD (C), AP (E), AT (G), ES (I), ME (K) and RI (M) events. The partial likelihood deviance of AA (B), AD (D), AP (F), AT (H), ES (J), ME (L) and RI (N) events. (O) The coefficient profiles of all seven types of AS events. (P) The partial likelihood deviance of all seven types of AS events.

prognosis of GC patients (Figure 6A). Kaplan-Meier plots indicated that GC patients in the high-risk group had significantly shorter overall survival than those in the low-risk group (Figure 6B), showing that this signature could effectively distinguish GC patients. The percent spliced in (PSI) values of AS events for

building the final prognostic model are shown in Figure 6C. After multivariate adjustment for clinical parameters, the prognostic signature can still act as an independent prognostic indicator (hazard ratio (HR)=1.136, 95% confidence interval (CI): 1.116~1.156, $P<0.001$) (Figure 7).

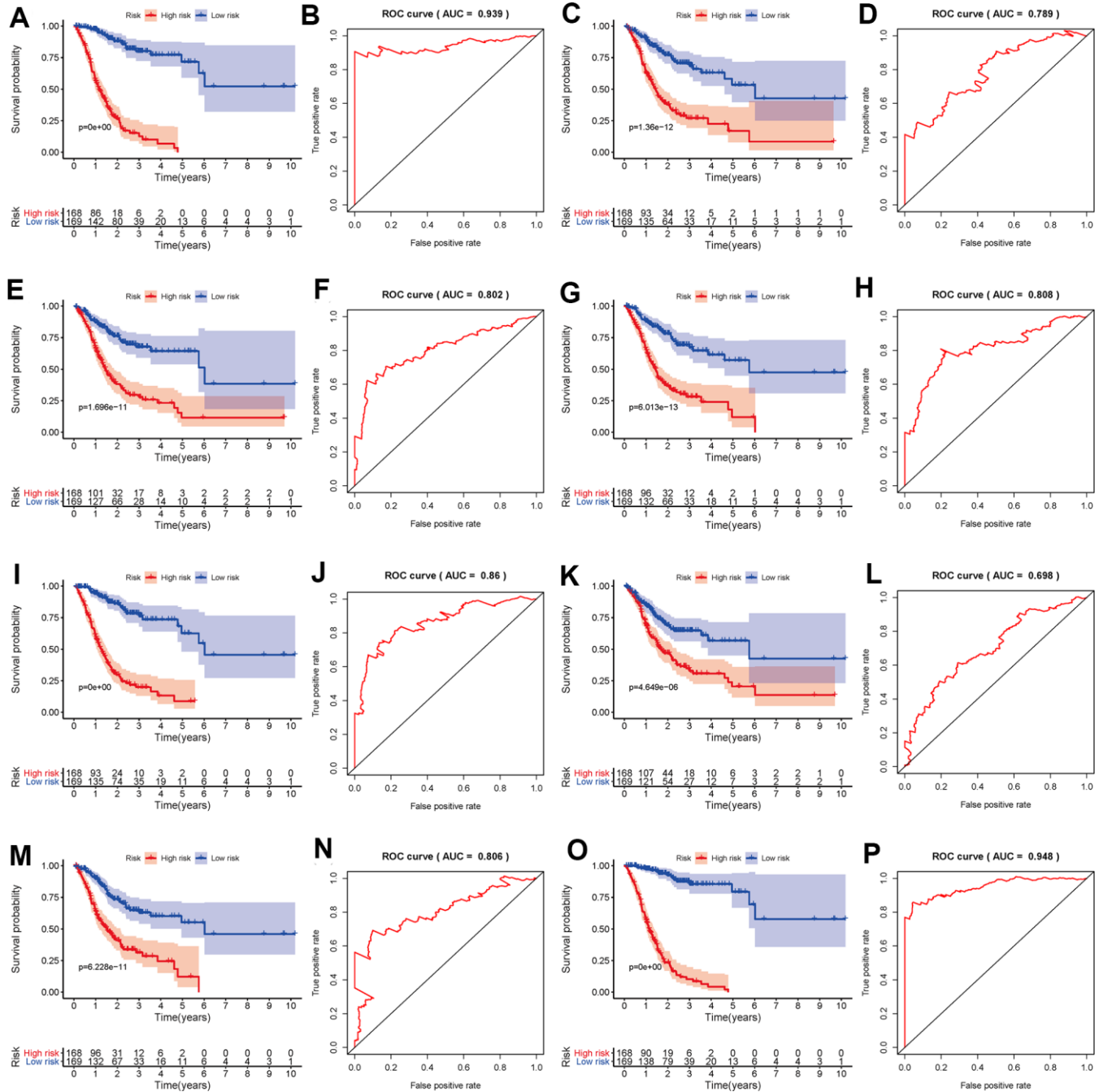


Figure 5. Kaplan-Meier and ROC curves of prognostic predictors in the GC cohort. Kaplan-Meier plot of the survival probability over time for prognostic predictors based on AA (A), AD (C), AP (E), AT (G), ES (I), ME (K) and RI (M) events with high (red) and low (blue) risk groups, respectively. ROC analysis for prognostic predictors based on AA (B), AD (D), AP (F), AT (H), ES (J), ME (L) and RI (N) events. (O) Kaplan-Meier plot depicting the survival probability over time for the final prognostic predictor with high (red) and low (blue) risk groups. (P) ROC analysis for the final prognostic predictor based on all seven types of AS events.

Correlation between survival-associated AS events and splicing factor expression

AS is mainly orchestrated by splicing factors, which recognize and bind to pre-mRNAs at specific positions and regulate RNA splicing. To explore the correlation

between the expression of splicing factors and the PSI values of AS events, Cytoscape software was applied to visualize the splicing-regulatory network of splicing factors and survival-associated AS events. As shown in Figure 8, 12 splicing factors (blue triangles) were significantly associated with 44 survival-associated AS

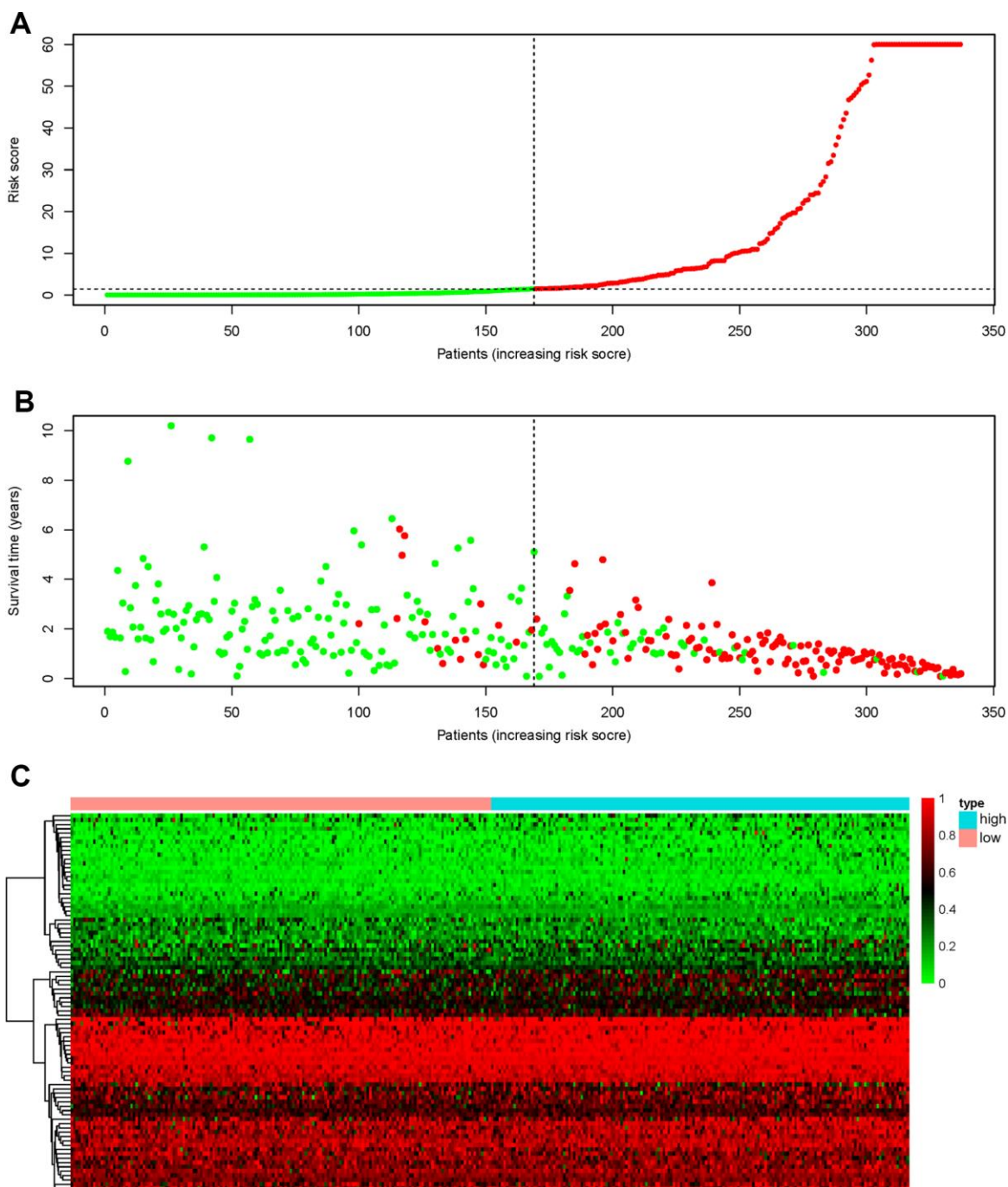


Figure 6. Recognition capability of the final prognostic signature for GC patients into low- and high-risk groups. (A) The risk scores of 337 patients. Green/red dots represent low/high risk groups that are distinguished using the dotted lines. (B) Overall survival status and survival duration of GC patients. Dotted lines were used to distinguish patients in the high- and low-risk groups. Green dots represent surviving patients, while red dots indicate dead patients. (C) Heatmap of PSI values of AS events for building the final prognostic signature.

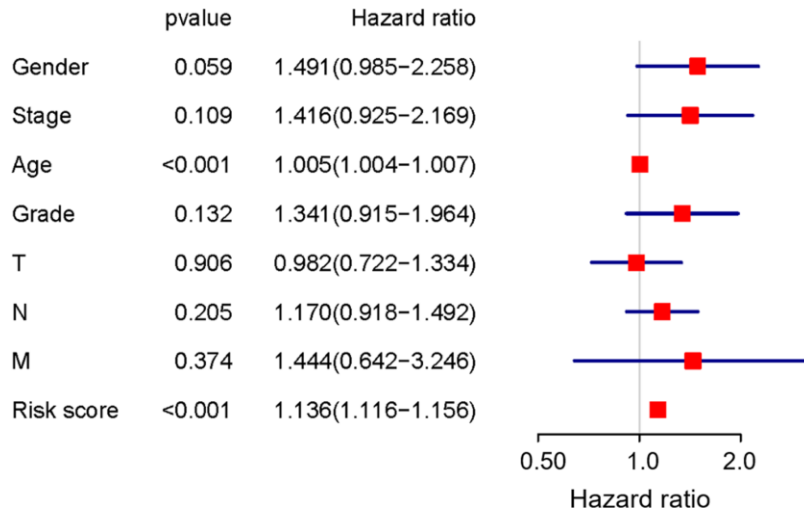


Figure 7. Prognostic value of the final prognostic signature adjusted by clinical parameters.

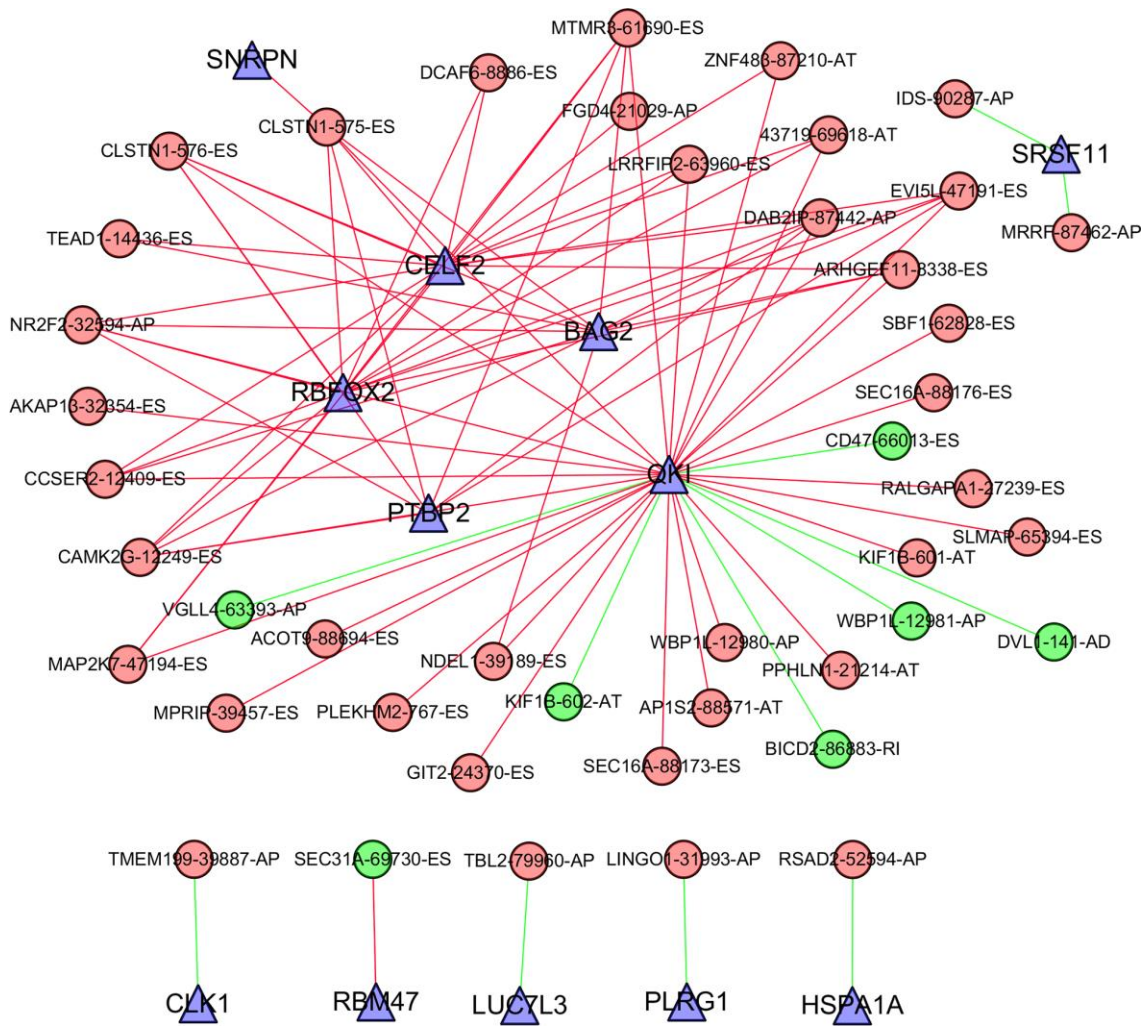


Figure 8. Correlation network between the expression of splicing factors and PSI values of AS events generated using Cytoscape. Triangle bubbles represent splicing factors. Red/green round bubbles represent adverse/favourable AS events. Red/green lines represent positive/negative correlations between substances.

events; among these events, seven were favourable prognosis AS events (green dots), while 37 were poor prognosis AS events (red dots). Moreover, the majority of adverse AS events were positively correlated (red lines) with splicing factor expression, whereas the majority of favourable AS events were negatively correlated (green lines) with splicing factor expression. In addition, five splicing factors, including *CELF2*, *BAG2*, *RBFOX2*, *PTBP2* and *QKI*, were associated with more than two AS events. Among these factors, the expression of *CELF2*, *BAG2*, *RBFOX2* and *PTBP2* was positively correlated with adverse AS events, whereas *QKI* expression was positively correlated with partial adverse AS events and negatively correlated with favourable AS events.

Functional enrichment analysis

To elucidate the function of genes with survival-associated AS events, Gene Ontology (GO) term enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out. As shown in Table 1, “positive regulation of GTPase activity”, “regulation of cell cycle”, and “DNA repair” were the three most significant biological process terms; “protein binding”, “guanyl-nucleotide exchange factor activity” and “microtubule binding” were the three most significant molecular function terms; “cytosol”, “nucleoplasm” and “cytoplasm” were the three most significant cellular component terms. Furthermore, we found that the ten signaling pathways significantly correlated with these genes were focal adhesion, the mTOR signaling pathway, the HIF-1 signaling pathway, the PI3K-Akt signaling pathway, the ErbB signaling pathway, chronic myeloid leukaemia, pathways in cancer, glyoxylate and dicarboxylate metabolism, glycosaminoglycan degradation, and the insulin signaling pathway (Figure 9).

DISCUSSION

In 2010, Miura et al. summarized aberrant splicing and altered variant expression in gastrointestinal malignancies and noted the significance of AS in normal and malignant tissues [13]. Growing evidence further demonstrates that aberrant AS events can play an important role in cancer development [6, 14]. At the same time, AS events have shown great prognostic value in various cancer patients. For example, Lin et al. developed a final prognostic signature that can act as an independent prognostic factor for papillary thyroid cancer patients [15]. Lin et al. also highlighted the prognostic value of AS events and splicing factors in gastrointestinal pan-adenocarcinomas [16]. The only analysis of AS events in GC patients has many limitations [12], and it remains urgent to conduct a

systematic analysis referring to alterations in the AS patterns and their clinical significance and underlying molecular function, which may help cancer researchers effectively recognize the widespread applicability of AS events in GC patients.

In this study, we downloaded AS profiles from TCGA SpliceSeq and comprehensively investigated the prognostic value of AS events in GC patients. First, we conducted SpliceSeq analyses to generate AS profiles in the GC cohort, which identified 48141 AS events of 10610 genes and verified more than one type of mRNA splicing event in a single gene. Then, a univariate Cox regression analysis was applied to estimate the association of overall survival and AS events; this analysis confirmed the prognostic value of AS events in GC patients. Subsequently, we proposed prognostic predictors for GC patients. All seven prediction models built by individual AS patterns showed considerable potential applications for the prognosis of GC patients. AA events displayed the highest efficiency in distinguishing survival outcome of GC patients, which is similar in colorectal cancer patients [17]. ES events have the highest incidence in GC patients. The AUCs of the ROC curves for the AA and ES models were 0.939 and 0.860, respectively, indicating that these prognostic models might be useful for risk stratification in GC patients. Moreover, a final prognostic predictor model was successfully constructed by the combination of all seven types of AS patterns. The AUC of ROC for this final high-performance model reached 0.948, implying that this model could be more precise in GC prognosis. This model was deployed to recognize GC patients, and the results clearly showed that it could distinguish the GC cohort with distinct clinical outcomes notably well. In addition, the White and Asian GC cohorts were verified to be properly fitted with the prognostic model (Supplementary Figure 2). Thus, the final prognostic signature was an ideal indicator to predict the prognosis of GC patients. Certainly, the present study inevitably had several limitations. First, the number of patients were limited and another independent validation in a larger clinical cohort should be performed in the future study. Second, more *in vitro* and *in vivo* experiments are required to elucidate the biological function of these AS events and splicing factors. In addition, this final model may not always effective in the clinical prediction owing to tumor heterogeneity, individual differences and the effects of various treatment options on GC patients.

Generally, the occurrence of AS events is regulated by *cis*-acting regulatory sequences and RNA-binding protein splicing factors. Splicing factors can act as activators or repressors, depending on their recognition and binding position in a pre-mRNA [1]. Errors in

Table 1. GO analysis of genes with survival-associated AS events.

Ontology	ID	Description	p. adjust	Count	
Biological process	GO:0043547	positive regulation of GTPase activity	7.66E-04	51	
	GO:0051726	regulation of cell cycle	0.001398	17	
	GO:0006281	DNA repair	0.001411	26	
	GO:0043407	negative regulation of MAP kinase activity	0.003236	8	
	GO:0030522	intracellular receptor signaling pathway	0.004442	8	
	GO:0043484	regulation of RNA splicing	0.004573	7	
	GO:0009612	response to mechanical stimulus	0.004988	10	
	GO:0007265	Ras protein signal transduction	0.005109	11	
	GO:0007050	cell cycle arrest	0.005168	17	
	GO:0006351	DNA-templated transcription	0.005416	135	
	GO:0005515	protein binding	3.77E-11	587	
	GO:0005085	guanyl-nucleotide exchange factor activity	9.02E-05	19	
	GO:0008017	microtubule binding	0.001328	24	
	GO:0003684	damaged DNA binding	0.002371	11	
Molecular function	GO:0042802	identical protein binding	0.004441	60	
	GO:0008307	structural constituent of muscle	0.007928	8	
	GO:0005089	Rho guanyl-nucleotide exchange factor activity	0.010129	11	
	GO:0070300	phosphatidic acid binding	0.01015	5	
	GO:0004672	protein kinase activity	0.010378	32	
	GO:0033130	acetylcholine receptor binding	0.011224	4	
	GO:0005829	cytosol	1.14E-11	261	
	GO:0005654	nucleoplasm	8.49E-11	224	
	GO:0005737	cytoplasm	8.81E-09	360	
	GO:0005634	nucleus	4.99E-05	346	
	Cellular component	GO:0005925	focal adhesion	6.65E-05	41
		GO:0016607	nuclear speck	0.001156	23
		GO:0045111	intermediate filament cytoskeleton	0.001594	10
		GO:0005875	microtubule associated complex	0.001853	8
GO:0035267		NuA4 histone acetyltransferase complex	0.002072	6	
GO:0005874		microtubule	0.002649	30	

splicing profiles can lead to disease states, which was also verified in multiple tumor types [1, 15–18]. Meanwhile, the importance of splicing factors in cancer was confirmed. Alterations in splicing factors in various cancers are considered independent molecules involved in the prediction of cancer outcome [19–21]. In GC patients, we found that 12 splicing factors were significantly associated with 44 survival-associated AS events. The splicing correlation network clearly showed that *CELF2*,

BAG2, *RBFOX2*, *PTBP2* and *QKI* were associated with more than two AS events, confirming that limited splicing factors can dominate numerous AS events [1]. According to existing studies, these splicing factors are highly correlated with the development, progression and prognosis of multiple cancers [22–26]. Notably, *CELF2*, *RBFOX2*, *PTBP2* and *QKI* were also involved in the misregulation of AS events in Epstein-Barr virus-associated gastric carcinomas [11]. The network in our analysis showed

that the expression of these five splicing factors was positively correlated with adverse AS events or negatively correlated with favourable AS events, indicating a strongly negative impact on the clinical outcome of GC patients. Thus, it may be valuable to study these splicing factors and their relevant survival-associated AS events, which may provide a new perspective in contrast to current approaches and may help to develop novel therapeutic targets for GC patients.

Further functional enrichment analysis showed that the “focal adhesion” pathway was most significantly correlated with these genes with survival-associated AS events. It is well-known that focal adhesions are stable integrin-mediated, cell-substrate adhesion structures that anchor cells via integrin receptors to the extracellular matrix (ECM) and intracellularly connect to actin stress fibres [27, 28]. These structures have been extensively characterized in cultured cells, in which they not only form a structural link between cells and their microenvironment but are also important for cell differentiation, proliferation, adhesion, and migration [28–32]. A key regulator in focal adhesions is the nonreceptor tyrosine kinase focal adhesion kinase (FAK), which triggers focal adhesion signals on cell adhesion to the ECM [28]. Interestingly, FAK is closely

related to relevant signaling pathways, including the mTOR signaling pathway, HIF-1 signaling pathway, and PI3K-Akt signaling pathway, which are synergistically involved in tumor migration, invasion, metastasis and recurrence [33, 34]. Thus, AS events occurring in GC patients might impact tumor invasion and metastasis via the focal adhesion pathway.

In summary, the current study established an ideal independent prognostic signature to predict the prognosis of GC patients based on systematic analyses of AS profiles and survival-associated AS events. The AUC of ROC for the final prognostic predictor was 0.948, showing considerable promise for predicting the outcome of GC patients. Further confirmation analysis clearly revealed good performance in distinguishing the GC cohort with distinct risk scores. Simultaneously, the constructed network reveals potential regulatory mechanisms between splicing factors and survival-associated AS events. In addition, GO enrichment and KEGG pathway analyses elucidated that the genes with survival-associated AS events were closely correlated with tumorigenesis and development, especially focal adhesion and its interdependent signaling pathway. These findings may facilitate ongoing efforts to develop novel diagnostic or prognostic biomarkers and exploit therapeutic targets for GC patients.

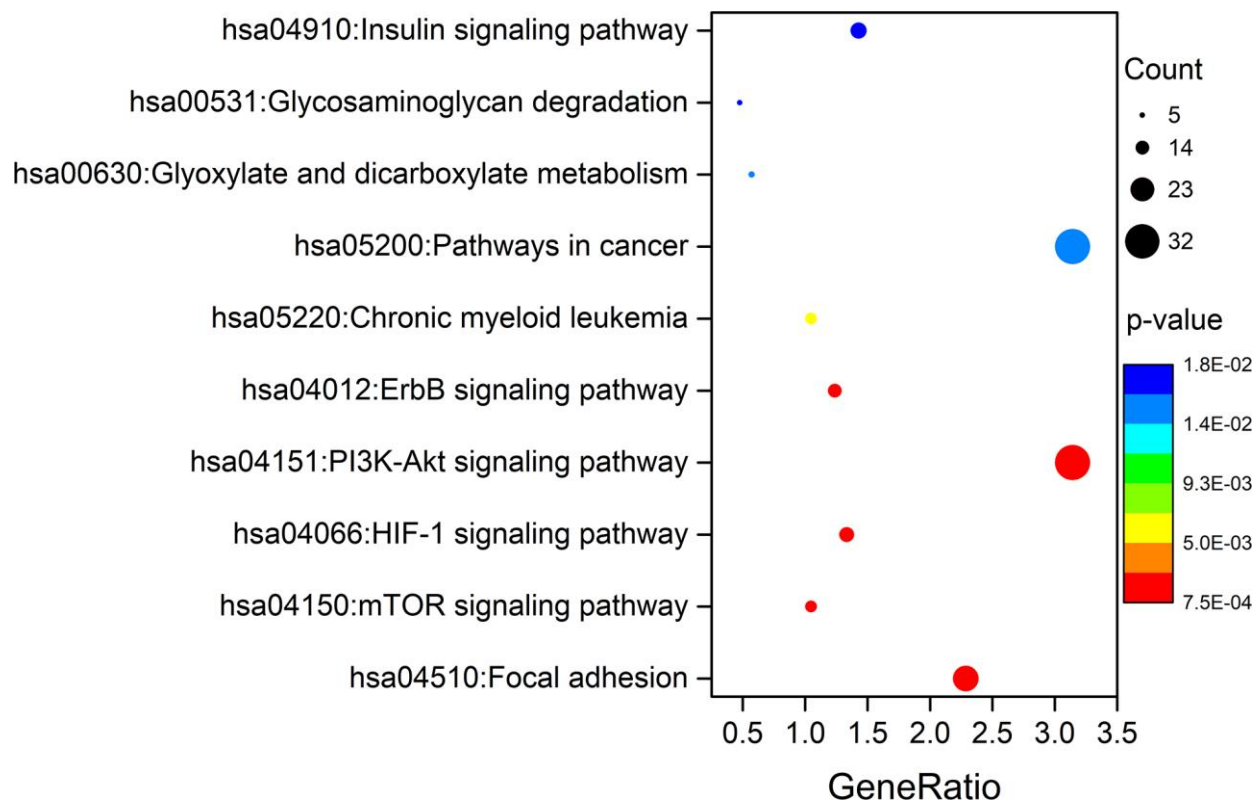


Figure 9. KEGG pathway analysis of genes with survival-associated AS events.

MATERIALS AND METHODS

Data acquisition

TCGA's SpliceSeq is a computational tool that provides AS profiles on the basis of RNA-seq data [35]. In GC cohort samples, AS events occurring in $\geq 75\%$ of samples were downloaded from the SpliceSeq database. Clinical information of GC patients was also downloaded and segregated from the TCGA pan-cancer atlas database [36]. AS event data and clinical information were confirmed using the same TCGA ID. The primary tumor characteristics and clinical information were listed in Supplementary Table 2. Subsequent analyses were conducted and the flow chart of data processing are shown in Supplementary Figure 1.

Identification of survival-associated AS events

In the survival analysis, only GC patients with both clinical follow-up and AS event data were finally enrolled in this study. Simultaneously, the cases in the following two criteria were excluded: (1) the overall survival time of patients less than 30 days; (2) the missing of AS events more than 20%. After AS events with a standard deviation (SD) <0.01 excluded, a univariate Cox analysis was performed to assess the relationships between each AS event and the overall survival of GC patients. UpSet was applied to visualize the associations between genes and each type of AS event. The Reactome was used to plot the protein networks to explore the interactions between the genes corresponding to survival-associated AS events.

Construction of prognostic predictors

The survival-associated AS events were selected for multivariate LASSO Cox analysis to develop prognostic signatures based on AA, AD, AP, AT, ES, ME and RI events. Subsequently, prognostic models were constructed, and Kaplan-Meier analysis was used to plot the survival probability over time for prognostic predictors of seven types of AS events. Then, the final prognostic model was constructed by the combination of all seven types of AS patterns. At the same time, ROC analyses were performed to assess the validity of prognostic predictors. The prognostic model was constructed to predict the clinical outcomes of GC patients. To further check the validity of the final prognostic model, the Kaplan-Meier method was deployed to plot the relationships between the survival outcome of GC patients with high and low risk scores. Finally, the final model was adjusted using a multivariate Cox regression analysis of sex, stage, age, tumor grade and the risk score.

Construction of the correlation network of survival-associated AS events and splicing factors

A catalogue of 404 splicing factor genes was referred to a previous study [37]. The expression profiles of splicing factor genes were obtained from the TCGA database. The count value of splicing factor level-3 mRNA data was also downloaded and converted to $\log_2(\text{count}+1)$. Cytoscape software version 3.7.1 was used to generate the correlation network between the expression of splicing factors and PSI values of survival-associated AS events. We set the parameters of $P<0.001$ and Pearson correlation coefficient (r) >0.6 .

Functional annotation

ClusterProfiler was applied to comprehensively perform GO term enrichment and KEGG analyses of 1088 genes corresponding to the significant survival-associated AS events. We considered the categories with a $P<0.05$ to be significant and displayed the top 10 enrichments of each sub-ontology and pathways.

Abbreviations

GC: gastric cancer; AS: alternative splicing; ES: exon skip; RI: retained intron; AD: alternate donor site; AA: alternate acceptor site; AP: alternate promoter; AT: alternate terminator; ME: mutually exclusive exons; TCGA: The Cancer Genome Atlas; LASSO: least absolute shrinkage and selection operator; ROC: receiver operating characteristic; AUC: area under the ROC curve; PSI: percent spliced in; HR: hazard ratio; CI: confidence interval; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

AUTHOR CONTRIBUTIONS

Conception and design: Shichao Zhang, Zuquan Hu, Xiaofeng Xu, Zhu Zeng, Yan Ouyang. Collection and assembly of data: Shichao Zhang, Yingwu Lan, Yan Ouyang. Data analysis and interpretation: Zuquan Hu, Yingwu Lan, Jinhua Long, Yun Wang, Xiaowen Chen, Xiaofeng Xu. Manuscript writing: Shichao Zhang, Zuquan Hu, Zhu Zeng. Final approval of manuscript: all authors.

ACKNOWLEDGMENTS

The authors thank the TCGA's Spliceseq and TCGA databases for the availability of the data.

CONFLICTS OF INTEREST

Authors declare no conflicts of interest for this article.

FUNDING

This study was jointly funded by the National Natural Science Foundation of China (31660258, 31960139, 31860244, 31771014, 11762006, 21906036), the Guizhou Provincial Natural Science Foundation (2019-1275, 2019-2787, 2018-1412, 2016-5676, 2020-1Z016, 2020-1Y087, 2020-5014, 2016-7381), the Science and Technology Foundation of Guizhou Health Commission (2019-1037), the Excellent Young Talents Plan of Guizhou Medical University (2020-105).

REFERENCES

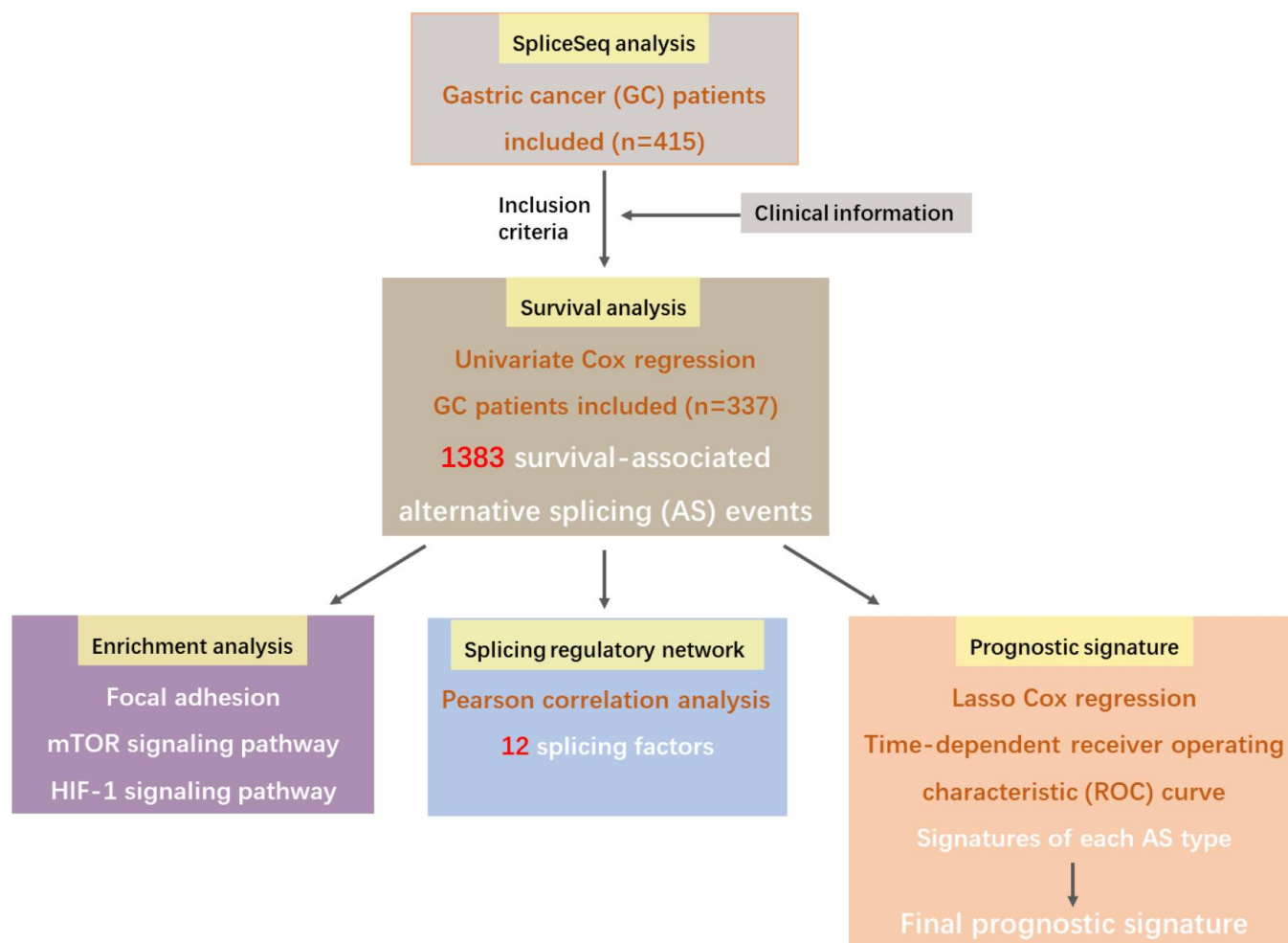
1. Lee Y, Rio DC. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu Rev Biochem*. 2015; 84:291–323.
<https://doi.org/10.1146/annurev-biochem-060614-034316> PMID:25784052
2. Chen QF, Li W, Wu P, Shen L, Huang ZL. Alternative splicing events are prognostic in hepatocellular carcinoma. *Aging (Albany NY)*. 2019; 11:4720–35.
<https://doi.org/10.18632/aging.102085> PMID:31301224
3. Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z, He J. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer Lett*. 2017; 393:40–51.
<https://doi.org/10.1016/j.canlet.2017.02.016> PMID:28223168
4. Nielsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010; 463:457–63.
<https://doi.org/10.1038/nature08909> PMID:20110989
5. Calabrese C, Davidson NR, Demircioğlu D, Fonseca NA, He Y, Kahles A, Lehmann KV, Liu F, Shiraishi Y, Soulette CM, Urban L, Greger L, Li S, et al, and PCAWG Transcriptome Core Group, and PCAWG Transcriptome Working Group, and PCAWG Consortium. Genomic basis for RNA alterations in cancer. *Nature*. 2020; 578:129–36.
<https://doi.org/10.1038/s41586-020-1970-0> PMID:32025019
6. Climente-González H, Porta-Pardo E, Godzik A, Eyra E. The functional impact of alternative splicing in cancer. *Cell Rep*. 2017; 20:2215–26.
<https://doi.org/10.1016/j.celrep.2017.08.012> PMID:28854369
7. Kouyama Y, Masuda T, Fujii A, Ogawa Y, Sato K, Tobo T, Wakiyama H, Yoshikawa Y, Noda M, Tsuruda Y, Kuroda Y, Eguchi H, Ishida F, et al. Oncogenic splicing abnormalities induced by DEAD-box helicase 56 amplification in colorectal cancer. *Cancer Sci*. 2019; 110:3132–44.
<https://doi.org/10.1111/cas.14163> PMID:31390121
8. Lee SC, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. *Nat Med*. 2016; 22:976–86.
<https://doi.org/10.1038/nm.4165> PMID:27603132
9. Song Z, Wu Y, Yang J, Yang D, Fang X. Progress in the treatment of advanced gastric cancer. *Tumour Biol*. 2017; 39:1010428317714626.
<https://doi.org/10.1177/1010428317714626> PMID:28671042
10. Ge S, Xia X, Ding C, Zhen B, Zhou Q, Feng J, Yuan J, Chen R, Li Y, Ge Z, Ji J, Zhang L, Wang J, et al. A proteomic landscape of diffuse-type gastric cancer. *Nat Commun*. 2018; 9:1012.
<https://doi.org/10.1038/s41467-018-03121-2> PMID:29520031
11. Armero VE, Tremblay MP, Allaire A, Boudreault S, Martenon-Brodeur C, Duval C, Durand M, Lapointe E, Thibault P, Tremblay-Létourneau M, Perreault JP, Scott MS, Bisailon M. Transcriptome-wide analysis of alternative RNA splicing events in epstein-barr virus-associated gastric carcinomas. *PLoS One*. 2017; 12:e0176880.
<https://doi.org/10.1371/journal.pone.0176880> PMID:28493890
12. Shi Y, Chen Z, Gao J, Wu S, Gao H, Feng G. Transcriptome-wide analysis of alternative mRNA splicing signature in the diagnosis and prognosis of stomach adenocarcinoma. *Oncol Rep*. 2018; 40:2014–22.
<https://doi.org/10.3892/or.2018.6623> PMID:30106437
13. Miura K, Fujibuchi W, Sasaki I. Alternative pre-mRNA splicing in digestive tract Malignancy. *Cancer Sci*. 2011; 102:309–16.
<https://doi.org/10.1111/j.1349-7006.2010.01797.x> PMID:21134075
14. Paronetto MP, Passacantilli I, Sette C. Alternative splicing and cell survival: from tissue homeostasis to disease. *Cell Death Differ*. 2016; 23:1919–29.
<https://doi.org/10.1038/cdd.2016.91> PMID:27689872
15. Lin P, He RQ, Huang ZG, Zhang R, Wu HY, Shi L, Li XJ, Li Q, Chen G, Yang H, He Y. Role of global aberrant alternative splicing events in papillary thyroid cancer prognosis. *Aging (Albany NY)*. 2019; 11:2082–97.
<https://doi.org/10.18632/aging.101902> PMID:30986203
16. Lin P, He RQ, Ma FC, Liang L, He Y, Yang H, Dang YW, Chen G. Systematic analysis of survival-associated alternative splicing signatures in gastrointestinal pan-adenocarcinomas. *EBioMedicine*. 2018; 34:46–60.
<https://doi.org/10.1016/j.ebiom.2018.07.040> PMID:30131306

17. 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. <https://doi.org/10.3389/fonc.2018.00537> PMID:30524964
18. Lee AR, Gan Y, Xie N, Ramnarine VR, Lovnicki JM, Dong X. Alternative RNA splicing of the GIT1 gene is associated with neuroendocrine prostate cancer. *Cancer Sci.* 2019; 110:245–55. <https://doi.org/10.1111/cas.13869> PMID:30417466
19. Wu F, Chen Q, Liu C, Duan X, Hu J, Liu J, Cao H, Li W, Li H. Profiles of prognostic alternative splicing signature in hepatocellular carcinoma. *Cancer Med.* 2020; 9:2171–80. <https://doi.org/10.1002/cam4.2875> PMID:31975560
20. Wang C, Zheng M, Wang S, Nie X, Guo Q, Gao L, Li X, Qi Y, Liu J, Lin B. Whole genome analysis and prognostic model construction based on alternative splicing events in endometrial cancer. *Biomed Res Int.* 2019; 2019:2686875. <https://doi.org/10.1155/2019/2686875> PMID:31355251
21. Li ZX, Zheng ZQ, Wei ZH, Zhang LL, Li F, Lin L, Liu RQ, Huang XD, Lv JW, Chen FP, He XJ, Guan JL, Kou J, et al. Comprehensive characterization of the alternative splicing landscape in head and neck squamous cell carcinoma reveals novel events associated with tumorigenesis and the immune microenvironment. *Theranostics.* 2019; 9:7648–65. <https://doi.org/10.7150/thno.36585> PMID:31695792
22. Piqué L, Martínez de Paz A, Piñeyro D, Martínez-Cardús A, Castro de Moura M, Llinàs-Arias P, Setien F, Gomez-Miragaya J, Gonzalez-Suarez E, Sigurdsson S, Jonasson JG, Villanueva A, Vidal A, et al. Epigenetic inactivation of the splicing RNA-binding protein CELF2 in human breast cancer. *Oncogene.* 2019; 38:7106–12. <https://doi.org/10.1038/s41388-019-0936-x> PMID:31409895
23. Qin L, Guo J, Zheng Q, Zhang H. BAG2 structure, function and involvement in disease. *Cell Mol Biol Lett.* 2016; 21:18. <https://doi.org/10.1186/s11658-016-0020-2> PMID:28536620
24. Quentmeier H, Pommerenke C, Bernhart SH, Dirks WG, Hauer V, Hoffmann S, Nagel S, Siebert R, Uphoff CC, Zaborski M, Drexler HG, and ICGC MMML-Seq Consortium. RBF2 and alternative splicing in b-cell lymphoma. *Blood Cancer J.* 2018; 8:77. <https://doi.org/10.1038/s41408-018-0114-3> PMID:30097561
25. Cheung HC, Hai T, Zhu W, Baggerly KA, Tsavachidis S, Krahe R, Cote GJ. Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. *Brain.* 2009; 132:2277–88. <https://doi.org/10.1093/brain/awp153> PMID:19506066
26. Mo HY, Jo YS, Yoo NJ, Kim MS, Song SY, Lee SH. Frameshift mutation of candidate tumor suppressor genes QK1 and TMEFF2 in gastric and colorectal cancers. *Cancer Biomark.* 2019; 24:1–6. <https://doi.org/10.3233/CBM-160559> PMID:30614793
27. Winograd-Katz SE, Fässler R, Geiger B, Legate KR. The integrin adhesome: from genes and proteins to human disease. *Nat Rev Mol Cell Biol.* 2014; 15:273–88. <https://doi.org/10.1038/nrm3769> PMID:24651544
28. Bauer MS, Baumann F, Daday C, Redondo P, Durner E, Jobst MA, Milles LF, Mercadante D, Pippig DA, Gaub HE, Gräter F, Lietha D. Structural and mechanistic insights into mechanoactivation of focal adhesion kinase. *Proc Natl Acad Sci USA.* 2019; 116:6766–74. <https://doi.org/10.1073/pnas.1820567116> PMID:30877242
29. Bosch-Fortea M, Martín-Belmonte F. Mechanosensitive adhesion complexes in epithelial architecture and cancer onset. *Curr Opin Cell Biol.* 2018; 50:42–49. <https://doi.org/10.1016/j.ceb.2018.01.013> PMID:29454273
30. Xu Z, Isaji T, Fukuda T, Wang Y, Gu J. O-GlcNAcylation regulates integrin-mediated cell adhesion and migration via formation of focal adhesion complexes. *J Biol Chem.* 2019; 294:3117–24. <https://doi.org/10.1074/jbc.RA118.005923> PMID:30587575
31. Wang W, Zuidema A, Te Molder L, Nahidiazar L, Hoekman L, Schmidt T, Coppola S, Sonnenberg A. Hemidesmosomes modulate force generation via focal adhesions. *J Cell Biol.* 2020; 219:e201904137. <https://doi.org/10.1083/jcb.201904137> PMID:31914171
32. Ye K, Ouyang X, Wang Z, Yao L, Zhang G. SEMA3F promotes liver hepatocellular carcinoma metastasis by activating focal adhesion pathway. *DNA Cell Biol.* 2020; 39:474–83. <https://doi.org/10.1089/dna.2019.4904> PMID:31968181
33. Neoh CA, Wu WT, Dai GF, Su JH, Liu CI, Su TR, Wu YJ. Flaccidoxide-13-acetate extracted from the soft coral cladiella kashmani reduces human bladder cancer cell migration and invasion through reducing activation of the FAK/PI3K/AKT/mTOR signaling pathway. *Molecules.* 2017; 23:58. <https://doi.org/10.3390/molecules23010058> PMID:29280977

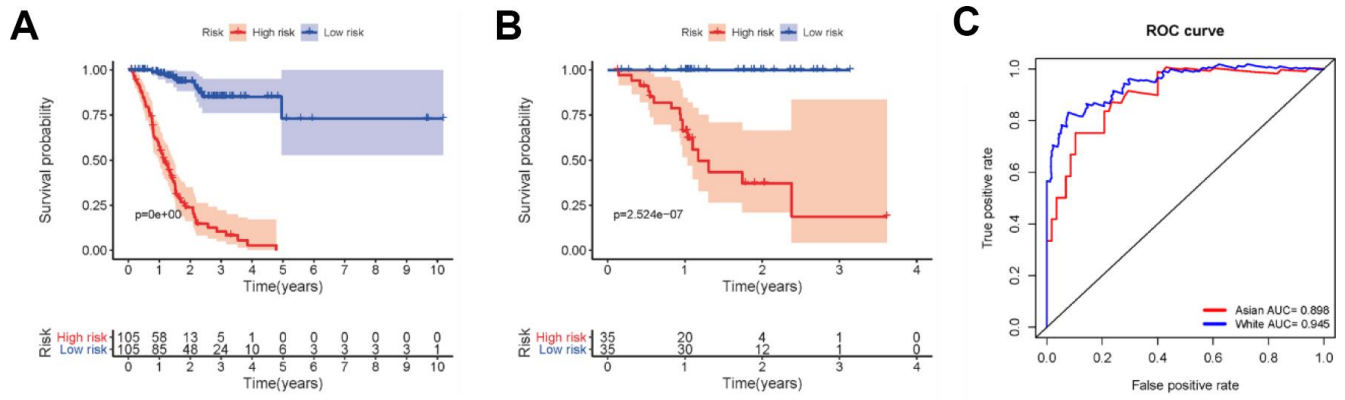
34. Gan L, Meng J, Xu M, Liu M, Qi Y, Tan C, Wang Y, Zhang P, Weng W, Sheng W, Huang M, Wang Z. Extracellular matrix protein 1 promotes cell metastasis and glucose metabolism by inducing integrin β 4/FAK/SOX2/HIF-1 α signaling pathway in gastric cancer. *Oncogene*. 2018; 37:744–55.
<https://doi.org/10.1038/onc.2017.363> PMID:[29059156](https://pubmed.ncbi.nlm.nih.gov/29059156/)
35. Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, Melott J, Weinstein J. TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res*. 2016; 44:D1018–22.
<https://doi.org/10.1093/nar/gkv1288> PMID:[26602693](https://pubmed.ncbi.nlm.nih.gov/26602693/)
36. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD, et al, and Cancer Genome Atlas Research Network. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018; 173:400–16.e11.
<https://doi.org/10.1016/j.cell.2018.02.052> PMID:[29625055](https://pubmed.ncbi.nlm.nih.gov/29625055/)
37. Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Buonamici S, Yu L, and Cancer Genome Atlas Research Network. Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Rep*. 2018; 23:282–96.e4.
<https://doi.org/10.1016/j.celrep.2018.01.088> PMID:[29617667](https://pubmed.ncbi.nlm.nih.gov/29617667/)

SUPPLEMENTARY MATERIALS

Supplementary Figure



Supplementary Figure 1. Flow chart of data processing in this study. In gastric cancer (GC) cohort samples, alternative splicing (AS) events occurring in $\geq 75\%$ of samples were downloaded from the SpliceSeq database. The patient samples with both clinical follow-up and AS event data were enrolled in our study. In result, we conducted a SpliceSeq analysis based on the 415 GC cases. In the survival analysis, 47 patients with an overall survival time of less than 30 days were excluded. The remaining 368 patients were then matched with their corresponding entries in the SpliceSeq database, and 31 cases were excluded once again due to $>20\%$ missing AS events. Thus, 337 patients were included in this study and their overall survival status was listed in Supplementary Table 2. A total of 1383 AS events were found to be significantly associated with the overall survival of GC patients, including 517 exon skip (ES) events, 354 alternate promoter (AP) events, 225 alternate terminator (AT) events, 98 alternate acceptor site (AA) events, 104 alternate donor site (AD) events, 72 retained intron (RI) events, and 13 mutually exclusive exons (ME) events. The functional enrichment analyses were carried out to elucidate the function of genes with survival-associated AS events. Cytoscape software was applied to visualize the splicing-regulatory network of splicing factors and survival-associated AS events to explore the correlation between the expression of splicing factors and the PSI values of AS events. LASSO Cox analysis was carried out to develop seven prognostic signatures based on AA, AD, AP, AT, ES, ME and RI events. Furthermore, these prognostic AS events in seven different types were combined to build the final prognostic signature.



Supplementary Figure 2. Kaplan-Meier and receiver operating characteristic (ROC) curves fitted with the White and Asian gastric cancer (GC) cohorts. Kaplan-Meier plot of the survival probability over time for White (A) and Asian (B) GC cohort with high (red) and low (blue) risk groups, respectively. (C) ROC analysis for White (blue) and Asian (red) GC cohort. The significance level of ROC curves between White and Asian cohort was compared and the p -value was 0.73.

Supplementary Tables

Supplementary Table 1. The overall survival status of the patients.

Supplementary Table 2. Primary tumor characteristics and clinical information.

Variable	Number of samples	Ratio (%)	Valid (%)
Age at diagnosis, y			
≤50	27	8.01	8.11
>50	306	90.80	91.89
Missing	4	1.19	
Tumor type			
Signet Ring Type	9	2.67	5.70
Diffuse Type	58	17.21	36.71
Tubular Type	66	19.58	41.77
Mucinous Type	19	5.64	12.02
Papillary Type	6	1.78	3.80
Not Otherwise Specified (NOS)	179	53.12	
Tumor grade			
1	8	2.37	2.45
2	122	36.20	37.31
3	197	58.46	60.24
Missing	10	2.97	
Gender			
Male	218	64.69	64.88
Female	118	35.01	35.12
Missing	1	0.30	
Race			
White	210	62.32	71.92
Asian	70	20.77	23.97
African American	11	3.26	3.77
Native Hawaiian or other pacific islander	1	0.30	0.34
Missing	45	13.35	
T-stage			
T1	18	5.34	5.42
T2	74	21.96	22.29
T3	154	45.70	46.39
T4	86	25.52	25.90
Missing	5	1.48	
N-stage			
N0	99	29.38	30.46

N1	93	27.60	28.62
N2	69	20.47	21.23
N3	64	18.99	19.69
Missing	12	3.56	
M-stage			
M0	302	89.62	93.50
M1	21	6.23	6.50
Missing	14	4.15	
Stage			
I	47	13.95	14.60
II	106	31.45	32.92
III	135	40.06	41.92
IV	34	10.09	10.56
Missing	15	4.45	
Antireflux treatment			
Yes	33	9.79	20.62
No	127	37.69	79.38
Missing	177	52.52	
Family history of stomach cancer			
Yes	15	4.45	5.62
No	252	74.78	94.38
Missing	70	20.77	
Radiation treatment adjuvant			
Yes	41	12.17	22.65
No	140	41.54	77.35
Missing	156	46.29	
Targeted molecular therapy			
Yes	87	25.82	47.54
No	96	28.48	52.46
Missing	154	45.70	
