Features of alternative splicing in stomach adenocarcinoma and their clinical implication: A research based on massive sequencing data

Yuanyuan Zhang (u201310506@hust.edu.cn)
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
https://orcid.org/0000-0001-9778-745X
Qian Niu
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Yun Han
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Xingyu Liu
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Jie Jiang
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Simiao Chen
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Haolong Lin
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Research article

Keywords: bioinformatic analysis, alternative splicing, stomach adenocarcinoma, survival, splicing factor, prognosis

Posted Date: January 3rd, 2020

DOI: https://doi.org/10.21203/rs.2.20011/v1

Version of Record: A version of this preprint was published on August 24th, 2020. See the published version at https://doi.org/10.1186/s12864-020-06997-x.

Abstract

Background: Alternative splicing (AS) offers a main mechanism to form protein polymorphism. A growing body of evidence indicates the correlation between splicing disorders and carcinoma. Nevertheless, an overall analysis of AS signatures in stomach adenocarcinoma (STAD) is absent and urgently needed.

Methods: Within this work, genetic expression and clinical data of STAD were queried from The Cancer Genome Atlas (TCGA), and profiles of AS events were searched from the SpliceSeq database. Cox regression analysis found survival associated AS events. Finally, the splicing network was constructed to reflect the correlation between survival associated AS events and splicing factors (SF).

Results: 2042 splicing events were confirmed as prognostic molecular events. Furthermore, the final prognostic signature constructed by 10 AS events gave good result with an area under the curve (AUC) of receiver operating characteristic (ROC) curve up to 0.902 for 5 years, showing high potency in predicting patient outcome. We built the splicing regulatory network to show the internal regulation mechanism of splicing events in STAD. QKI may play a significant part in the prognosis induced by splicing events.

Conclusions: In our study, a high-efficiency prognostic prediction model was built for STAD patients, and the results showed that AS events could become potential prognostic biomarkers for STAD. Meanwhile, QKI may become an important target for drug design in the future.

1. Introduction

Gastric cancer (GC) is the fourth major cancer threat to human health in the world whose etiology remains unclear, with 989,000 new cases and 738,000 deaths every year [1]. Most (about 90%) of gastric cancers are adenocarcinomas, which originate from the glands of the most superficial layer or the mucosa of the stomach and is caused by malignant changes in gastric gland cells. Although morbidity and mortality are on the decline, and significant progress has been made in the study of epidemiology, pathological mechanisms and treatment options, the medical burden still exceeds expectations [2]. It is not optimistic that complete surgical resection is still the only solution for doctors to treat gastric cancer [3, 4]. The widespread implementation and application of adjuvant and neoadjuvant therapy have increased the 5-year overall survival rate by 10–15%, but it is worth thinking that there is no global consensus on the best treatment scheme [2]. Therefore, it is urgent to explore new and accurate biomarkers to evaluate the diagnosis and prognosis of STAD patients.

Eukaryotic cells produce various regulatory changes and perform complex functions to adapt to changes in the environment, largely due to the diversity of proteins. A common mechanism is that a limited number of gene sets produce a large number of mRNA isoforms through alternative splicing of pre-mRNA [5]. Alternative splicing actually regulates gene expression at the intron/exon level [6–8]. In addition, alternative splicing causes the premature occurrence of termination codon in mRNA, which degrades immediately upon discovery to prevent its translation [9]. Therefore, alternative splicing is a key biological

process in cells, and different mRNA splicing isoforms make the final protein products perform different functions.

More and more studies have found that splicing disorder can be used as a marker of tumor development [10] and as a key mechanism involved in the broad biological process of cancer [11, 12]. It is noteworthy that some important splicing factors can change the alternative splicing mode of target genes, thus forming a favorable environment for promoting the occurrence and development of cancer [13]. In general, comprehensive and in-depth analysis of alternative splicing can dig out potential biomarkers of malignant tumors, so as to assist physicians in clinical diagnosis and prognosis judgment [12, 14].

We used a variety of bioinformatics analysis methods to explore prognostic factors in STAD. COX regression analysis helped us screen out significant prognostic markers for further study. According to the regulatory relationship between AS events and splicing factors in STAD, a clear network diagram was drawn to find out the potential mechanism. These results provide a basic direction for further exploration of the molecular mechanism of STAD and exploration of STAD diagnostic markers.

2. Materials And Methods

2.1. Data acquisition

TCGA SpliceSeq [15] is a data portal that provides AS profiles across 33 tumors based on TCGA RNA-seq data. SpliceSeq evaluates seven types of splice events, including alternate acceptor (AA), alternate donor (AD), alternate promoter (AP), alternate terminator (AT), exon skip (ES), mutually exclusive exon (ME) and retained intron (RI). TCGA SpliceSeq processed the percent spliced in (PSI) value for cancer research analysis, which indicates the inclusion of a transcript element divided by the total number of reads for that AS event. Alterations in PSI values range from 0 to 100 (%), which suggests a shift in splicing events. AS events with a PSI value of more than 75% in stomach adenocarcinoma cohort samples were downloaded from the TCGA SpliceSeq database. The AS events with standard diversion < 1 were removed.

Clinical information of STAD patients was also obtained from the TCGA database. Only pathologically confirmed STAD patients with both follow-up and AS event data were included for our analysis. The same TCGA ID was used to integrate clinical information and AS events data.

2.2. Survival analysis

In the survival analysis, the follow-up periods ranged from 90 days to 3720 days after removal of patients with survival less than 90 days. Univariate Cox analysis was conducted to assess the correlations between the PSI value (from 0 to 100) of each AS event and the survival data of STAD patients ($P < 0.05$). The least absolute shrinkage and selection operator (LASSO) method is a widely used regression analysis method of high-dimensional predictors [16]. LASSO has been extended for use in Cox regression survival analysis and is ideal for high- dimensional data. We selected the LASSO Cox regression model to

determine the accurate coefficient for each prognostic feature and to estimate the deviance likelihood via 1-standard error (SE) criteria. The coefficients and partial likelihood deviance were calculated with the “glmnet” package in R.

2.3. Prognostic signature construction

The significant AS events in univariate Cox analysis were submitted to LASSO regression Cox analysis to develop prognostic signatures based on seven types of AS events. Finally, prognostic signatures for survival prediction were calculated by multiplying the PSI values of prognostic indictors and the coefficient assigned by LASSO Cox analysis. By incorporating the following parameters into multivariate Cox regression analysis, splicing-based prognostic signature was evaluated as independent predictors: age, gender, grade, stage, TMN stage.

2.4. SF-AS regulatory network

A compendium of 404 splicing factors was obtained from a previous study [17]. The expression profiles of SF genes were curated from the TCGA dataset. We selected axes between the expression value of SFs and PSI values of prognosis-related AS events to construct the SF-AS regulatory network according to the following conditions: P value less than 0.001 and the absolute value of Pearson's correlation coefficient more than 0.6. Then, we built the correlation plots via Cytoscape version 3.7.1.

3. Results

3.1. Survival associated AS events

As a whole, there are 4006 AA events in 2799 genes, 3450 AD events in 2401 genes, 10004 AP events in 4025 genes, 8390 AT events in 3666 genes, 19121 ES events in 6973 genes, 226 ME events in 219 genes, and 2944 RI events in 1956 genes for evaluation of prognostic value (Fig. 1A). A total of 157 AA events in 153 genes, 174 AD events in 164 genes, 461 AP events in 304 genes, 297 AT events in 203 genes, 805 ES events in 660 genes, 18 ME events in 18 genes, and 130 RI events in 113 genes were identified as prognostic AS events ($P < 0.05$) (Fig. 1B). Thus, one gene might have two or more AS events that were markedly related to the survival of STAD patients. The ES which was vividly revealed by the UpSet plot was the most common prognosis-related event, and a gene could have up to seven prognosis-related events (Fig. 1C).

3.2. Molecular characteristics of survival related AS events

The distributions of AS events significantly correlated with patient survival are displayed in Fig. 2A. The 20 most significant prognosis-related AS events are also shown (Fig. 2B-H). To reveal the molecular characteristics of genes with survival-associated AS events, several bioinformatics analyses were conducted. First, a protein-protein interaction (PPI) network was constructed to demonstrate the relationships among these genes. UBA52, STAT3 and PLK4 ranked at the core in the network (Fig. 3). According to the functional annotations, "intracellular transport", "negative regulation of biological

process" and "negative regulation of cellular process" were the three most significant biological process terms (Fig. 4A). “intracellular part”, “intracellular” and “intracellular organelle” were the three most significant cellular component terms (Fig. 4B). For molecular function, “binding”, “enzyme binding” and “protein binding” were three most enriched categories (Fig. 4C).

3.3. Prognostic signatures for STAD patients

By applying the LASSO Cox analysis following univariate Cox, we developed seven types of prognostic signatures based on AA, AD, AP, AT, ES, ME and RI (Fig. 5, Table 1). Interestingly, we found that the seven prognostic signatures could predict the clinical outcome of STAD patients (Fig. 6). ROC curves validated the performance of prognostic signatures in prognosis prediction (Fig. 7). Figure 8 shows the patient's survival status and risk score, as well as the splicing pattern of AS signatures for each AS type or a combination of seven AS types. In univariate Cox analysis, the risk Score we constructed was correlated with prognosis (Fig. 9). After multivariate adjustment by clinical factors, the prognostic signature remained a moderate and independent prognostic indicator (Fig. 10).

Table 1

Prognostic signatures for STAD.

Type	Formula	Hazard ratio (95%CI)	AUC
AA	ST5-14270-AA*(-5.351641) + PLK4-70545-AA*(-50.075071) + BDKRB2-29192-AA*(-2.251191) + NAT6-64990-AA*2.558247 + APOBEC3B-62269-AA*(-58.382634) + ECT2-67658-AA*(-9.149068) + MORF4L2-89778-AA*(-2.400196) + STAT3-41041-AA*3.052117 + PARPBP-24045-AA*(-3.451740) + CBX7-62286-AA*5.771950 + TRAPPC2L-38043-AA*5.892537 + C19orf60-48492-AA*(-2.837545) + DHPS-47831-AA*(-3.486833) + TROAP-21565-AA*2.928574 + ZNF410-28332-AA*13.901426 + HNRNPR-1047-AA*2.702840	1.034(1.021 − 1.047)	0.843
AD	UBA52-48486-AD*(-3.817624) + PHRF1-13700-AD*(-2.535869) + TCIRG1-17286-AD*(-15.912327) + CCDC51-64653-AD*(-1.376775) + HMBS-19096-AD*(-1.858043) + SNX27-7647-AD*(-10.039140) + RAP1B-22959-AD*(-5.215765) + SERPINA3-29154-AD*(-4.624829) + RPS6KA4-16651-AD*14.672505 + MBD4-66720-AD*(-3.903928) + NKG7-51322-AD*(-13.306358) + RALGPS1-87614-AD*(-4.395804) + TMEM106C-21404-AD*(-5.264665) + FAH-32181-AD*(-0.998076) + SMIM19-83739-AD*(-1.958156) + HYI-2185-AD*5.019022	1.175(1.136 − 1.215)	0.841
AP	KIAA1147-82046-AP*(-26.932098) + CDKN3-27569-AP*(-16.593603) + WEE1-14328-AP*(-4.589803) + RCAN1-60494-AP*0.979582 + MID1-88465-AP*3.529023 + TUBA1A-21538-AP*7.031309 + PDZK1IP1-2893-AP*(-77.071138)	1.072(1.047 − 1.099)	0.71
AT	PPHLN1-21214-AT*4.712066 + ABCB5-78909-AT*1.406777 + TLN2-30978-AT*(-3.649182) + TET2-70188-AT*2.736038 + MFSD2B-52798-AT*(-4.976167) + BRSK1-52060-AT*(-3.402968) + ZC3H12D-78076-AT*1.091190 + IL7R-71774-AT*(-4.704211) + ZNF846-47399-AT*1.093000 + ZFYVE28-68559-AT*(-1.498791)	1.183(1.130 − 1.239)	0.77
ES	CD44-14986-ES*(-5.431100) + RASSF4-11351-ES*(-15.542342) + PPP2R5D-76200-ES*(-9.142504) + LOH12CR1-20507-ES*(-14.620773) + CBWD3-86515-ES*2.727266 + GNPDA2-69151-ES*(-13.336665) + EIF3K-49681-ES*(-5.515884) + CLEC4A-20178-ES*(-4.701354) + FANCA-38149-ES*(-378.778875) + ZNF106-30164-ES*(-34.393642) + NME6-64602-ES*(-23.597673) + PAOX-13555-ES*(-2.845972) + DYNC2H1-18489-ES*(-7.577542) + CYP2B6-50020-ES*(-1.561095) + TMX2-15906-ES*(-4.527987) + D2HGDH-58425-ES*(-16.861342) + DUSP22-75134-ES*(-16.679166)	1.037(1.028 − 1.046)	0.816
ME	ANK3-11852-ME*0.617693 + AMT-64866-ME*9.363879 + C4orf21-70379-ME*8.722888 + MCFD2-102349-ME*(-2.786243) + CS-22420-ME*1.961533 + MTHFSD-102413-ME*3.738460 + KDM6A-98323-ME*(-1.285022) + ATE1-91855-ME*(-2.737012) + USP10-37863-ME*1.511545 + RAB6A-17707-ME*2.671001 + GRB10-79717-ME*1.261053 + ITGB1-126615-ME*1.273771	1.636(1.436 − 1.863)	0.781
RI	SRSF7-53276-RI*(-1.721712) + RPS15-46490-RI*(-2.387697) + DTD2-27118-RI*(-6.362278) + DUSP18-61796-RI*(-5.322993) + ADRA2C-68651-RI*(-5.026148) + LGALS3BP-43940-RI*(-2.549496) + BICD2-86883-RI*(-1.137099) + ALS2CL-64462-RI*1.664472 + THAP7-61211-RI*(-1.872850) + KAT5-16914-RI*(-1.868716)	1.138(1.097 − 1.180)	0.902

Type	Formula	Hazard ratio (95%CI)	AUC
All	CD44-14986-ES*(-6.883160) + PPHLN1-21214-AT*5.396354 + RASSF4-11351-ES*(-14.293402) + KIAA1147-82046-AP* (-22.278949) + PPP2R5D-76200-ES*(-6.035515) + LOH12CR1-20507-ES*(-10.075293) + CDKN3-27569-AP*(-18.653016) + UBA52-48486-AD*(-3.222755) + CADPS-65499-AT*(-1.957771) + SRSF7-53276-RI* (-2.217781) + WEE1-14328-AP*(-5.279043)	1.043(1.030 – 1.057)	0.882

3.4. Survival-associated SF-AS network

Because events are primarily orchestrated by SFs that often bind with pre-mRNAs and regulate RNA splicing via influencing exon selection and splicing site. Therefore, exploration of the SF-AS regulatory network is imperative in STAD. Next, correlation analyses between the SFs' expression and the most significant AS events' PSI value ($P < 0.001$) were conducted (Fig. 11A). We observed that QKI was most significantly connected in the network, so we compared the influence of QKI expression on STAD's survival rate. The consequence showed that low QKI expression significantly improved the survival rate of patients with STAD, and the five-year survival rate of the patients with low QKI expression was almost twice that of the patients with high QKI expression (Fig. 11B, Fig. 11C).

4. Discussion

Currently, scientific research on the role of AS events in STAD still has many unanswered questions owing to the dearth of available large-sample public AS profiles and the paucity of systematic analysis referring to their clinical significance and deep molecular function. These bottlenecks have prevented cancer researchers from effectively recognizing the widespread applicability of AS events in STAD. Exploration of AS patterns broadens our vision and our understanding in traditional transcriptome molecular biomarkers. In this project, we adopted several biomedical computational approaches, which integrate the AS event profiles and clinical information of STAD patients to mine prognosis-related AS and construct splicing prognostic signatures that could stratify STAD patients into subgroups with distinct survival outcomes. Moreover, the SF-AS network could provide further insights into regulatory mechanisms in patients with STAD from the perspective of splicing.

Gastric cancer is a highly heterogeneous malignant tumor. Therefore, single drug is not significantly useful for various types of gastric cancer. Classical cytotoxic therapy cannot be fully effective because of the presence of patients resistant to specific drugs. At present, the diagnosis and treatment of gastric cancer rely on histopathological diagnosis and definite classification. Therefore, in addition to targeted treatment with trastuzumab, we need to develop new targeted drugs to provide better treatment for patients. Potential biomarkers can be mined and used to predict patient outcomes, and treatment strategies can be developed for specific tumor parameters.

The next-generation sequencing technology developed in recent years adopts the whole- genome sequencing method, which has great advantages in exploring alternative splicing. Previously, several studies conducted SpliceSeq analyses to generate alterative splicing profiles for some types of cancer, as well as to construct prognostic signatures for cancer prognosis monitoring, including non-small cell lung cancer [18], colorectal cancer [19], and esophageal cancer [20]. This computational bioinformatics analysis could open up different perspectives on the clinical application and potential pathological mechanism of AS on a macro level. Previously, several studies have proposed transcriptomic signatures related with epithelial-to-mesenchymal transition and diagnosis of gastric cancer [21, 22]. The present in-depth study further explored alterations of transcriptomes used as prognostic predictors and could broaden our horizons in the clinical significance of transcriptomic signatures.

Given the multitude of AS events impacted by their own pre-mRNAs, the downstream functional impact is partly used to describe the molecular function of AS alteration events. In the PPI network analysis, UBA52, STAT3 and PLK4 were the hub genes. Previous studies have shown that UBA52 and STAT3 are all considered to be related molecules involved in the biological process of STAD. For example, bioinformatics analysis has verified the correlation between UBA52 and GC progress and metastasis [23]. STAT3 signaling drives transcription activation of EZH2 and mediates poor prognosis in gastric cancer [24]. STAT3 promotes the increased expression of lncRNA HAGLROS, which leads to further progress of gastric cancer [25]. These findings also pave the way for future clinical applications. Functional enrichment analysis showed that in STAD, the main molecular function of AS event gene related to prognosis is to bind to GTPase, so it may provide selective advantages for cancer cells by regulating GTPase. Increased RhoA activity leads to poorer survival outcomes for the Lauren diffuse type of gastric adenocarcinoma (DGA), and inhibition of RhoA can correct the drug resistance of DGA [26]. RacGAP1 is closely associated with malignant progression and poor survival [27]. Leptin promotes GC migration through the Rho/ROCK mechanism [28]. RASSF6 partially regulates the effect of mir-181a-5p on GC progression through MAKP pathway [29]. It is worth considering that, in gastric cancer cells, RhoA promotes cell proliferation and RhoC stimulates cell migration and invasion, while RhoB functions contrary to RhoA and/or RhoC [30]. Therefore, targeted GTPase therapy is also being explored. For example, ALEX1 functions in gastric cancer through the PAR-1/Rho GTPase signaling pathway, becoming a new target for tumor inhibition [31]. RhoA-mediated Fbxw7 regulates the apoptosis of tumor cells and other phenotypes in gastric cancer [32]. Similarly, Gastrokine 1's inhibition of gastric cancer progression may also be dependent on RhoA [33]. Our findings suggest that a group of AS events play a biological role in the alteration of GTPase in STAD.

The highlight of the current study was that we proposed prognostic signatures based on AS events for monitoring the prognosis of STAD patients. Recently, some prognostic signatures in STAD have been proposed. Zhang H et al. found that the efficacy of postoperative adjuvant chemotherapy for gastric cancer was affected by the degree of neutrophil infiltration of the tumor [34]. Jiang Y et al. developed an immune score GC classifier that can effectively predict the recurrence and survival of patients with gastric cancer, which plays a good role in complementation of the prognosis judgment for the TNM staging system [35]. The clinical management of STAD patients still needs to be improved, and the above

mentioned molecular biomarkers have broad prospects. In order to facilitate clinical practice, we selected a group of AS events using the LASSO Cox regression model, and the prognostic model proposed on which showed satisfactory results. Of course, this also requires a separate cohort for validation.

A large number of AS events are programmed by finite SFs in cells [36]. The altered profile of AS events in multiple tumor types emphasizes the important mechanism of splicing factors in cancer which is disordered splicing [37]. It is increased believed that changes of SFs in STAD can be involved in tumorigenesis and progression through various mechanisms [38–40]. The splicing correlation network analysis has also found out the larger regulated nodes, indicating that they occupy a significant position in the SF-AS network. QKI, which is recognized as a tumor suppressor in a wide range of cancers, is highly connected in the network, which can play a significant part in the prognosis induced by splicing events [41–43]. But the role of QKI in STAD has not been fully discussed yet. Our study indicates that the level of QKI expression is significantly correlated with the survival rate of patients with SATD, and it can become an important target for drug design in the future. Nevertheless, our algorithm suggested deregulated AS events as a hallmark of STAD. However, there are some limitations inevitably affecting the reliability of the study. Firstly, we didn't use a separate cohort for more validation. Secondly, more functional experiments are needed to further investigate the impact of dysregulated AS events and SFs on carcinogenesis.

In conclusion, the current study has found out a phenomenological relationship between AS events and prognosis in STAD patients, which is the base of unscrambling the functional contribution of AS events in STAD. These findings are conducive to develop new genomic models for clinical cancer management. In addition, the further identification of predictive splicing factors for prognosis and the construction of SF-AS networks will pave the way for further exploration of splicing related mechanisms.

Declarations

Declaration of interest

The authors declare no potential conflicts of interest.

Acknowledgments

The authors would like to thank the TCGA Spliceseq and TCGA databases for the availability of the data.

Abbreviations

Alternative splicing (AS)

Stomach adenocarcinoma (STAD)

The Cancer Genome Atlas (TCGA)

Splicing factor (SF)

Area under the curve (AUC)

Receiver operating characteristic (ROC)

Gastric cancer (GC)

Alternate acceptor (AA)

Alternate donor (AD)

Alternate promoter (AP)

Alternate terminator (AT)

Exon skip (ES)

Mutually exclusive exon (ME)

Retained intron (RI)

Percent spliced in (PSI)

The least absolute shrinkage and selection operator (LASSO)

Standard error (SE)

Protein-protein interaction (PPI)

The Lauren diffuse type of gastric adenocarcinoma (DGA)

Funding

No funding was received.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Author contributions

YZ conceived of and designed the study. YZ, QN, YH and XL performed the literature search, generated the figures and tables, and wrote the manuscript. YZ, QN, YH and JJ collected and analyzed the data, and critically reviewed the manuscript. YZ, XL, SC and HL supervised the study and reviewed the manuscript.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics**. *CA Cancer J Clin* 2011, **61**(2):69-90.
2. Cutsem EV, Sagaert X, Topal B, Haustermans K, Prenen H: **Gastric cancer**. *Lancet* 2016, **388**(26):2654-2664.
3. Van Cutsem E, Dicato M, Geva R, Arber N, Bang Y, Benson A, Cervantes A, Diaz-Rubio E, Ducreux M, Glynne-Jones R *et al*: **The diagnosis and management of gastric cancer: expert discussion and recommendations from the 12th ESMO/World Congress on Gastrointestinal Cancer, Barcelona, 2010**. *Ann Oncol* 2011, **22 Suppl 5**:v1-9.
4. Lutz MP, Zalcberg JR, Ducreux M, Ajani JA, Allum W, Aust D, Bang YJ, Cascinu S, Holscher A, Jankowski J *et al*: **Highlights of the EORTC St. Gallen International Expert Consensus on the primary therapy of gastric, gastroesophageal and oesophageal cancer - differential treatment strategies for subtypes of early gastroesophageal cancer**. *Eur J Cancer* 2012, **48**(16):2941-2953.
5. Nilsen TW, Graveley BR: **Expansion of the eukaryotic proteome by alternative splicing**. *Nature* 2010, **463**(7280):457-463.
6. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes**. *Nature* 2008, **456**(7221):470-476.
7. Salton M, Misteli T: **Small Molecule Modulators of Pre-mRNA Splicing in Cancer Therapy**. *Trends Mol Med* 2016, **22**(1):28-37.
8. Wahl MC, Will CL, Luhrmann R: **The spliceosome: design principles of a dynamic RNP machine**. *Cell* 2009, **136**(4):701-718.
9. Ge Y, Porse BT: **The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression**. *Bioessays* 2014, **36**(3):236-243.
10. Song X, Zeng Z, Wei H, Wang Z: **Alternative splicing in cancers: From aberrant regulation to new therapeutics**. *Semin Cell Dev Biol* 2018, **75**:13-22.
11. David CJ, Manley JL: **Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged**. *Genes Dev* 2010, **24**(21):2343-2364.

12. Oltean S, Bates DO: **Hallmarks of alternative splicing in cancer**. *Oncogene* 2014, **33**(46):5311-5318.
13. 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* 2016, **35**(19):2413-2427.
14. Ladomery M: **Aberrant alternative splicing is another hallmark of cancer**. *Int J Cell Biol* 2013, **2013**:463786.
15. 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**(D1):D1018-1022.
16. Tibshirani R: **The lasso method for variable selection in the cox model**. *Statistics In Medicine* 1997, **16**:385-395.
17. Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Cancer Genome Atlas Research N, Buonamici S, Yu L: **Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types**. *Cell Rep* 2018, **23**(1):282-296 e284.
18. 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.
19. Xiong Y, Deng Y, Wang K, Zhou H, Zheng X, Si L, Fu Z: **Profiles of alternative splicing in colorectal cancer and their clinical significance: A study based on large-scale sequencing data**. *EBioMedicine* 2018, **36**:183-195.
20. Mao S, Li Y, Lu Z, Che Y, Sun S, Huang J, Lei Y, Wang X, Liu C, Zheng S *et al*: **Survival-associated alternative splicing signatures in esophageal carcinoma**. *Carcinogenesis* 2019, **40**(1):121-130.
21. Xu B, Bai Z, Yin J, Zhang Z: **Global transcriptomic analysis identifies SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer**. *PeerJ* 2019, **7**:e7091.
22. Cui J, Chen Y, Chou WC, Sun L, Chen L, Suo J, Ni Z, Zhang M, Kong X, Hoffman LL *et al*: **An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer**. *Nucleic Acids Res* 2011, **39**(4):1197-1207.
23. Tian X, Ju H, Yang W: **An ego network analysis approach identified important biomarkers with an association to progression and metastasis of gastric cancer**. *J Cell Biochem* 2019, **120**(9):15963-15970.
24. Pan YM, Wang CG, Zhu M, Xing R, Cui JT, Li WM, Yu DD, Wang SB, Zhu W, Ye YJ *et al*: **STAT3 signaling drives EZH2 transcriptional activation and mediates poor prognosis in gastric cancer**. *Mol Cancer* 2016, **15**(1):79.
25. Chen JF, Wu P, Xia R, Yang J, Huo XY, Gu DY, Tang CJ, De W, Yang F: **STAT3-induced lncRNA HAGLROS overexpression contributes to the malignant progression of gastric cancer cells via mTOR signal-mediated inhibition of autophagy**. *Mol Cancer* 2018, **17**(1):6.
26. Yoon C, Cho SJ, Aksoy BA, Park DJ, Schultz N, Ryeom SW, Yoon SS: **Chemotherapy Resistance in Diffuse-Type Gastric Adenocarcinoma Is Mediated by RhoA Activation in Cancer Stem-Like Cells**. *Clin Cancer Res* 2016, **22**(4):971-983.

27. Saigusa S, Tanaka K, Mohri Y, Ohi M, Shimura T, Kitajima T, Kondo S, Okugawa Y, Toiyama Y, Inoue Y *et al*: **Clinical significance of RacGAP1 expression at the invasive front of gastric cancer**. *Gastric Cancer* 2015, **18**(1):84-92.
28. Dong Z, Fu S, Xu X, Yang Y, Du L, Li W, Kan S, Li Z, Zhang X, Wang L *et al*: **Leptin-mediated regulation of ICAM-1 is Rho/ROCK dependent and enhances gastric cancer cell migration**. *Br J Cancer* 2014, **110**(7):1801-1810.
29. Mi Y, Zhang D, Jiang W, Weng J, Zhou C, Huang K, Tang H, Yu Y, Liu X, Cui W *et al*: **miR-181a-5p promotes the progression of gastric cancer via RASSF6-mediated MAPK signalling activation**. *Cancer Lett* 2017, **389**:11-22.
30. Zhou J, Zhu Y, Zhang G, Liu N, Sun L, Liu M, Qiu M, Luo D, Tang Q, Liao Z *et al*: **A distinct role of RhoB in gastric cancer suppression**. *Int J Cancer* 2011, **128**(5):1057-1068.
31. Pang L, Li JF, Su L, Zang M, Fan Z, Yu B, Wu X, Li C, Yan M, Zhu ZG *et al*: **ALEX1, a novel tumor suppressor gene, inhibits gastric cancer metastasis via the PAR-1/Rho GTPase signaling pathway**. *J Gastroenterol* 2018, **53**(1):71-83.
32. Li H, Wang Z, Zhang W, Qian K, Xu W, Zhang S: **Fbxw7 regulates tumor apoptosis, growth arrest and the epithelial-to-mesenchymal transition in part through the RhoA signaling pathway in gastric cancer**. *Cancer Lett* 2016, **370**(1):39-55.
33. Yoon JH, Choi WS, Kim O, Choi BJ, Nam SW, Lee JY, Park WS: **Gastrokine 1 inhibits gastric cancer cell migration and invasion by downregulating RhoA expression**. *Gastric Cancer* 2017, **20**(2):274-285.
34. Zhang H, Liu H, Shen Z, Lin C, Wang X, Qin J, Qin X, Xu J, Sun Y: **Tumor-infiltrating Neutrophils is Prognostic and Predictive for Postoperative Adjuvant Chemotherapy Benefit in Patients With Gastric Cancer**. *Ann Surg* 2018, **267**(2):311-318.
35. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, Ye G, Deng H, Mou T, Cai S *et al*: **ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer**. *Ann Surg* 2018, **267**(3):504-513.
36. Lee Y, Rio DC: **Mechanisms and Regulation of Alternative Pre-mRNA Splicing**. *Annu Rev Biochem* 2015, **84**:291-323.
37. Zhang J, Manley JL: **Misregulation of pre-mRNA alternative splicing in cancer**. *Cancer Discov* 2013, **3**(11):1228-1237.
38. Zhu S, Chen Z, Katsha A, Hong J, Belkhiri A, El-Rifai W: **Regulation of CD44E by DARPP-32-dependent activation of SRp20 splicing factor in gastric tumorigenesis**. *Oncogene* 2016, **35**(14):1847-1856.
39. Butkyte S, Ciupas L, Jakubauskiene E, Vilys L, Mocevicius P, Kanopka A, Vilkaitis G: **Splicing-dependent expression of microRNAs of mirtron origin in human digestive and excretory system cancer cells**. *Clin Epigenetics* 2016, **8**:33.
40. Park WC, Kim HR, Kang DB, Ryu JS, Choi KH, Lee GO, Yun KJ, Kim KY, Park R, Yoon KH *et al*: **Comparative expression patterns and diagnostic efficacies of SR splicing factors and HNRNPA1 in gastric and colorectal cancer**. *BMC Cancer* 2016, **16**:358.
41. Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R, Levanon EY: **Identification of recurrent regulated alternative splicing events across human solid tumors**. *Nucleic Acids Res* 2015,

43(10):5130-5144.

42. Chen AJ, Paik JH, Zhang H, Shukla SA, Mortensen R, Hu J, Ying H, Hu B, Hurt J, Farny N *et al*: **STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA**. *Genes Dev* 2012, **26**(13):1459-1472.

43. Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, Fang Z, Fang R, Lu D, Ji H *et al*: **The RNA-binding protein QKI suppresses cancer-associated aberrant splicing**. *PLoS Genet* 2014, **10**(4):e1004289.

Figures

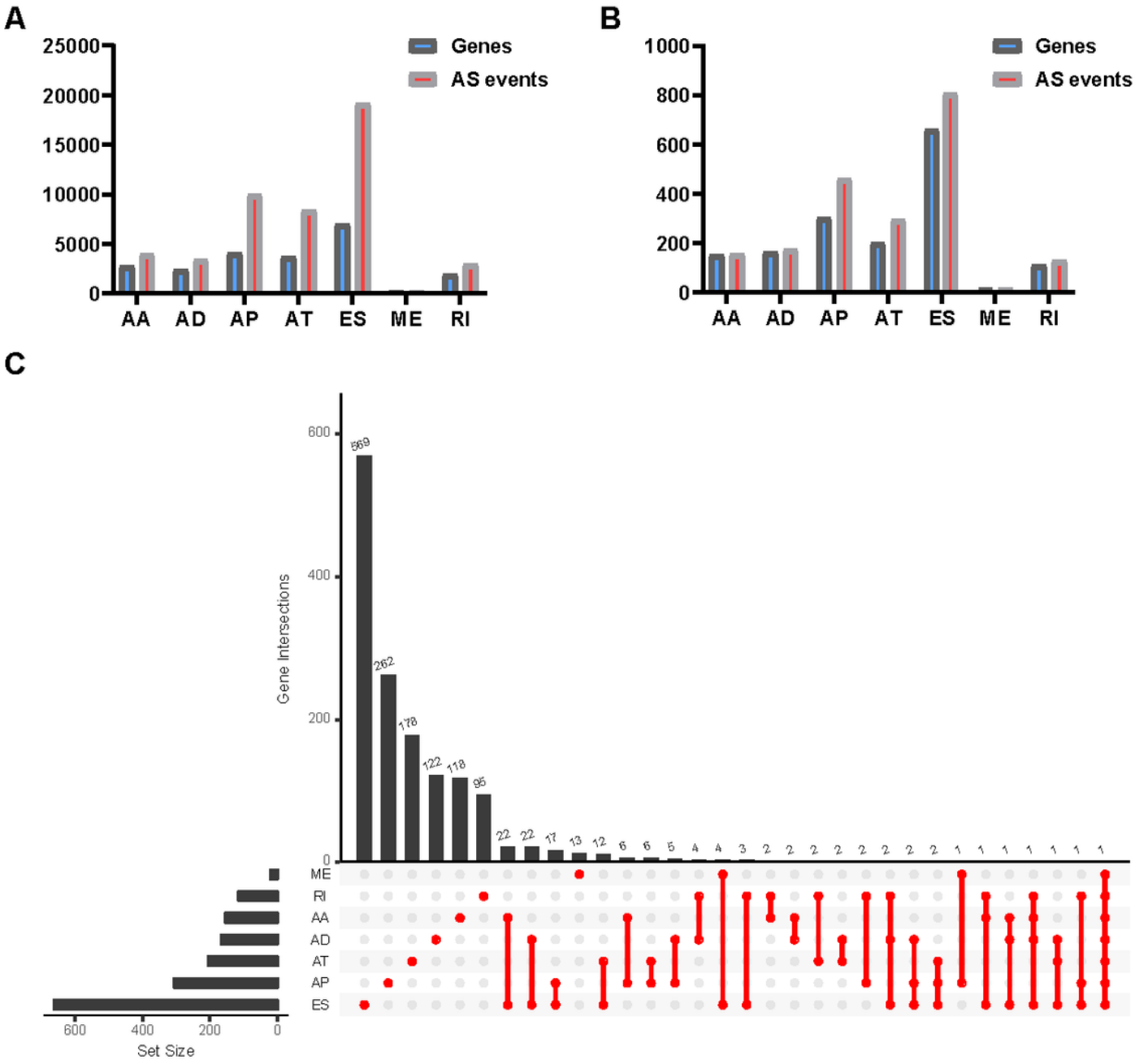

Figure 1

Prognosis-related alternative splicing (AS) events. (A) The number of AS events and corresponding genes included in the present study; (B) The number of prognosis-related AS events and corresponding genes obtained by using univariate COX analysis; (C) UpSet plot of interactions between the seven types of survival associated AS events in STAD. One gene may have up to seven types of alternative splicing to be associated with patient survival.

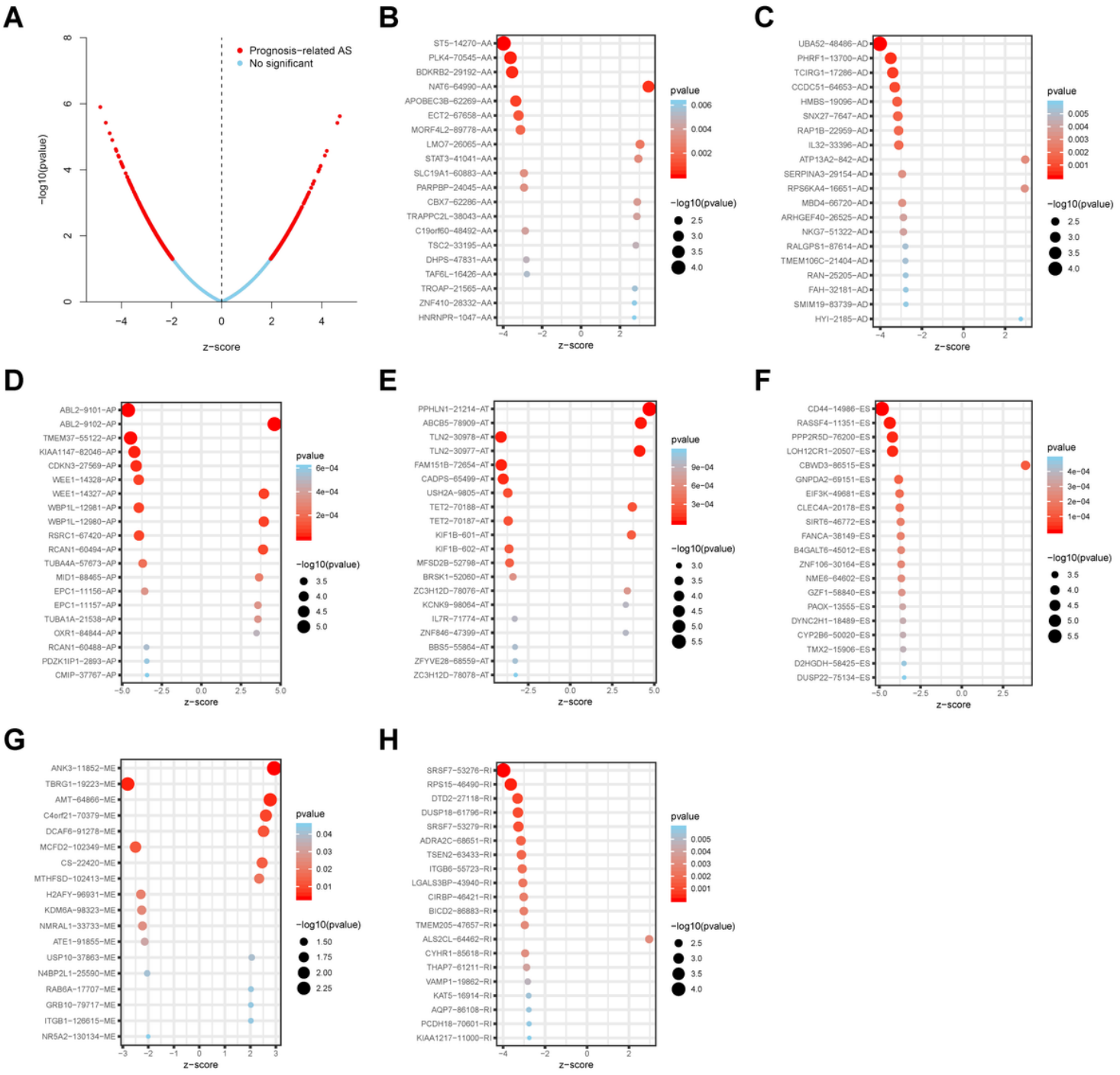

Figure 2

Top 20 most significant alternative splicing (AS) events in STAD. (A) The red dots represent AS events that are significantly correlated with patient survival. The blue dots represent AS events without

correlation. The top 20 AS events correlated with clinical outcome based on acceptor sites (B), alternate donor sites (C), alternate promoters (D), alternate terminators (E), exon skips (F), mutually exclusive exons (G), and retained introns (H).

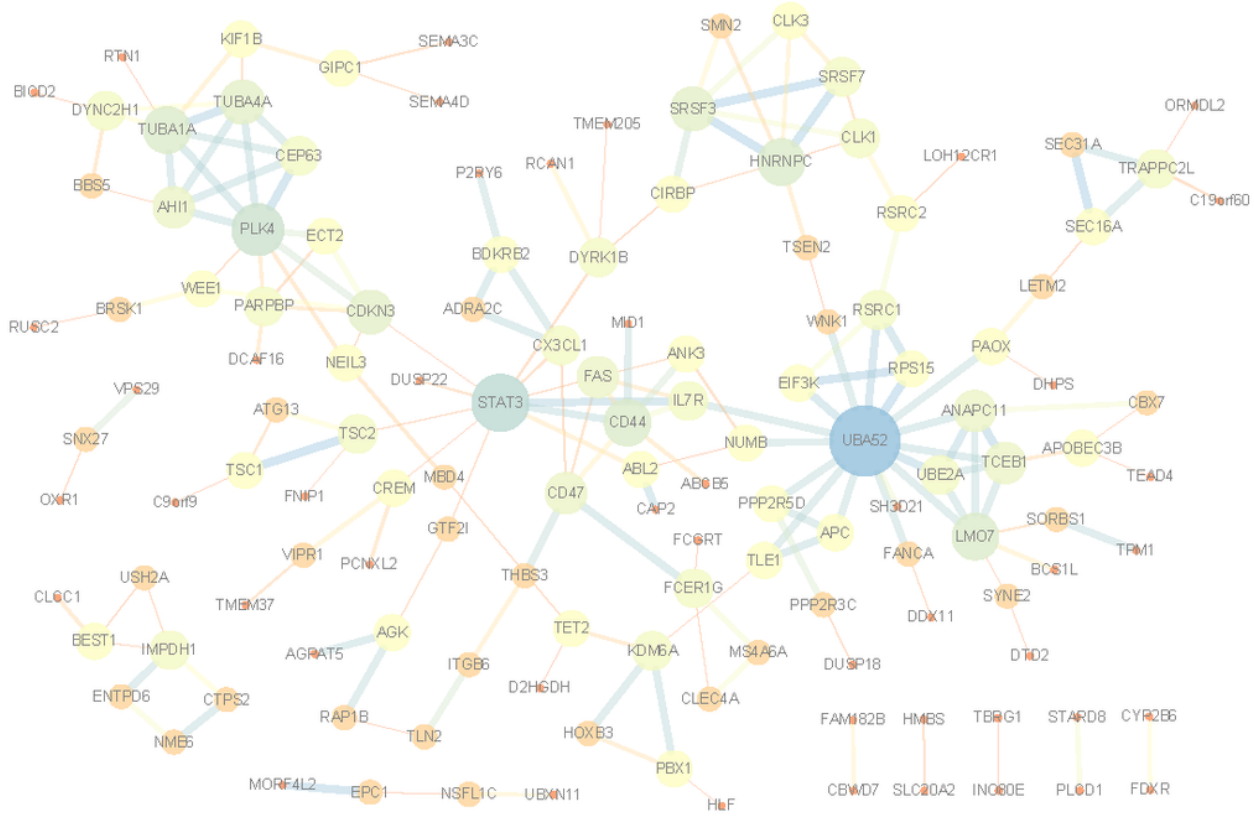

Figure 3

Protein-protein interaction network of genes with survival-associated alternative splicing events in STAD. For nodes, low degrees correspond to small sizes and bright colors; For edges, low combined_scores correspond to small sizes and bright colors.

A

ID	Biological process	false discovery rate	Count
GO:0046907	intracellular transport	0.0024	91
GO:0048519	negative regulation of biological process	0.0024	256
GO:0048523	negative regulation of cellular process	0.0024	234
GO:0051641	cellular localization	0.0024	130
GO:0051726	regulation of cell cycle	0.0024	79
GO:0010629	negative regulation of gene expression	0.003	104
GO:0006996	organelle organization	0.0052	169
GO:0008104	protein localization	0.0052	116
GO:0009987	cellular process	0.0052	621
GO:0010564	regulation of cell cycle process	0.0052	52

B

ID	Cellular component	false discovery rate	Count
GO:0044424	intracellular part	2.67E-13	643
GO:0005622	intracellular	1.52E-12	649
GO:0043229	intracellular organelle	1.70E-11	574
GO:0043227	membrane-bounded organelle	1.18E-10	536
GO:0005737	cytoplasm	1.44E-10	535
GO:0043226	organelle	1.97E-10	577
GO:0044422	organelle part	3.17E-08	443
GO:0044446	intracellular organelle part	3.17E-08	434
GO:0044444	cytoplasmic part	1.09E-07	450
GO:0043231	intracellular membrane-bounded organelle	1.41E-07	487

C

ID	Molecular function	false discovery rate	Count
GO:0005488	binding	5.91E-06	541
GO:0019899	enzyme binding	0.00015	134
GO:0005515	protein binding	0.00016	325
GO:0051020	GTPase binding	0.00016	52
GO:0005085	guanyl-nucleotide exchange factor activity	0.00021	33
GO:0017137	Rab GTPase binding	0.0034	21
GO:0017016	Ras GTPase binding	0.004	41
GO:0031267	small GTPase binding	0.004	42
GO:0005088	Ras guanyl-nucleotide exchange factor activity	0.0099	24
GO:0044877	protein-containing complex binding	0.0117	63

Figure 4

Gene ontology analysis of genes with survival-associated alternative splicing events. (A) Biological process; (B) Cellular component; (C) Molecular function.

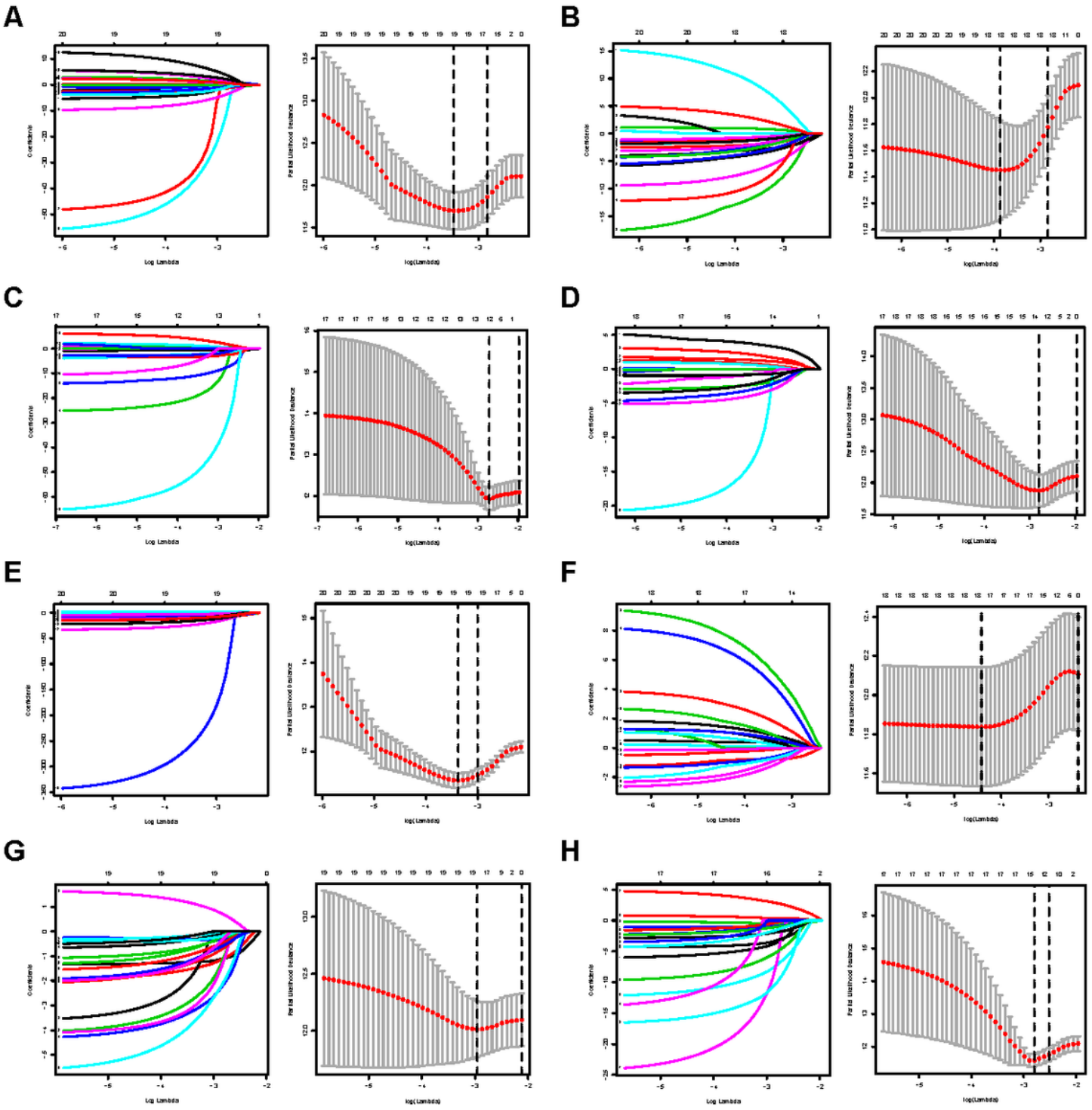

Figure 5

Construction of prognostic signatures based on LASSO COX analysis. (A) Alternate Acceptor site (AA); (B) Alternate Donor site (AD); (C) Alternate Promoter (AP); (D) Alternate Terminator (AT); (E) Exon Skip (ES); (F) Mutually Exclusive Exons (ME); (G) Retained Intron (RI); and (H) All types of AS.

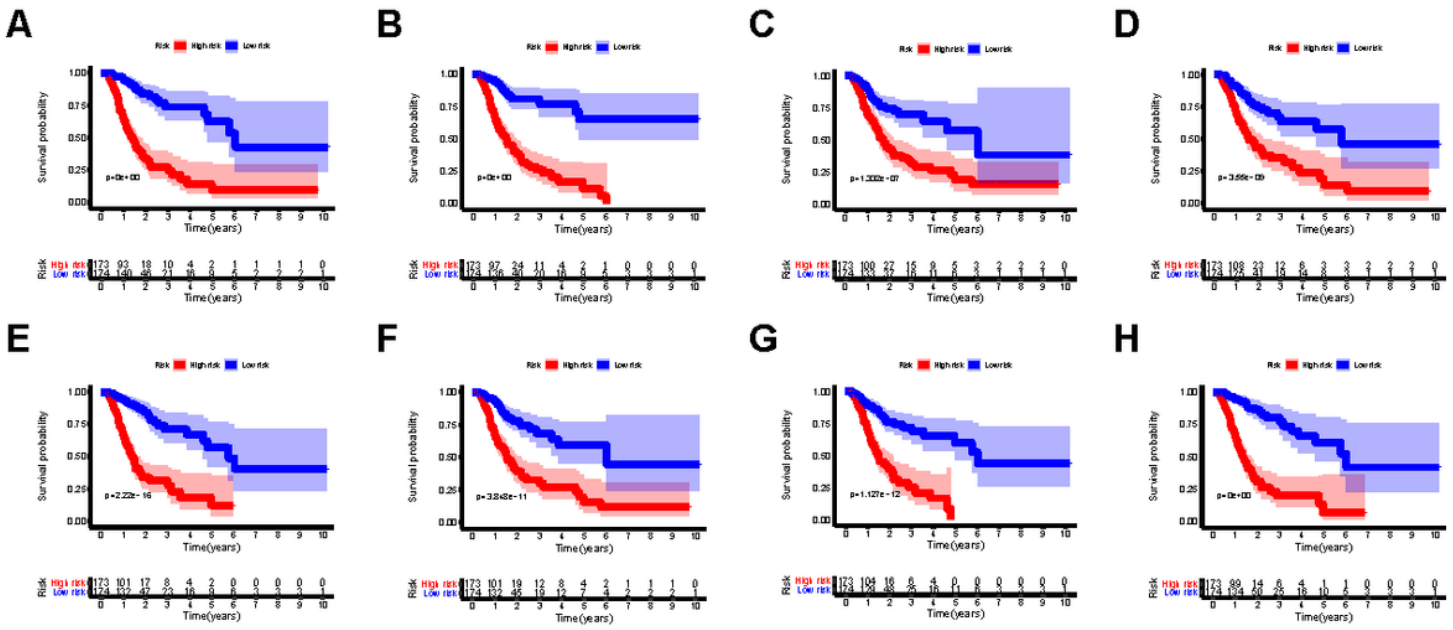

Figure 6

Kaplan-Meier curves of prognostic predictors for STAD. (A) Alternate Acceptor site (AA); (B) Alternate Donor site (AD); (C) Alternate Promoter (AP); (D) Alternate Terminator (AT); (E) Exon Skip (ES); (F) Mutually Exclusive Exons (ME); (G) Retained Intron (RI); and (H) All types of AS.

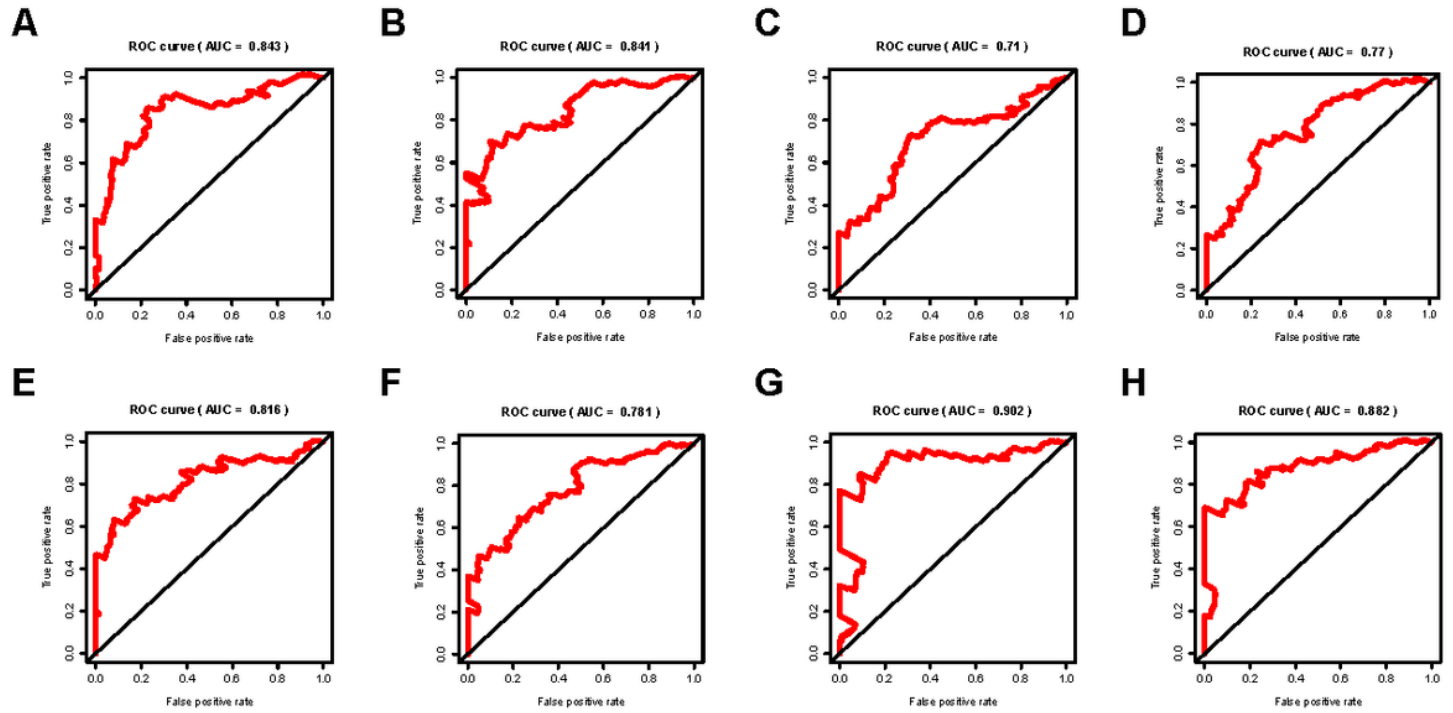

Figure 7

ROC curves of prognostic predictors for STAD. (A) Alternate Acceptor site (AA); (B) Alternate Donor site (AD); (C) Alternate Promoter (AP); (D) Alternate Terminator (AT); (E) Exon Skip (ES); (F) Mutually Exclusive Exons (ME); (G) Retained Intron (RI); and (H) All types of AS.

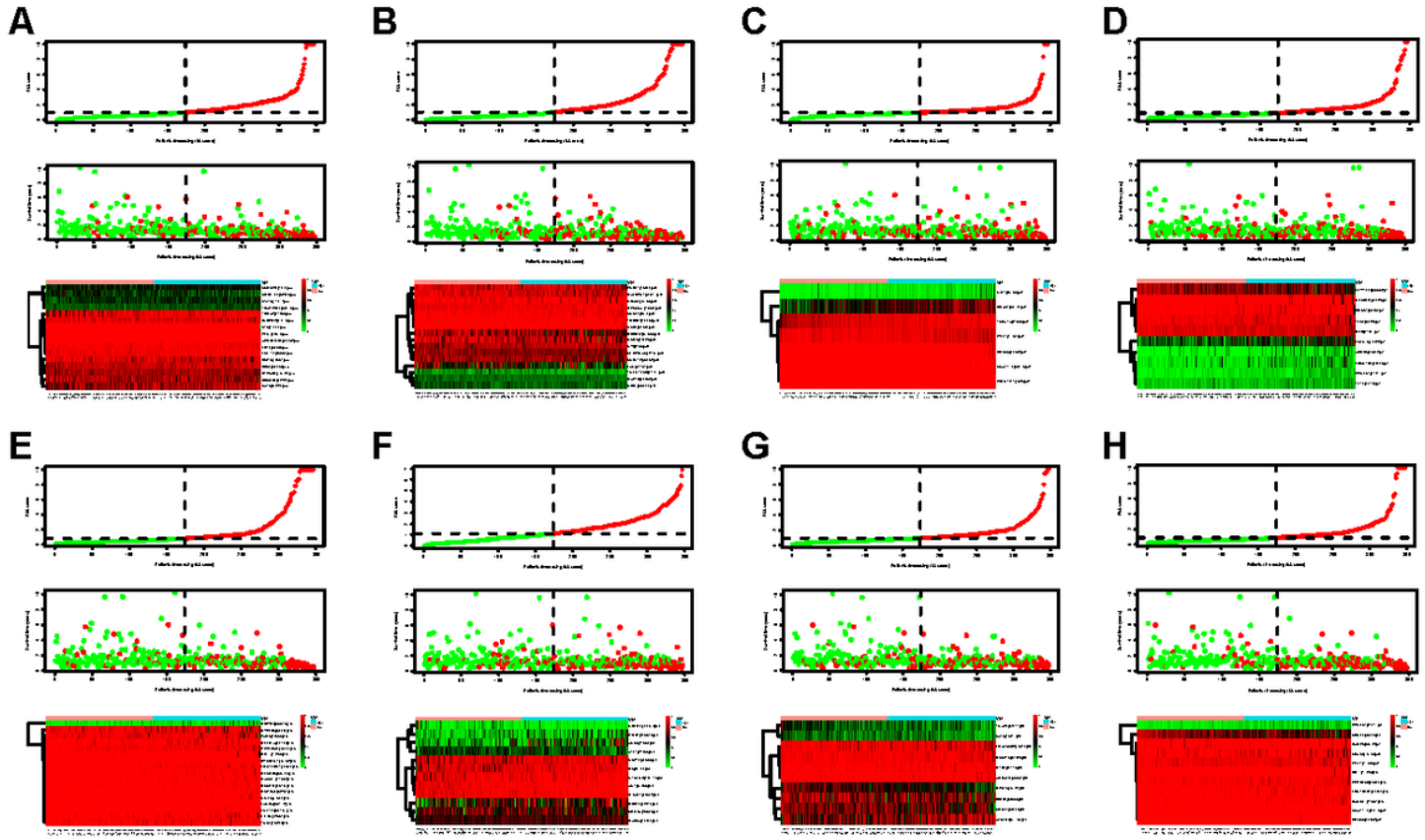

Figure 8

Determination and analysis of the prognostic AS signatures in STAD cohort. STAD patients were divided into high- and low-risk subgroups based on the median cut of risk score calculated separately. The upper part of each assembly represents the risk score curve, the middle part indicates distribution of patients' survival status and survival times ranked by risk score, and the bottom heatmap displays splicing pattern of the AS signature from each AS type or all seven AS types. Color transition from green to red indicates the increasing PSI score of corresponding AS event from low to high. (A) Alternate Acceptor site (AA); (B) Alternate Donor site (AD); (C) Alternate Promoter (AP); (D) Alternate Terminator (AT); (E) Exon Skip (ES); (F) Mutually Exclusive Exons (ME); (G) Retained Intron (RI); and (H) All types of AS.

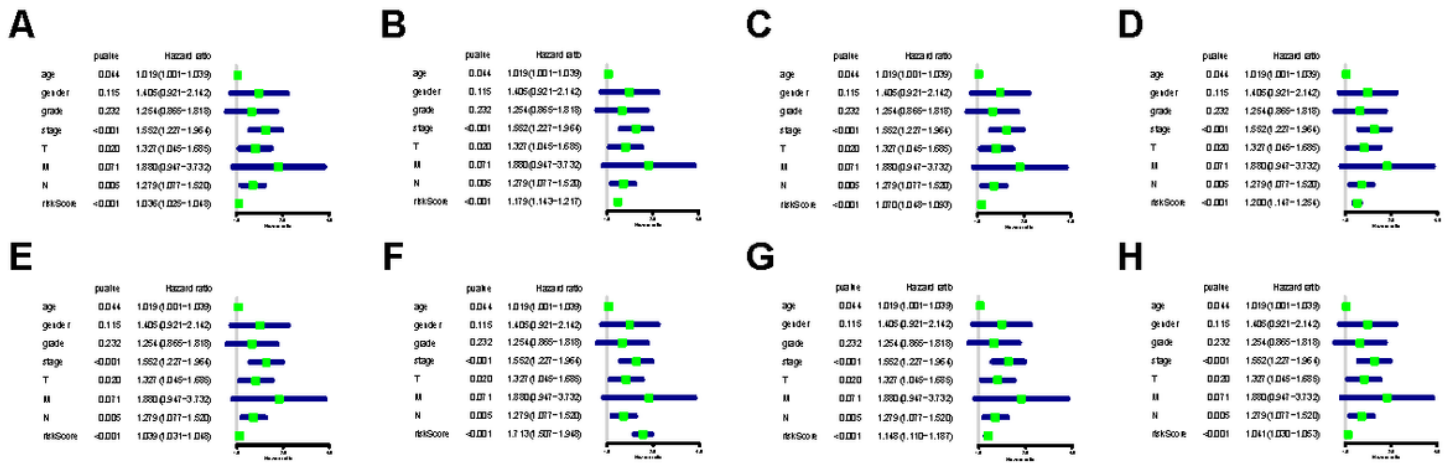

Figure 9

Univariate Cox regression analysis of clinical parameters and riskScore in STAD. (A) Alternate Acceptor site (AA); (B) Alternate Donor site (AD); (C) Alternate Promoter (AP); (D) Alternate Terminator (AT); (E) Exon Skip (ES); (F) Mutually Exclusive Exons (ME); (G) Retained Intron (RI); and (H) All types of AS.

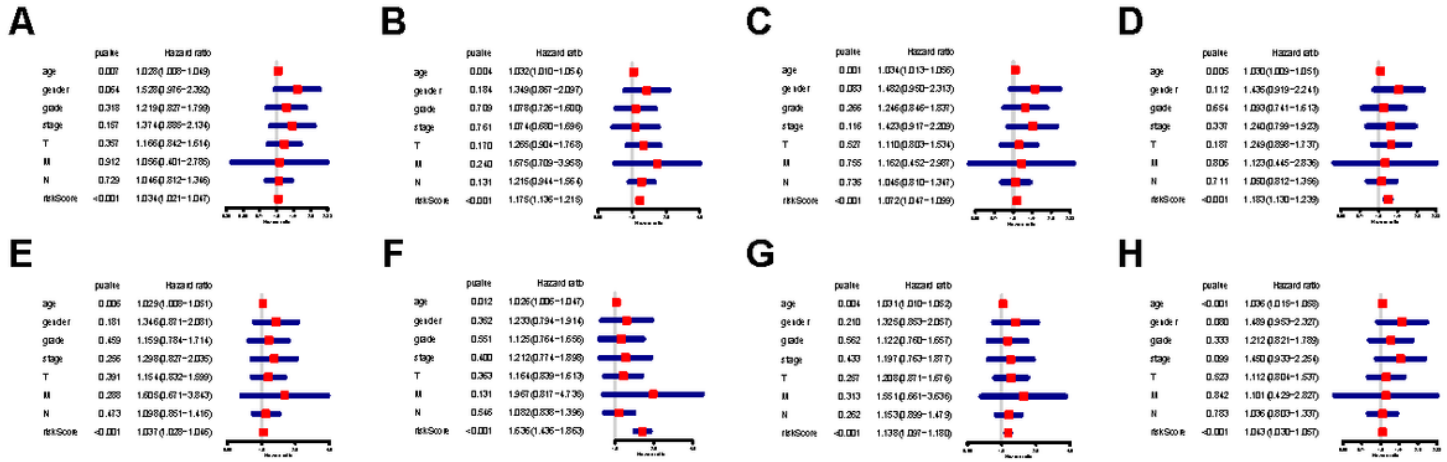

Figure 10

Multivariate Cox regression analysis of clinical parameters and riskScore in STAD. (A) Alternate Acceptor site (AA); (B) Alternate Donor site (AD); (C) Alternate Promoter (AP); (D) Alternate Terminator (AT); (E) Exon Skip (ES); (F) Mutually Exclusive Exons (ME); (G) Retained Intron (RI); and (H) All types of AS.

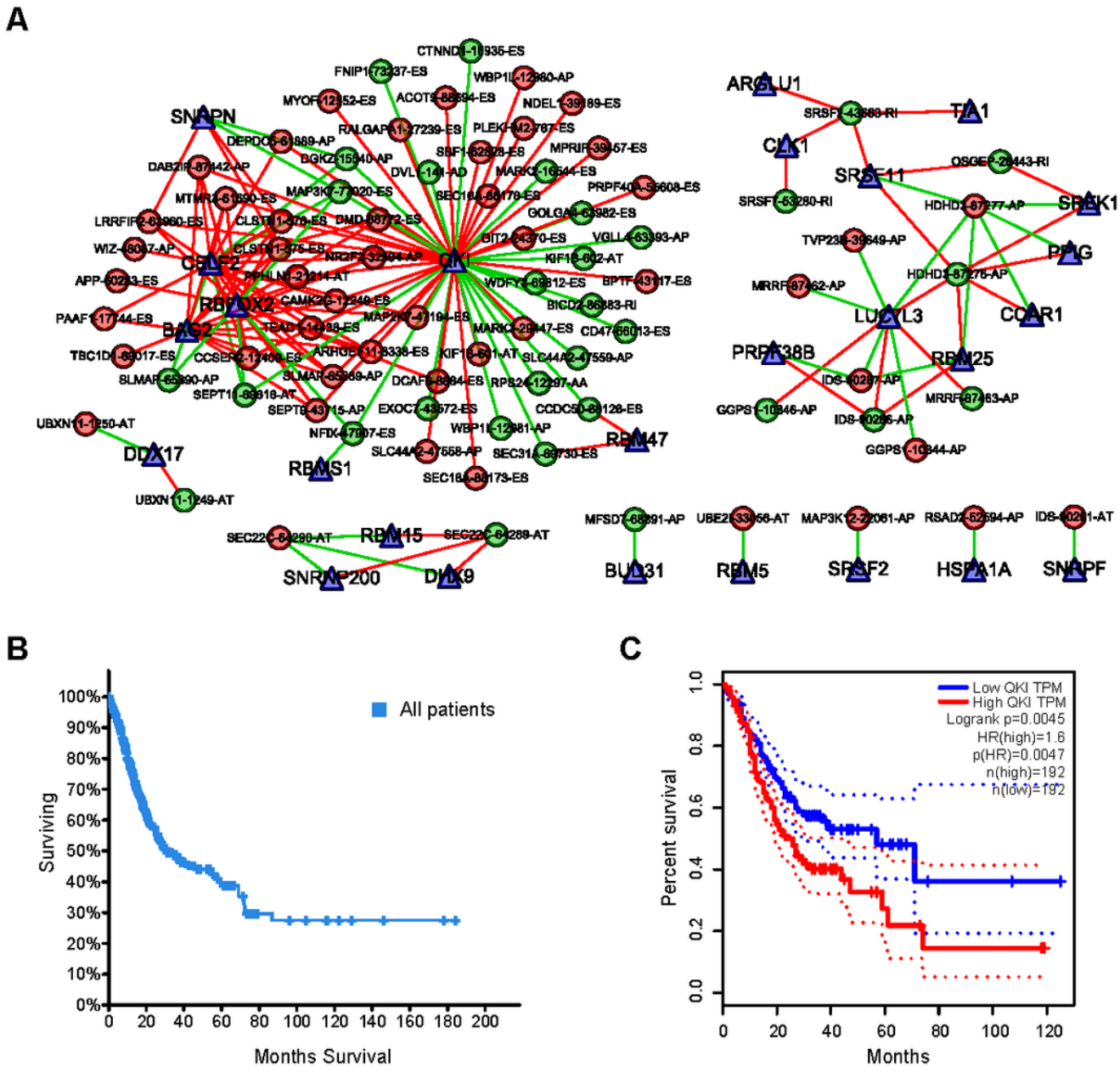

Figure 11

SF-AS network and survival analysis. (A) Survival-associated SF-AS network in STAD. AS is represented by circles (red for high-risk AS, green for low-risk AS), and SF is represented by triangles. A line between AS and SF indicates a regulatory relationship between them (the red line represents positive regulation and the green line represents negative regulation); (B) Overall survival of STAD patients; (C) Correlation analysis between the expression of QKI and the survival rate of STAD patients.