Identification of prognostic splicing factors and exploration of their potential regulatory mechanisms in pancreatic adenocarcinoma

Min-hua Rong [1], **Zhan-hui Zhu** [1], **Ying Guan** [2], **Mei-wei Li** [1], **Jia-shuo Zheng** [1], **Yue-qi Huang** [1], **Dan-ming Wei** [3], **Ying-mei Li** [3], **Xiao-ju Wu** [3], **Hui-ping Bu** [1], **Hui-liu Peng** [1], **Xiao-lin Wei** [1], **Guosheng Li** [1], **Ming-xuan Li** [1], **Ming-hui Chen** [1], **Su-ning Huang** [Corresp. 2]

[1] Affiliated Cancer Hospital, Guangxi Medical University, Research Department, Nanning, Guangxi Zhuang Autonomous Region, P.R. China

[2] Affiliated Cancer Hospital, Guangxi Medical University, Department of Radiotherapy, Nanning, Guangxi Zhuang Autonomous Region, P.R. China

[3] First Affiliated Hospital, Guangxi Medical University, Department of Pathology, Nanning, Guangxi Zhuang Autonomous Region, P.R. China

Corresponding Author: Su-ning Huang
Email address: huangsuning2000@163.com

Pancreatic adenocarcinoma (PAAD), the most common subtype of pancreatic cancer, is a highly lethal disease. In this study, we integrated the expression profiles of splicing factors (SFs) of PAAD from RNA-sequencing data to provide a comprehensive view of the clinical significance of SFs. A prognostic index (PI) based on SFs was developed using the least absolute shrinkage and selection operator (LASSO) COX analysis. The PI exhibited excellent performance in predicting the status of overall survival of PAAD patients. We also used the percent spliced in (PSI) value obtained from SpliceSeq software to quantify different types of alternative splicing (AS). The prognostic value of AS events was explored using univariate COX and LASSO COX analyses; AS-based PIs were also proposed. The integration of prognosis-associated SFs and AS events suggested the potential regulatory mechanisms of splicing processes in PAAD. This study defined the markedly clinical significance of SFs and provided novel insight into their potential regulatory mechanisms.

Identification of prognostic splicing factors and exploration of their potential regulatory mechanisms in pancreatic adenocarcinoma

Min-hua Rong[1], Zhan-hui Zhu[1], Ying Guan[2] , Mei-wei Li[1], Jia-shuo Zheng[1], Yue-qi Huang[1], Dan-ming Wei[3], Ying-mei Li[3], Xiao-ju Wu[3], Hui-ping Bu[1], Hui-liu Peng[1], Xiao-lin Wei[1], Guo-sheng Li[1], Ming-xuan Li[1], Ming-hui Chen[1], Su-ning Huang[2]*

1 Research Department, Affiliated Cancer Hospital, Guangxi Medical University, 71 Hedi Road, Nanning, Guangxi Zhuang Autonomous Region, 530021, P.R. China
2 Department of Radiotherapy, Affiliated Cancer Hospital, Guangxi Medical University, 71 Hedi Road, Nanning, Guangxi Zhuang Autonomous Region, 530021, P.R. China
3 Department of Pathology, First Affiliated Hospital, Guangxi Medical University, 6 Shuangyong Road, Nanning, Guangxi Zhuang Autonomous Region, 530021, P.R. China

Correspondence: Su-ning Huang, Department of Radiotherapy, Affiliated Cancer Hospital, Guangxi Medical University, 71 Hedi Road, Nanning, Guangxi Zhuang Autonomous Region, P.R. China. Tel +86 771 533 1466 Fax +86 771 533 1466 Email huangsuning2000@163.com

Abstract

Pancreatic adenocarcinoma (PAAD), the most common subtype of pancreatic cancer, is a highly lethal disease. In this study, we integrated the expression profiles of splicing factors (SFs) of PAAD from RNA-sequencing data to provide a comprehensive view of the clinical significance of SFs. A prognostic index (PI) based on SFs was developed using the least absolute shrinkage and selection operator (LASSO) COX analysis. The PI exhibited excellent performance in predicting the status of overall survival of PAAD patients. We also used the percent spliced in (PSI) value obtained from SpliceSeq software to quantify different types of alternative splicing (AS). The prognostic value of AS events was explored using univariate COX and LASSO COX analyses; AS-based PIs were also proposed. The integration of prognosis-associated SFs and AS

events suggested the potential regulatory mechanisms of splicing processes in PAAD. This study defined the markedly clinical significance of SFs and provided novel insight into their potential regulatory mechanisms.

Keywords: Pancreatic adenocarcinoma, splicing factors, RNA-sequencing, overall survival

Introduction

Pancreatic cancer, the seventh most common cause of cancer-related death worldwide, is a highly lethal disease (Bray et al. 2018; Liang et al. 2018; Wang et al. 2018). According to epidemiological estimates in the United States, approximately 56,770 new pancreatic cancer cases were diagnosed and 45,750 people died from the disease in 2019 (Siegel et al. 2019). Pancreatic adenocarcinoma (PAAD) is the predominant subtype of pancreatic cancer and remains a health priority (Chen et al. 2019; Kamisawa et al. 2016). Current treatments for PAAD include surgery, chemotherapy, radiation therapy, and palliative care; surgery is regarded as the only option for cure. However, most PAAD patients experience no symptoms in the early stages, which precludes surgical removal (Strobel et al. 2019). Hence, molecular biomarkers that can effectively monitor the onset and prognosis of PAAD are indispensable. In addition, the complex mechanisms underlying the development of PAAD remains poorly understood.

Splicing is an important process *in vivo* and is responsible for transcript diversity (Dvinge & Bradley 2015; Kim et al. 2018). Splicing factors (SFs) are a powerful manipulator in modulating RNA processing and maintaining cellular homeostasis (Dvinge et al. 2016). More importantly, intricate splicing events are orchestrated by a limited number of SFs. Many studies have found

48 links between the turbulences of SFs and the onset and progression of cancers (Cieply & Carstens.
49 2015 ; Shilo et al. ; Silipo et al. 2015). In PAAD, SFs also exhibit potential effective functions in
50 many ways. Adesso L et al (Adesso et al. 2013) found that silencing SRSF1, a member of the
51 arginine/serine-rich splicing factor protein family, could facilitate apoptosis induced by
52 gemcitabine via the MNK/eIF4E pathway (Adesso et al. 2013). This finding offers an alternative
53 way to enhance gemcitabine efficiency in PAAD. However, studies with a focus on the functions
54 of SFs in PAAD are still scarce. A comprehensive analysis to determine the clinical value of SFs
55 in PAAD is urgently needed.

56 Here, we systematically analyzed the clinical significance of SFs in PAAD and provided
57 clinically practicable molecular biomarkers. More importantly, a prognostic index (PI) based on
58 the expression profiles of SFs was proposed, which offers excellent survival prediction. Moreover,
59 we also explored the clinical significance of alternative splicing (AS) events. The PI based on AS
60 events also demonstrated a satisfactory prognosis prediction performance. In addition, the SF-AS
61 regulatory network also provides novel insight into the molecular function of SFs in PAAD.

62

63 **Methods**

64 **Data acquisition**

65 A catalog of 404 SF genes was obtained from a previous study (Seiler et al. 2018). The
66 fragments per kilobase of transcript per million mapped reads (FPKM) data of PAAD patients
67 were downloaded from the Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/)
68 database using the TCGAbiolinks R software package (Colaprico et al. 2016). The corresponding

clinical annotation had also been downloaded and extracted from the TCGA database. Gene name annotation was performed using an ensemble database (GRCh38.95). Next, the FPKM expression data were quantified to "transcripts per million" (TPM) data and normalized to the log2 (TPM+1) data type. Then, normalized TPM data was used for subsequent analysis.

Survival analysis

The R package survival outputs were used for univariate COX analysis of selected prognosis-associated SFs. To obtain more accurate results, only PAAD patients with an overall survival (OS) greater than 90 days were included in the survival analysis. Then, we further conducted gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway functional enrichment analysis to reveal the potential molecular functions of prognosis-related SFs. The GO analysis mainly includes biological processes (BPs), cellular components (CCs), and molecular function (MF). The gene functional enrichment analysis was conducted using the "clusterProfiler" package in R software (Yu et al. 2012)

Survival-associated alternative splicing events

SFs performed their molecular function mainly by regulating the AS events process (Papasaikas & Valcárcel 2016). We further systematically analyzed the prognostic value of alterations in AS events in PAAD and the associations between SFs and AS events. Transcript and splicing event details of cross-tumors of TCGA RNA-seq data were downloaded from the TCGA SpliceSeq database (https://bioinformatics.mdanderson.org/TCGASpliceSeq/) (Gao et al. 2019; Lin et al. 2019; Lin et al. 2018; Ryan et al. 2016; Zhang et al. 2018). The SpliceSeq database

quantified the seven AS events, including Alternate Acceptor Site (AA), Alternate Donor Site (AD), Alternate Promoter (AP), Alternate Terminator (AT), Exon Skip (ES), Mutually Exclusive Exons (ME), and Retained Intron (RI), by calculating a percent-splice-in (PSI) value. The PSIs ranged from 0–1. A PSI value of an ES event of 0.8 indicates that the exon is contained in approximately 80% of the transcripts in the sample. We used splice event filters according to the following conditions: 1) Percentage of samples with a PSI > 75% and 2) a minimum PSI standard deviation > 0.1. The missing value was filled using the k-Nearest Neighbor (KNN) method. The KNN was conducted with the Impute package in R software. Next, we integrated the PSI values of AS events and the survival data of PAAD and conducted a univariate COX analysis to identify prognosis-associated AS events. AS events with a P-value < 0.005 were identified as prognosis-associated AS events.

Construction of a PI

To develop a PI based on the expression profiles of SFs genes, a least absolute shrinkage and selection operator (LASSO) was conducted. Any SFs genes with P-values <0.005 were identified as most the significant prognosis-related genes. Then, the selected most significant prognosis-related SFs were further screened and confirmed by the LASSO regression. The classifier was trained using 10-fold cross-validation to determine the optimal parameter configuration. The PI was established with the following formula: PI = expression level of SF 1 * β1 + expression of SF 2 * β2 +… expression of SF n * βn. We generated a risk score for each patient based on the PI. Then, PAAD patients were placed into groups of two according to the median value of PI (Qin et

111 al. 2019).Furthermore, we used another gene expression dataset that were publicly available and
112 reported clinical outcome information to be used as validation cohort. Gene expression matrix of
113 pancreatic tumors patients in GSE62452 dataset was downloaded from the Gene Expression
114 Omnibus (https://www.ncbi.nlm.nih.gov/geo/).

115 Similarly, the top 20 AS values that were closely related to the prognosis (except the number
116 of ME <20) were subjected to a LASSO COX analysis to develop a PI based on AA, AD, AP, AT,
117 ES, ME, and RI, respectively. Then, a final PI was generated by submitting the top 20 AS events
118 for a LASSO COX analysis. The time-dependent incident dynamic ROCs with area under the
119 curve (AUC) values were calculated to estimate the performance of each model (Blanche et al.
120 2013).

121

122 **SF-AS regulatory network**

123 To construct an SF-AS regulatory network and learn more about the PI we proposed, we
124 analyzed the relationships between SFs genes included in the PI and OS associated AS events. Co-
125 expression relationships were identified by Pearson correlation analysis, and the threshold was set
126 to correlation coefficient $r > |0.6|$ with a P-value <0.05.

127

128 **Results**

129 **Identification of prognosis-associated SFs**

130 After removing those with an OS of less than 90 days, 166 total PAAD patients were included
131 in the present study and were comprised of 90 (54.2%) male and 76 (45.8%) female patients. By

integrating 404 SF gene expression profiles and the survival data, we conducted a univariate COX analysis and found 93 SFs genes were correlated with the OS of PAAD patients ($P < 0.05$). The top 20 most significant SFs are listed in Figure 1.

Gene functional enrichment analysis revealed that prognosis-related SFs genes were classified into 61 BPs, 21 CCs, 24 MF, and 3 KEGG pathways. For BPs, the three most significant categories were "RNA splicing," "mRNA processing," and "RNA splicing via transesterification reactions with bulged adenosine as nucleophile" (Figure 2A). For CCs, the three most significant terms were "spliceosomal complex," "small nuclear ribonucleoprotein complex," and "spliceosomal snRNP complex" (Figure 2B). For MF, these genes were mainly involved in "snRNA binding," "mRNA binding," and "pre-mRNA binding" (Figure 2C). Furthermore, we found these SFs genes mainly participated in "spliceosome," "mRNA surveillance pathway," and "RNA transport pathways" (Figure 2D).

Development of a PI based on SFs

We suspected that a gene set could exhibit more accurate survival prediction performance than a single gene. Therefore, we constructed an SF-based PI according to the results of the LASSO COX analysis (Figure 3). This analysis was conducted using the most significant SFs ($P < 0.005$). Finally, 12 SFs were included in the PI, including DDX21, GPATCH3, IGF2BP3, MYEF2, NRIP2, PTBP3, RBM10, RBM14, RBM5, SRPK1, XAB2, and YBX3. The constructed PI based on the 12 SFs = [DDX21 * 0.204800595 + GPATCH3 * (-0.075547356) + IGF2BP3 *0.060551219 + MYEF2 * (-0.16140842) + NRIP2 * (-0.274848438) + PTBP3 *0.217746846 +

RBM10 * (-0.096000129) + RBM14 * (-0.147396111) + RBM5 * (-0.289524669) + SRPK1 *0.031528808 + XAB2 * (-0.051783325) + YBX3 * 0.30845434]. Each patient was generated a PI (Figure 4A). We found that the patients could be separated into two groups with distinct clinical outcomes based on the median PI (Figure 4B). The heatmap also showed that the included SFs were differentially expressed between the high- and low-risk groups (Figure 4C). K-M plots were generated to reveal the survival significance of genes included in the prognostic signature (Figure 5). Based on the SF-based PI median value, PAAD patients could be separated into two groups with distinct clinical outcomes (Figure 6A). The AUC was 0.734 in 3 year (Figure 6B). In the validation cohort, patients in high-risk group suffered poorer survival near to statistical significance (Figure 6C). The AUC was 0.681 in 3 year (Figure 6D).

Identification of prognosis-associated AS events

We obtained 10,354 AS events for the survival analysis, including 656 AA, 705 AD, 3181 AP, 1394 AT, 3494 ES, 62 ME, and 862 RI. We found that the 26 AA, 35AD, 297 AP, 122 AT, 230 ES, 6 ME, and 70 RI events were most significantly correlated with the OS of PAAD patients ($P < 0.005$). LASSO COX analyses were conducted based on the top 20 most significant OS-associated SFs. Seven PIs based on AA, AD, AP, AT, ES, ME, and RI were finally constructed (Figure 7). According to the final PI based on AS events, a PI was generated for each patient (Figure 8A). We found that the patients could be separated into two groups with distinct clinical outcomes based on the median PI (Figure 8B). The heatmap also showed that the included AS events were differentially expressed between the high- and low-risk groups (Figure 8C). The time-

174 dependent ROC of PI based on AS events indicated that the final PI possessed the highest AUC
175 (Figure 9A). The AUCs of SF-based PI, AS-based PI, and TNM are also displayed (Figure 9B).
176
177 **SF-AS regulatory network**
178 AS events are mainly regulated by just a few SFs. Therefore, we decided to explore the
179 prospective regulatory mechanism between SFs and AS events in PAAD. A Pearson correlation
180 analysis was performed and suggested that 33 favorable AS events (blue dots) and 6 risky AS
181 events (red dots) were closely related to the 4 SFs (green dots) (Figure 10).
182 **Discussion**
183 We performed a survival analysis focused on the clinical significance of SFs in PAAD based
184 on one of the largest available cancer genomics datasets to develop an excellent prognostic risk
185 score. Although systematic analyses of somatic mutations, copy numbers, gene expression
186 patterns, and associated AS events have been reported (Neelamraju et al. 2018; Sebestyén et al.
187 2016), many important issues in the field remain unresolved, especially the unique clinical value
188 of SFs in PAAD. Moreover, the AS events related to SFs could also provide novel insight into the
189 molecular function of SFs in PAAD.
190 PAAD is one of the most lethal cancers and causes a high morbidity. Hence, exploration of
191 the impact of multiple molecular biomarkers is crucial for a prognosis evaluation. Previously,
192 several studies have proposed prognostic signatures for survival prediction. Previously, several
193 studies have proposed prognostic signatures for survival prediction. For example, Yu Y et al
194 integrated the miRNA-expression profiles and clinical information of 168 PAAD patients in the

TCGA database and developed a two-microRNA signature for the diagnosis and prognosis assessment.((Yu et al. 2018). Similarly, Shi X et al proposed a three-lncRNA signature for potential survival prediction, and this signature served as an independent prognostic predictor in PAAD.((Shi et al. 2018). ROC of the 3-lncRNA signature was 0.716, which is slightly lower than the present study. Researchers have also provided a five-lncRNAs signature that could act as a potential prognostic indicator for PAAD patients by mining the TCGA database.((Song et al. 2018). The AUC for the six-lncRNA biomarkers prognostic model was 0.727 at 5 years of OS. These findings provide alternative clinically selectable indicators for PAAD surveillance. The prognostic signature we proposed have well performance when compared with previous molecular index. However, the prognostic signatures proposed based on the global alterations of genes could only provide limited information. In the present study, we proposed a risk score that was mainly focused on the expression profiles of SFs in PAAD. Because the roles of SFs in PAAD have not been fully explored, more studies are needed to reveal their clinical significance. New findings about the relationships between SFs and AS events could offer a broader insight into the molecular process of PAAD.

We finally proposed a prognostic signature based on 12 SFs. Interestingly, some of these 12 SFs have been reported in PAAD. Schaeffer DF et al concluded that IGF2BP3 was upregulated in PAAD, and its overexpression indicated poor survival based upon an immunohistochemical analysis of 127 PAAD patients (Schaeffer et al. 2010). This result was consistent with our findings. Subsequent analyses revealed that IGF2BP3 could promote cell invasiveness and the metastasis of pancreatic cancer (Taniuchi et al. 2014). RBM5 has also been proven to be downregulated in

pancreatic cancer, and reduced RBM5 expression has a close association with poor clinicopathological features (Peng et al. 2013). Furthermore, Hayes GM et al. reported that knockdown of the expression levels of SRPK1 in pancreatic tumor cells could decrease the proliferative capacity and increase the apoptotic potential of pancreatic tumor cells (Hayes et al. 2006). These results suggest that SRPK1 could be an effective therapeutic target for pancreatic cancer. Previous reports about the SFs we proposed provide some evidence about their crucial functions. Although previous studies have mentioned their clinical significance and molecular function, a comprehensive exploration is still needed.

The prognosis anticipating value of AS events was widely explored recently for its limitless potential for clinical applications. Several studies have attempted to investigate the prognostic value of AS events in several types of cancer. For example, some researchers have explored the prognostic value of AS events in lung cancer (Li et al. 2017). This groundbreaking research pointed out the well-known value of AS events. Subsequently, researchers found that AS exhibited an effective prognosis prediction value in thyroid cancer (Lin et al. 2019), gastrointestinal pan-adenocarcinomas (Lin et al. 2018), and diffuse large B-cell lymphoma (Zhang et al. 2018). Algorithmically, the established prognostic models were based on the PSI value. This value is a useful method for quantifying AS events and demonstrating its clinical value. To the best of our knowledge, we are the first group to integrate the clinical parameters and PSI values of AS events. In this study, we also constructed an SF-AS potential regulatory network, which provides the underlying mechanisms of SFs in PAAD. Indeed, many survival associated splicing events has been validated. For example, previous study has reported that VEGFA/76336/ES significantly

237 associated poor survival in pancreatic cancer (Zhang et al. 2017). Furthermore, many SFs have
238 been validated important in splicing regulation and regulate the processes of tumors. RBM5 could
239 promote exon 4 skipping of AID pre-mRNA (Jin et al. 2012). YBX3 was found to be related to
240 spliceosomes in large-scale spliceosome capture and mass spectrometry analyses(Rappsilber et al.
241 2002).

242 As the present study was based on an *in silico* analysis, there are several inevitable limitations
243 that should be mentioned. First, no other independent study, especially a prospective study, has
244 validated the prognostic signatures we proposed. Second, the clinical parameters related to the
245 prognosis of PAAD have not been fully investigated.

246 **Conclusions**

247 In conclusion, we systematically explored the clinical significance of SFs in PAAD patients.
248 More importantly, a prognostic signature based on the prognosis-associated SFs was constructed
249 to separate PAAD patients into two groups with distinct clinical outcomes. These findings could
250 provide more information about the clinical value of SFs. The SF-AS regulatory network provided
251 information regarding the molecular functions of SFs.

252

253 **Acknowledgments**

254 The study was supported by the Funds of National Natural Science Foundation of China
255 (NSFC81860596), Guangxi Natural Scientific Research (2016GXNSFAA380022), Future
256 Academic Star of Guangxi Medical University (WLXSZX19118), and Guangxi Zhuang
257 Autonomous Region University Student Innovative Plans (2018090, 201910598418). The authors

sincerely appreciate the public access to TCGA database.

Figure Legends

Figure 1. The top 20 most significant survival-associated splicing factors.

Figure 2. Gene ontology and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of survival associated splicing factors.

"RNA splicing," "mRNA processing," and "RNA splicing via transesterification reactions with bulged adenosine as a nucleophile" are the three most significant biological process terms. (B) "Spliceosomal complex," "small nuclear ribonucleoprotein complex," and "spliceosomal snRNP complex" are the three most significant cellular component terms. (C) The three most significant molecular function terms were "snRNA binding," "mRNA binding," and "pre-mRNA binding." (D) The KEGG pathway analysis revealed that these genes were mainly involved in "spliceosome," "mRNA surveillance pathway," and "RNA transport."

Figure 3. Construction of the prognostic index based on the most significant survival-associated splicing factor genes (P <0.005) using the LASSO COX regression model.

(A) The LASSO coefficient profiles of the splicing factors. A vertical line is drawn at the value chosen by 10-fold cross-validation.

(B) Tuning parameter (λ) selection cross-validation error curve. The vertical lines were drawn at

279 the optimal values by the minimum criteria and the 1-SE criteria.

280

281 **Figure 4. The development of a prognostic index based on splicing factors.**

282 (A) Distribution of a risk score by the prognostic signature based on splicing factors. (B) The
283 patients were separated into two groups with distinct survival statuses according to the median
284 value of the risk score. (C) A heatmap shows the expression profiles of the included genes in the
285 high- and low-risk groups.

286

287 **Figure 5. Kaplan-Meier survival plots showed the clinical significance of the splicing**
288 **factors included in the PI.**

289 (A) DDX21, (B) NRIP2, (C) RBM5, (D) GPATCH3, (E) PTBP3, (F) SRPK1, (G) IGF2BP3, (H)
290 RBM10, (I) XAB2, (J) MYEF2, (K) RBM14, and (L) YBX3.

291

292 **Figure 6: The survival prediction performance of the prognostic index.**

293 Kaplan-Meier survival plots suggested that patients in the high-risk group could expect a poor
294 survival in the TCGA database (A) and GSE62452 (C). ROC curves with calculated AUCs of
295 prognostic signatures built in the PAAD cohort for risk prediction over 3 years in the TCGA
296 database (B) and GSE62452 (D).

297 **Figure 7. Kaplan-Meier survival plots showed the stratification of the prognostic index based**
298 **on alternative splicing events.**

299 (A) Acceptor Site, (B) Alternate Donor Site, (C) Alternate Promoter, (D) Alternate Terminator,

(E) Exon Skip, (F) Mutually Exclusive Exons, (G) Retained Intron, and (H) all types of AS events.

Figure 8. The development of a PI based on alternative splicing events.

(A) Distribution of the risk score by the prognostic signature based on alternative splicing events. (B) The patients were separated into two groups with distinct survival statuses by the median value of the risk score. (C) A heatmap displays the expression profiles of the included genes in the high- and low-risk groups.

Figure 9. Time-dependent receiver operating characteristic (ROC) curves of the survival prediction systems.

(A) The area under the curve (AUC) of the ROC at 500, 1000, 1500, and 2000 days, respectively. The AUC of the prognostic index that was based on all types of alternative splicing events was the highest.

(B) The AUC of the ROC at 500, 1000, 1500, and 2000 days, respectively. The AUC of the prognostic index that was based on all types of alternative splicing events was the highest.

Figure 10. Prognostic splicing factors and the splicing correlation network in PAAD.

The construction of an SF-AS regulatory network. Green dots represent the SFs; red dots indicate risky alternative splicing events, and blue dots represent protective events.

References

Adesso L, Calabretta S, Barbagallo F, Capurso G, Pilozzi E, Geremia R, Delle Fave G, and Sette C. 2013. Gemcitabine triggers a pro-survival response in pancreatic cancer cells through activation of the MNK2/eIF4E pathway. *Oncogene* 32:2848-2857. 10.1038/onc.2012.306

Blanche P, Dartigues JF, and Jacqmin-Gadda H. 2013. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 32:5381-5397. 10.1002/sim.5958

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68:394-424. 10.3322/caac.21492

Chen H, Kong Y, Yao Q, Zhang X, Fu Y, Li J, Liu C, and Wang Z. 2019. Three hypomethylated genes were associated with poor overall survival in pancreatic cancer patients. *Aging (Albany NY)* 11:885-897. 10.18632/aging.101785

Cieply B, and Carstens RP. 2015. Functional roles of alternative splicing factors in human disease. *Wiley Interdiscip Rev RNA* 6:311-326. 10.1002/wrna.1276

Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, and Noushmehr H. 2016. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 44:e71. 10.1093/nar/gkv1507

Dvinge H, and Bradley RK. 2015. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med* 7:45. 10.1186/s13073-015-0168-9

Dvinge H, Kim E, Abdel-Wahab O, and Bradley RK. 2016. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* 16:413-430. 10.1038/nrc.2016.51

Gao L, Xie ZC, Pang JS, Li TT, and Chen G. 2019. A novel alternative splicing-based prediction model for uteri corpus endometrial carcinoma. *Aging (Albany NY)* 11:263-283. 10.18632/aging.101753

Hayes GM, Carrigan PE, Beck AM, and Miller LJ. 2006. Targeting the RNA splicing machinery as a novel treatment strategy for pancreatic carcinoma. *Cancer Res* 66:3819-3827.

Jin W, Niu Z, Xu D, and Li X. 2012. RBM5 promotes exon 4 skipping of AID pre-mRNA by competing with the binding of U2AF65 to the polypyrimidine tract. FEBS Lett 586:3852-3857. 10.1016/j.febslet.2012.09.006

Kamisawa T, Wood LD, Itoi T, and Takaori K. 2016. Pancreatic cancer. *Lancet* 388:73-85. 10.1016/s0140-6736(16)00141-0

Kim HK, Pham MHC, Ko KS, Rhee BD, and Han J. 2018. Alternative splicing isoforms in health and disease. *Pflugers Arch* 470:995-1016. 10.1007/s00424-018-2136-x

Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z, and He J. 2017. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer Lett* 393:40-51.

Liang L, Wei DM, Li JJ, Luo DZ, Chen G, Dang YW, and Cai XY. 2018. Prognostic microRNAs and their potential molecular mechanism in pancreatic cancer: A study based on The Cancer Genome Atlas and bioinformatics investigation. *Mol Med Rep* 17:939-951. 10.3892/mmr.2017.7945

Lin P, He RQ, Huang ZG, Zhang R, Wu HY, Shi L, Li XJ, Li Q, Chen G, Yang H, and He Y. 2019. Role of global aberrant alternative splicing events in papillary thyroid cancer prognosis. *Aging (Albany NY)* 11:2082-2097. 10.18632/aging.101902

359 Lin P, He RQ, Ma FC, Liang L, He Y, Yang H, Dang YW, and Chen G. 2018. Systematic Analysis of Survival-
360 Associated Alternative Splicing Signatures in Gastrointestinal Pan-Adenocarcinomas. *EBioMedicine* 34:46-
361 60.
362 Neelamraju Y, Gonzalez-Perez A, Bhat-Nakshatri P, Nakshatri H, and Janga SC. 2018. Mutational landscape of RNA-
363 binding proteins in human cancers. *RNA Biol* 15:115-129. 10.1080/15476286.2017.1391436
364 Papasaikas P, and Valcárcel J. 2016. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem*
365 *Sci* 41:33-45.
366 Peng J, Valeshabad AK, Li Q, and Wang Y. 2013. Differential expression of RBM5 and KRAS in pancreatic ductal
367 adenocarcinoma and their association with clinicopathological features. *Oncol Lett* 5:1000-1004.
368 Qin XG, Zeng JH, Lin P, Mo WJ, Li Q, Feng ZB, Luo DZ, Yang H, Chen G, and Zeng JJ. 2019. Prognostic value of
369 small nuclear RNAs (snRNAs) for digestive tract pan- adenocarcinomas identified by RNA sequencing data.
370 *Pathol Res Pract* 215:414-426.
371 Rappsilber J, Ryder U, Lamond AI, and Mann M. 2002. Large-scale proteomic analysis of the human spliceosome.
372 Genome Res 12:1231-1245. 10.1101/gr.473902
373 Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, Melott J, and Weinstein J. 2016. TCGASpliceSeq a
374 compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res* 44:D1018-1022.
375 10.1093/nar/gkv1288
376 Schaeffer DF, Owen DR, Lim HJ, Buczkowski AK, Chung SW, Scudamore CH, Huntsman DG, Ng SS, and Owen
377 DA. 2010. Insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3) overexpression in pancreatic
378 ductal adenocarcinoma correlates with poor survival. *BMC Cancer* 10:59. 10.1186/1471-2407-10-59
379 Sebestyén E, Singh B, Miñana B, Pagès A, Mateo F, Pujana MA, Valcárcel J, and Eyras E. 2016. Large-scale analysis
380 of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks.
381 *Genome Res* 26:732-744. 10.1101/gr.199935.115
382 Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Buonamici S, and Yu L. 2018. Somatic
383 Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types.
384 *Cell Rep* 23:282-296.e284.
385 Shi X, Zhao Y, He R, Zhou M, Pan S, Yu S, Xie Y, Li X, Wang M, Guo X, and Qin R. 2018. Three-lncRNA signature
386 is a potential prognostic biomarker for pancreatic adenocarcinoma. *Oncotarget* 9:24248-24259.
387 10.18632/oncotarget.24443
388 Shilo A, Siegfried Z, and Karni R. The role of splicing factors in deregulation of alternative splicing during
389 oncogenesis and tumor progression. *Mol Cell Oncol* 2:e970955. 10.4161/23723548.2014.970955
390 Siegel RL, Miller KD, and Jemal A. 2019. Cancer statistics, 2019. *CA Cancer J Clin* 69:7-34. 10.3322/caac.21551
391 Silipo M, Gautrey H, and Tyson-Capper A. 2015. Deregulation of splicing factors and breast cancer development. *J*
392 *Mol Cell Biol* 7:388-401. 10.1093/jmcb/mjv027
393 Song J, Xu Q, Zhang H, Yin X, Zhu C, Zhao K, and Zhu J. 2018. Five key lncRNAs considered as prognostic targets
394 for predicting pancreatic ductal adenocarcinoma. *J Cell Biochem* 119:4559-4569. 10.1002/jcb.26598
395 Strobel O, Neoptolemos J, Jäger D, and Büchler MW. 2019. Optimizing the outcomes of pancreatic cancer surgery.
396 *Nat Rev Clin Oncol* 16:11-26. 10.1038/s41571-018-0112-1
397 Taniuchi K, Furihata M, Hanazaki K, Saito M, and Saibara T. 2014. IGF2BP3-mediated translation in cell protrusions
398 promotes cell invasiveness and metastasis of pancreatic cancer. *Oncotarget* 5:6832-6845.
399 Wang X, Song Z, Chen F, Yang X, Wu B, Xie S, Zheng X, Cai Y, Chen W, and Zhong Z. 2018. AMPK-related kinase

5 (ARK5) enhances gemcitabine resistance in pancreatic carcinoma by inducing epithelial-mesenchymal transition. *Am J Transl Res* 10:4095-4106.

Yu G, Wang LG, Han Y, and He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16:284-287. 10.1089/omi.2011.0118

Yu Y, Feng X, and Cang S. 2018. A two-microRNA signature as a diagnostic and prognostic marker of pancreatic adenocarcinoma. *Cancer Manag Res* 10:1507-1515. 10.2147/cmar.S158712

Zhang L, Wang H, Li C, Zhao Y, Wu L, Du X, and Han Z. 2017. VEGF-A/Neuropilin 1 Pathway Confers Cancer Stemness via Activating Wnt/beta-Catenin Axis in Breast Cancer Cells. Cell Physiol Biochem 44:1251-1262. 10.1159/000485455

Zhang R, Lin P, Yang X, He RQ, Wu HY, Dang YW, Gu YY, Peng ZG, Feng ZB, and Chen G. 2018. Survival associated alternative splicing events in diffuse large B-cell lymphoma. *Am J Transl Res* 10:2636-2647.

Figure 1

The top 20 most significant survival-associated splicing factors.

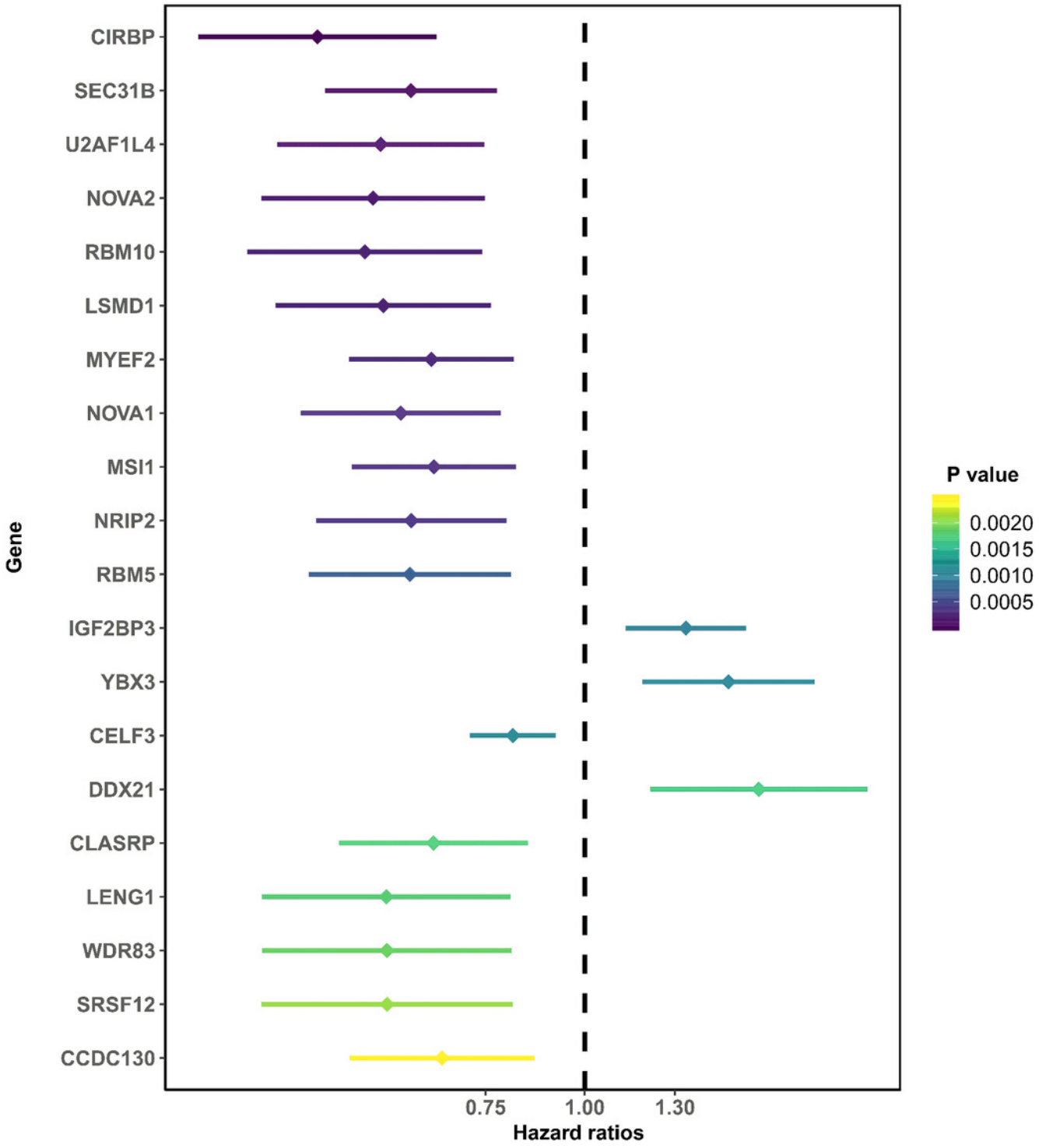

CIRBP
SEC31B
U2AF1L4
NOVA2
RBM10
LSMD1
MYEF2
NOVA1
MSI1
NRIP2
RBM5
IGF2BP3
YBX3
CELF3
DDX21
CLASRP
LENG1
WDR83
SRSF12
CCDC130
Gene
0.75
1.00
1.30
Hazard ratios
P value
0.0020
0.0015
0.0010
0.0005

Figure 2

Gene ontology and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of survival associated splicing factors.

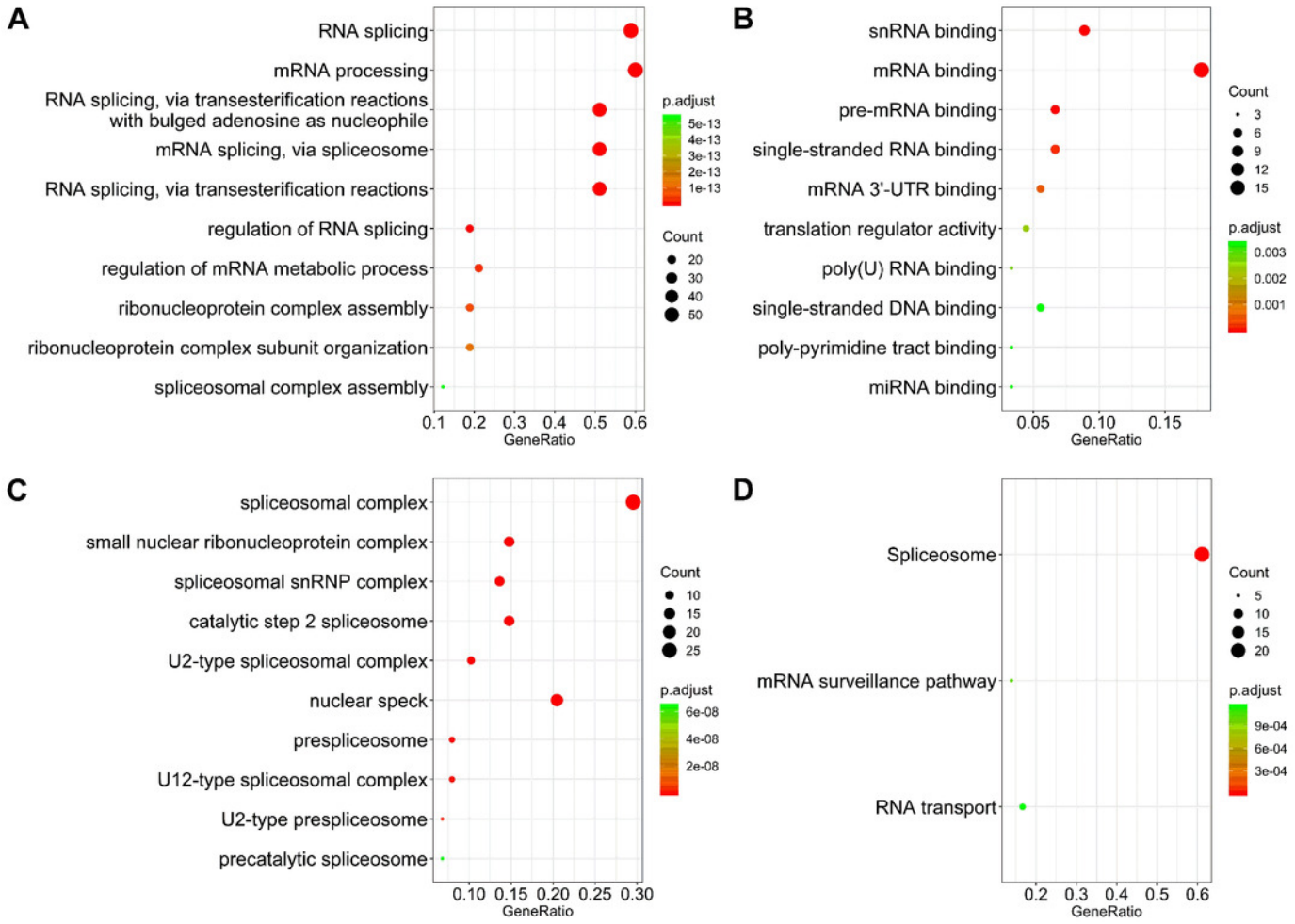


Figure 3

Construction of the prognostic index based on the most significant survival-associated splicing factor genes (P <0.005) using the LASSO COX regression model.

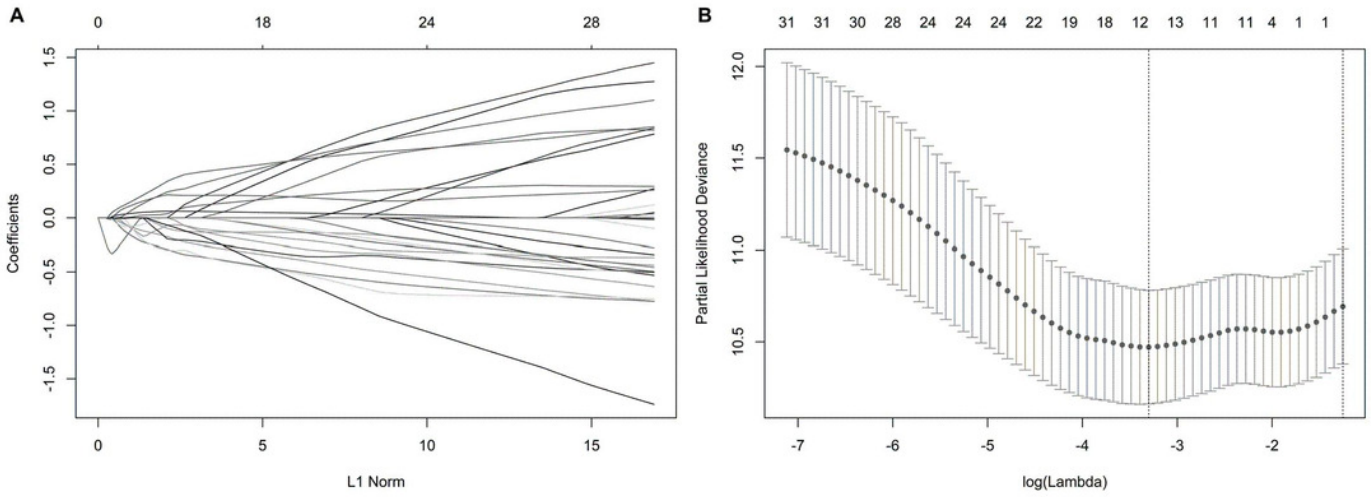


Figure 4

The development of a prognostic index based on splicing factors.

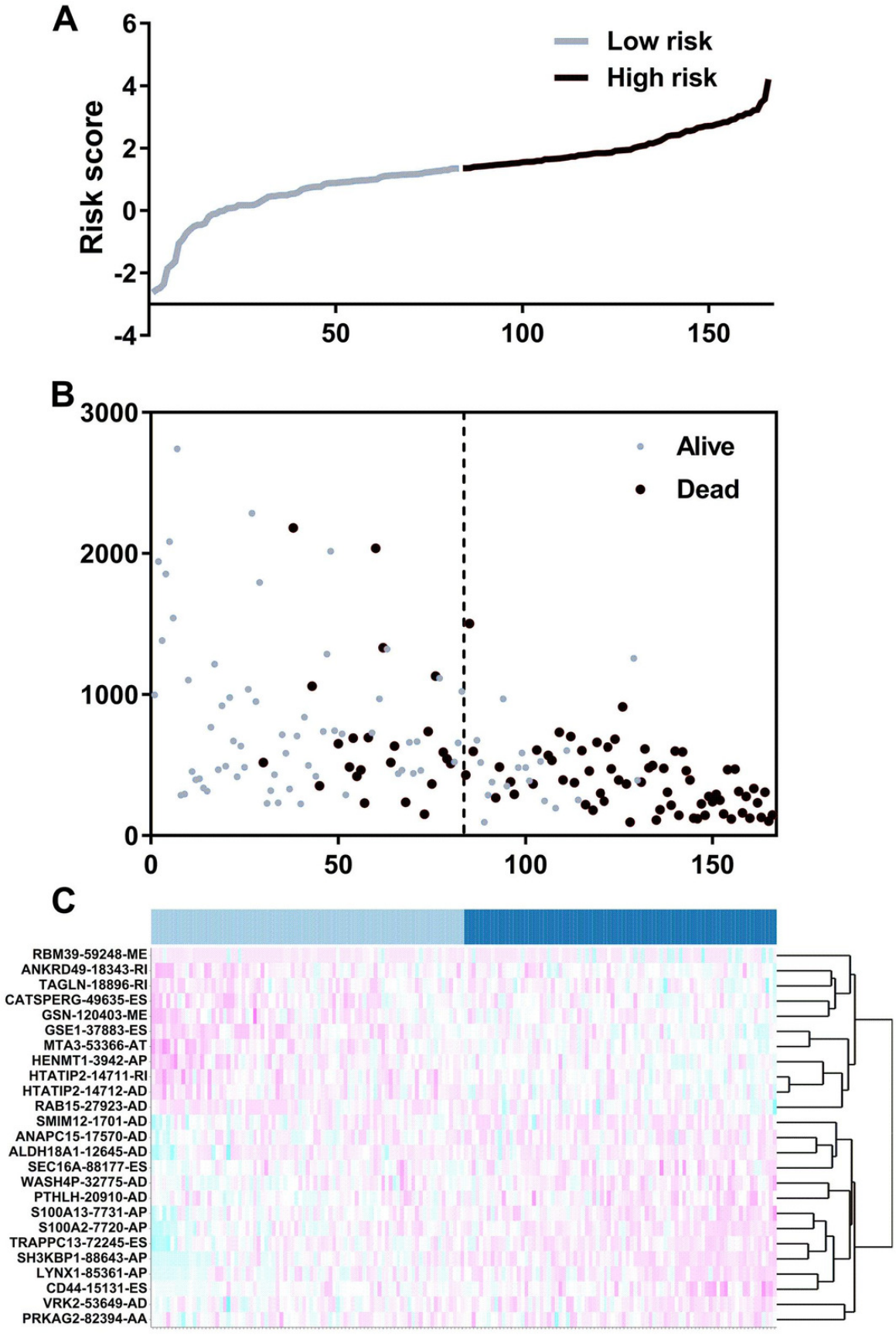

A
Risk score
6
4
2
0
-2
-4
50
100
150
Low risk
High risk
B
3000
2000
1000
0
0
50
100
150
Alive
Dead
C
RBM39-59248-ME
ANKRD49-18343-RI
TAGLN-18896-RI
CATSPERG-49635-ES
GSN-120403-ME
GSE1-37883-ES
MTA3-53366-AT
HENMT1-3942-AP
HTATIP2-14711-RI
HTATIP2-14712-AD
RAB15-27923-AD
SMIM12-1701-AD
ANAPC15-17570-AD
ALDH18A1-12645-AD
SEC16A-88177-ES
WASH4P-32775-AD
PTHLH-20910-AD
S100A13-7731-AP
S100A2-7720-AP
TRAPPC13-72245-ES
SH3KBP1-88643-AP
LYNX1-85361-AP
CD44-15131-ES
VRK2-53649-AD
PRKAG2-82394-AA

Figure 5

Kaplan-Meier survival plots showed the clinical significance of the splicing factors included in the PI.

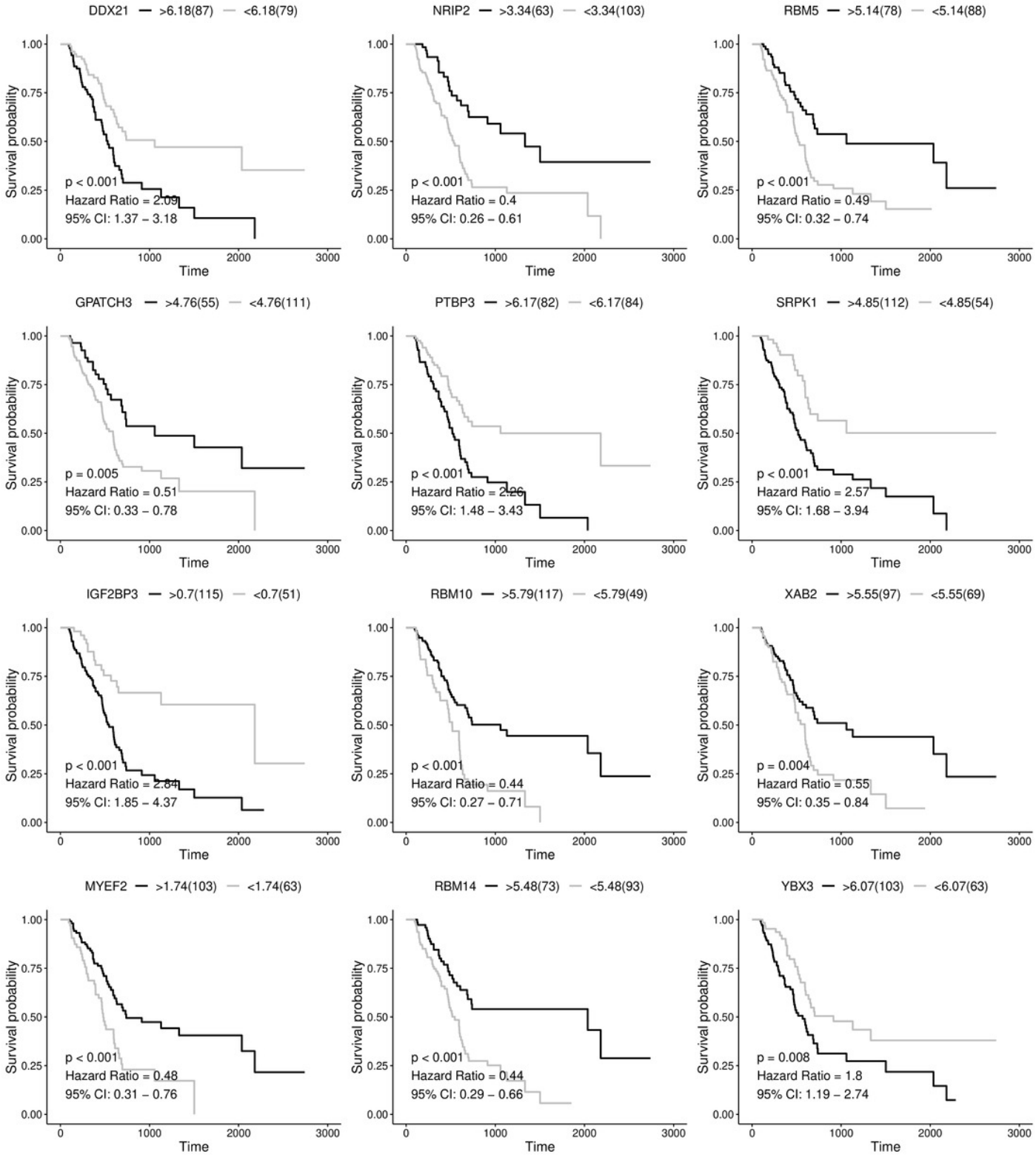

DDX21 >6.18(87) <6.18(79)
p < 0.001
Hazard Ratio = 2.09
95% CI: 1.37 – 3.18
NRIP2 >3.34(63) <3.34(103)
p < 0.001
Hazard Ratio = 0.4
95% CI: 0.26 – 0.61
RBM5 >5.14(78) <5.14(88)
p < 0.001
Hazard Ratio = 0.49
95% CI: 0.32 – 0.74
GPATCH3 >4.76(55) <4.76(111)
p = 0.005
Hazard Ratio = 0.51
95% CI: 0.33 – 0.78
PTBP3 >6.17(82) <6.17(84)
p < 0.001
Hazard Ratio = 2.26
95% CI: 1.48 – 3.43
SRPK1 >4.85(112) <4.85(54)
p < 0.001
Hazard Ratio = 2.57
95% CI: 1.68 – 3.94
IGF2BP3 >0.7(115) <0.7(51)
p < 0.001
Hazard Ratio = 2.84
95% CI: 1.85 – 4.37
RBM10 >5.79(117) <5.79(49)
p < 0.001
Hazard Ratio = 0.44
95% CI: 0.27 – 0.71
XAB2 >5.55(97) <5.55(69)
p = 0.004
Hazard Ratio = 0.55
95% CI: 0.35 – 0.84
MYEF2 >1.74(103) <1.74(63)
p < 0.001
Hazard Ratio = 0.48
95% CI: 0.31 – 0.76
RBM14 >5.48(73) <5.48(93)
p < 0.001
Hazard Ratio = 0.44
95% CI: 0.29 – 0.66
YBX3 >6.07(103) <6.07(63)
p = 0.008
Hazard Ratio = 1.8
95% CI: 1.19 – 2.74
Survival probability
Time

Figure 6

The survival prediction performance of the prognostic index.

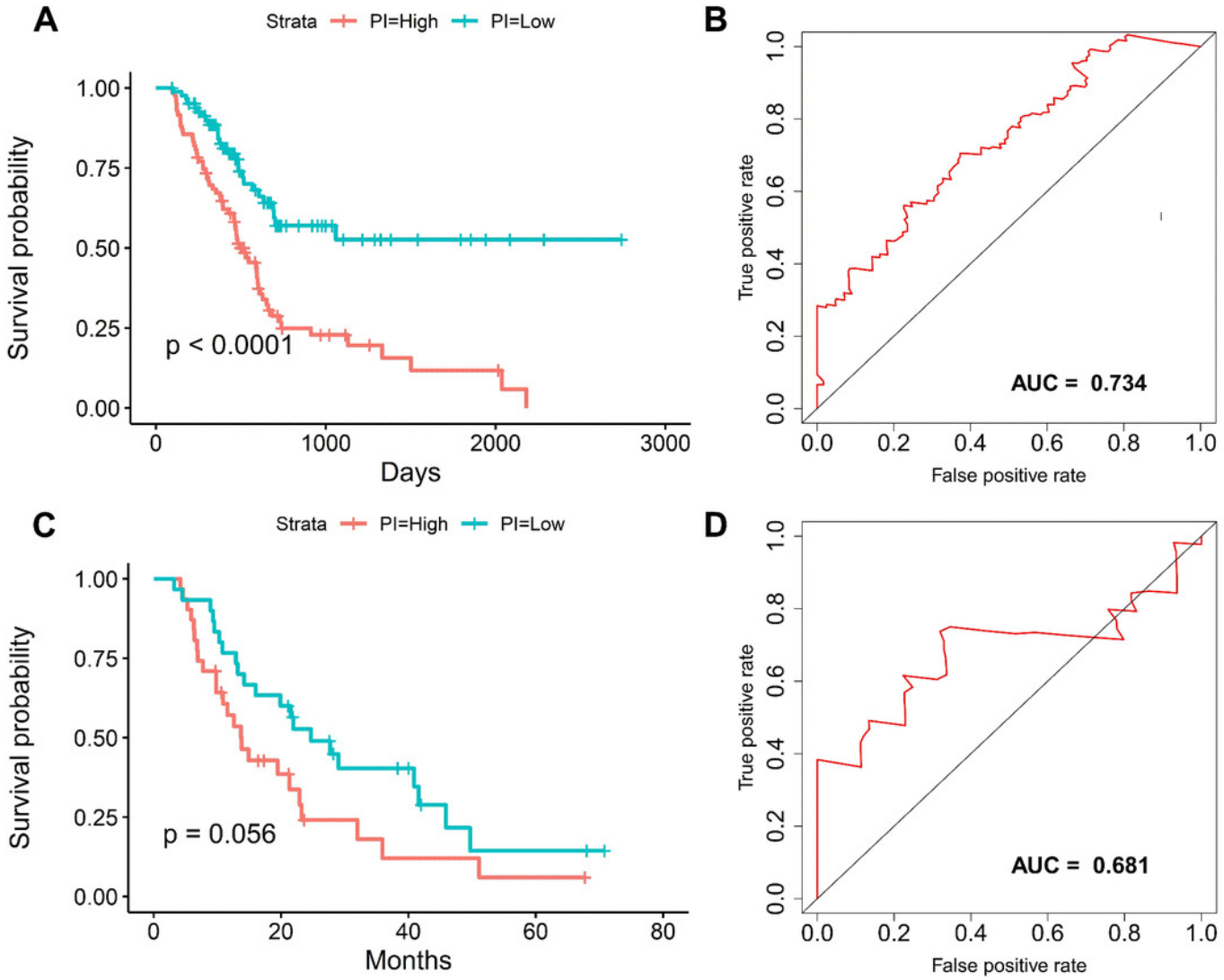


Figure 7

Kaplan-Meier survival plots showed the stratification of the prognostic index based on alternative splicing events

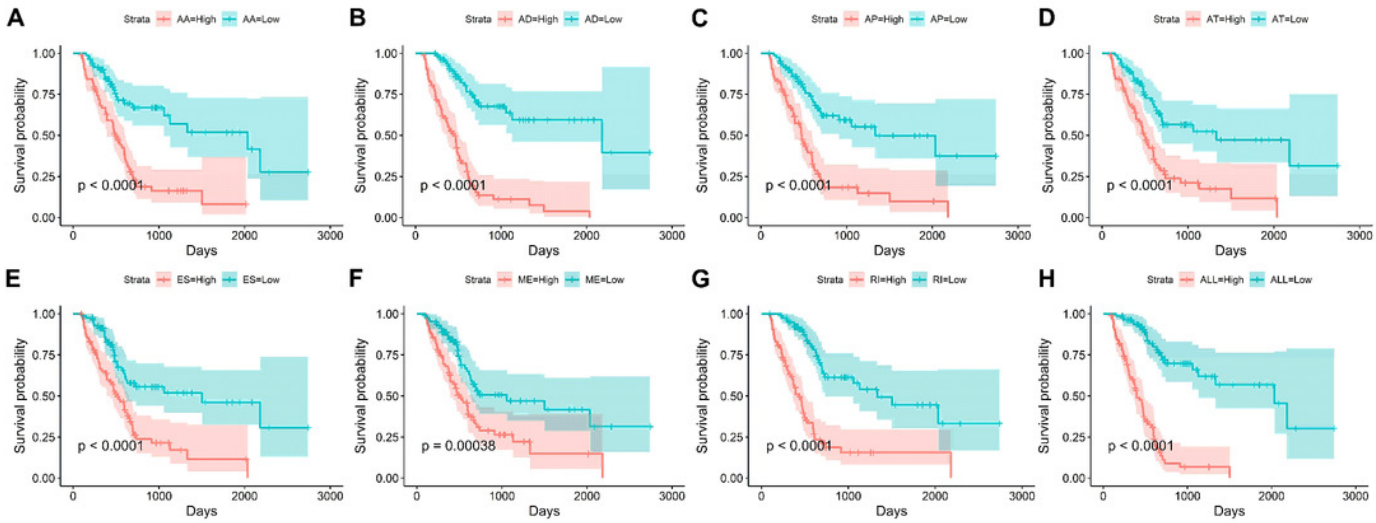


Figure 8

The development of a PI based on alternative splicing events.

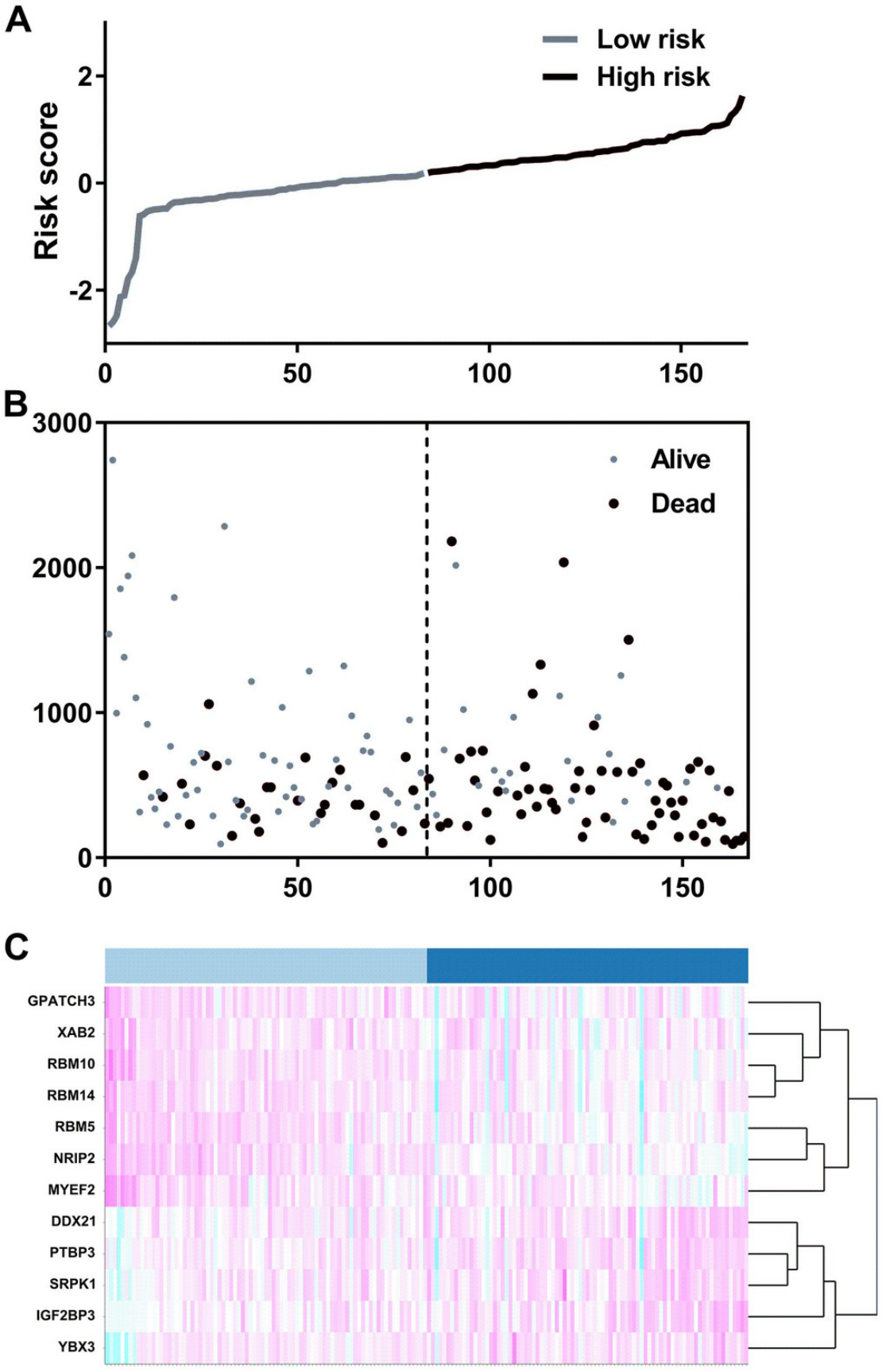
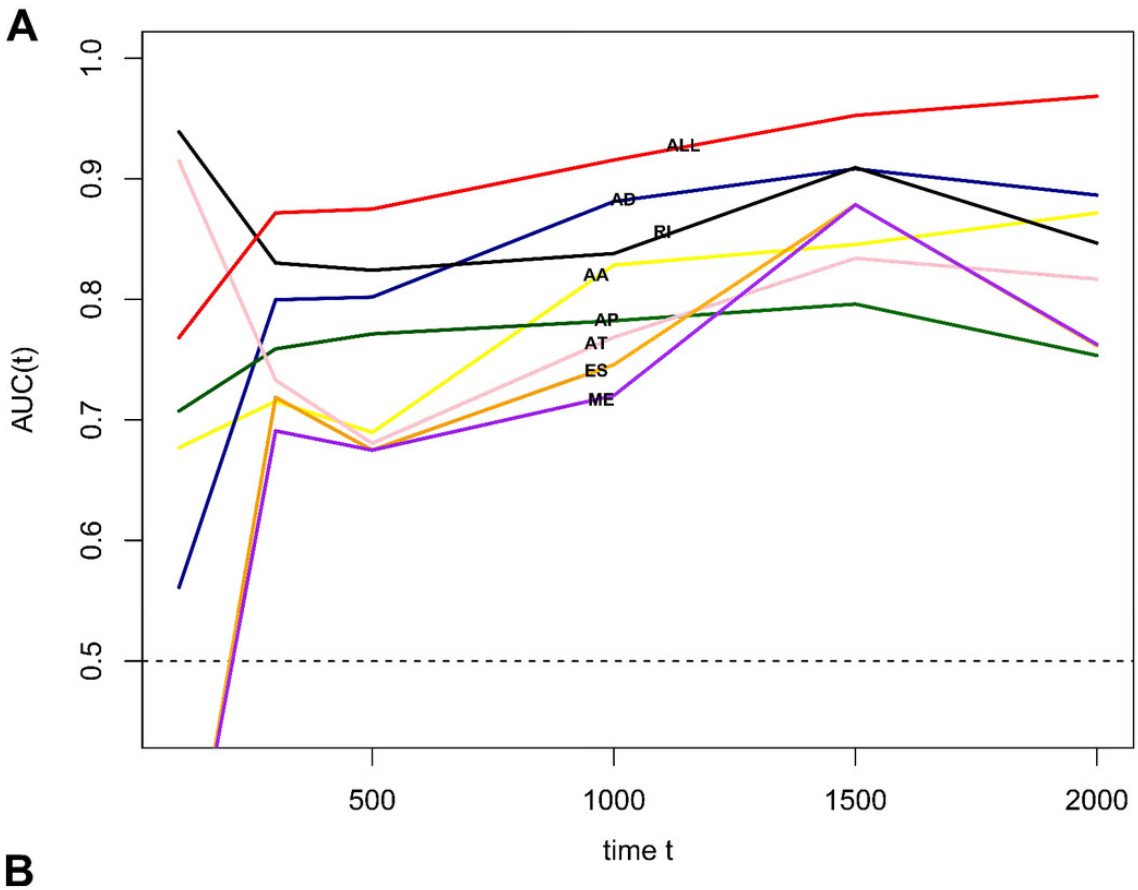

A
Low risk
High risk
Risk score
2
0
-2
0
50
100
150
B
3000
2000
1000
0
Alive
Dead
0
50
100
150
C
GPATCH3
XAB2
RBM10
RBM14
RBM5
NRIP2
MYEF2
DDX21
PTBP3
SRPK1
IGF2BP3
YBX3

Figure 9

Time-dependent receiver operating characteristic (ROC) curves of the survival prediction systems.


A
AUC(t)
1.0
0.9
0.8
0.7
0.6
0.5
ALL
AD
RI
AA
AP
AT
ES
ME
500
1000
1500
2000
time t

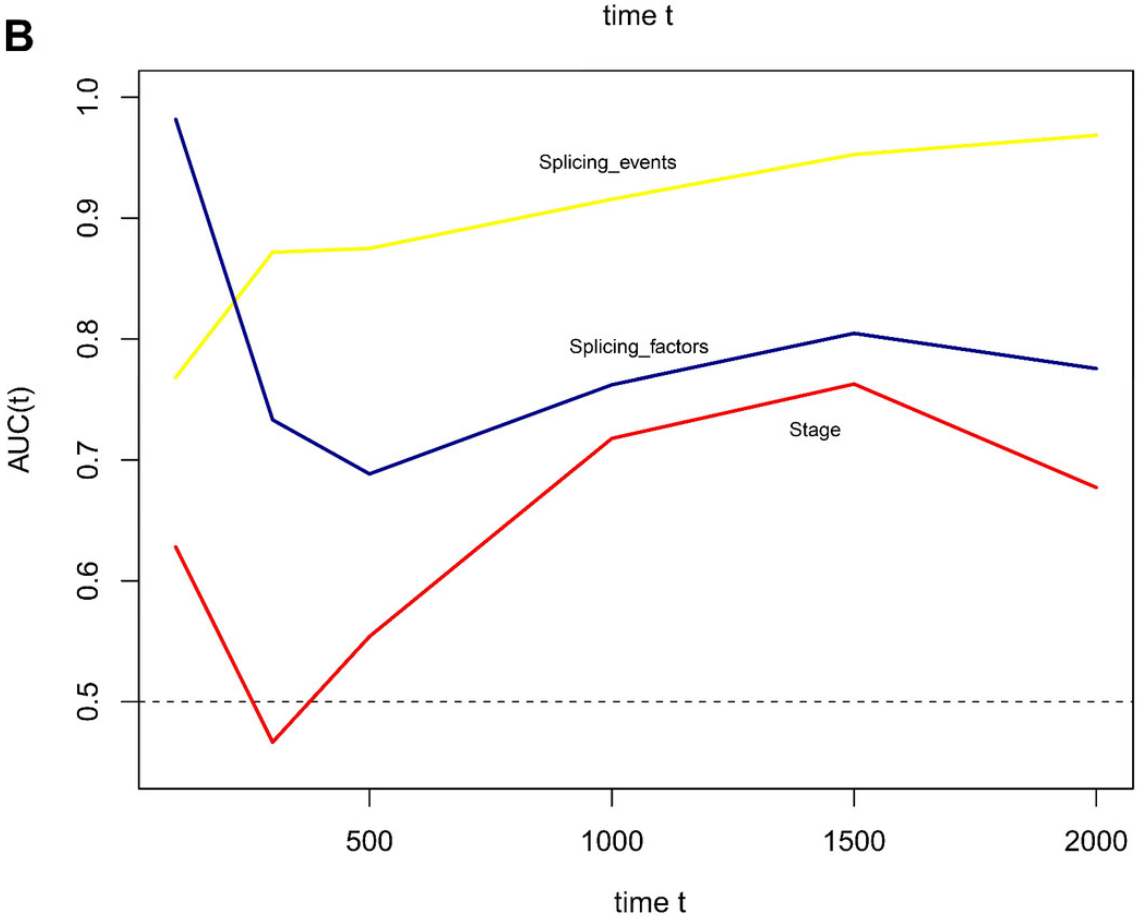

B
AUC(t)
1.0
0.9
0.8
0.7
0.6
0.5
Splicing_events
Splicing_factors
Stage
500
1000
1500
2000
time t

Figure 10

Prognostic splicing factors and the splicing correlation network in PAAD.