

Original Article

Prognostic alternative splicing signatures and underlying regulatory network in esophageal carcinoma

Zu-Cheng Xie¹, Hua-Yu Wu⁴, Fu-Chao Ma¹, Yi-Wu Dang², Zhi-Gang Peng¹, Hua-Fu Zhou³, Gang Chen²

Departments of ¹Medical Oncology, ²Pathology, ³Cardio-Thoracic Surgery, First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning 530021, Guangxi Zhuang Autonomous Region, P. R. China;

⁴Department of Cell Biology and Genetics, School of Pre-Clinical Medicine, Guangxi Medical University, 22 Shuangyong Road, Nanning 530021, Guangxi Zhuang Autonomous Region, P. R. China

Received February 24, 2019; Accepted June 9, 2019; Epub July 15, 2019; Published July 30, 2019

Abstract: Alternative splicing (AS) has been widely reported to play an important role in cancers, including esophageal carcinoma (ESCA). However, no study has comprehensively investigated the clinical use of combination of prognostic AS events and clinicopathological parameters. Therefore, we collected 165 ESCA patients including 83 esophageal adenocarcinoma (EAC) and 82 esophageal squamous cell carcinoma (ESCC) patients from The Cancer Genome Atlas to explore the survival rate associated with seven types of AS events. Prognostic predictors for the clinical outcomes of ESCA patients were built. Predictive prognosis models of the alternative acceptor site in ESCA (area under the curve [AUC] = 0.83), alternative donor site in EAC (AUC = 0.99), and alternative terminator site in ESCC (AUC = 0.974) showed the best predictive efficacy. A novel combined prognostic model of AS events and clinicopathological parameters in ESCA was also constructed. Combined prognostic models of ESCA all showed better predictive efficacy than independent AS models or clinicopathological parameters model. Through constructing splicing regulatory network, the expression of AS factor was found to be negatively correlated with the most favorable AS events. Moreover, gene amplification, mutation, and copy number variation of AS genes were commonly observed, which may indicate the molecular mechanism of how the AS events influence survival. Conclusively, the constructed prognostic models based on AS events, especially the combined prognostic models of AS signatures and clinicopathological parameters could be used to predict the outcome of ESCA patients. Moreover, the splicing regulatory network and genomic alteration in ESCA could be used for illuminating the potential molecular mechanism.

Keywords: Alternative splicing, esophageal carcinoma, prognosis, splicing factor

Introduction

Esophageal carcinoma (ESCA), consisting of esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC) subtypes, is a prevalent malignancy worldwide [1]. In 2017, an estimated total of 16,940 new cases and an estimated 15,690 deaths from ESCA were reported [2]. ESCA patients have a dismal prognosis with 5-year survival rates from 40% to 50% [3].

Over the last few decades, there has been great progress in research on new ESCA therapies. Numerous novel predictive prognosis biomarkers have been found. For example, mRNAs [4-6], MicroRNAs [7-10], and long non-coding RNAs [11-14] have been widely reported to be

associated with the survival of ESCA patients. It has become widely accepted that gene regulation dysfunction is a critical factor in the initiation and progression of tumors. Aberrant expression of genes has been shown to be remarkably significant in ESCA diagnosis and therapy [15, 16].

Alternative splicing (AS), a sophisticated post-transcriptional process, expands gene expression patterns, resulting in increased protein diversity. AS events are estimated to occur in up to 94% of human genes [17]. AS is widely involved in biological processes such as cell differentiation, proliferation, and apoptosis [18-20], has also been found to play an important role in oncogenesis, and may provide opportunities for novel cancer therapeutics [21, 22].

Recently, the crucial roles of AS in the diagnosis and prognosis of various cancers, including gastrointestinal adenocarcinomas [23-26] and urogenital malignancies [27, 28], have been gradually uncovered. These discoveries have provided inspiring perspectives in the field of splicing research.

With the rapid development of next-generation sequencing, it has become easier to obtain the RNA-seq data in clinical samples. Therefore, research on AS and the clinical outcomes of patients can be accomplished. Some studies have reported an association between AS and the prognosis of patients in several cancers, such as breast cancer [29], lung cancer [30], and ovarian cancer [31]. In ESCA, Mao's study has investigated the prognostic AS events in EAC and ESCC separately [32]. However, it still lacks sufficient work in the way of clinicopathological parameters and underlying molecular mechanism. Therefore, it was necessary to carry out this study to provide more information about the combination of splicing and clinical parameters, as well as potential mechanism of the survival-associated splicing events in ESCA.

The Cancer Genome Atlas (TCGA) database provides valuable AS data for 33 different cancers as well as for normal tissues. In this study (**Figure 1**), we comprehensively analyzed the prognosis-related AS events in 165 ESCA patients including 83 EAC and 82 ESCC patients. We constructed prognostic models of AS events to evaluate the prognostic value of the AS events. More importantly, we constructed novel prognostic models for ESCA via combining splicing signatures and clinicopathological parameters. We revealed the possible regulatory mechanisms in ESCA and its histological subtypes by constructing key splicing networks. We also investigated the underlying genetic alterations including gene amplification, mutation, and copy number variation of splicing genes. We hope this study will provide insights into understanding the regulatory mechanisms of AS events in ESCA and will facilitate the therapeutic clinical practice of ESCA patients.

Materials and methods

Collection of AS profiles

TCGASpliceSeq (<http://bioinformatics.mdanderson.org/TCGASpliceSeq>), a web-based resource,

provides a user-friendly interface for detailed views of alternative mRNA splicing based on the TCGA database [33]. Using TCGASpliceSeq, we downloaded the alternative mRNA splicing data of ESCA. Seven alternative events, namely, the alternative acceptor site (AA), alternative donor site (AD), alternative promoter site (AP), alternative terminator site (AT), exon skip (ES), mutually exclusive exons (ME), and retained intron (RI), were quantified with the percent spliced in (PSI) value, which ranges from zero to one hundred [34].

Identification of survival-associated splicing events and clinical parameters

Related clinical data of ESCA patients were also downloaded from the TCGA database. Only patients with an overall survival (OS) of 90 days or longer were enrolled in this study. A total of 165 ESCA patients consisting of 83 EAC and 82 ESCC patients were included. To identify the association between AS events and survival, we divided the patients into two groups according to a median cut-off value. Univariate Cox regression was used to analyze the relationship between AS events and OS. Multivariate Cox regression was used to identify independent prognostic predictors of AS events. For ESCA as a whole, the top 15 in each type of splicing and seven combined events were selected. For each subtype of ESCA, including EAC and ESCC, the top 10 splicing events were likewise selected using a multivariate Cox proportional hazards regression. The prognostic value of demographic and clinicopathological parameters of ESCA was also investigated using univariate and multivariate Cox regression analysis. Univariate Cox regression was conducted using a survival package in R while multivariate Cox regression was done using SPSS Statistics ver. 23.0 (IBM, Armonk, NY). A *p*-value less than 0.05 was adopted for identifying significant splicing events.

Construction and comparison of prognostic models based on survival-associated AS events and clinical parameters

Prognostic models for ESCA, EAC, and ESCC patients were built using prognostic signatures, namely, the independent survival-associated AS events in the multivariate Cox regression. Each prognostic signature was evaluated with the risk score [28], which was calculated using the following formula:

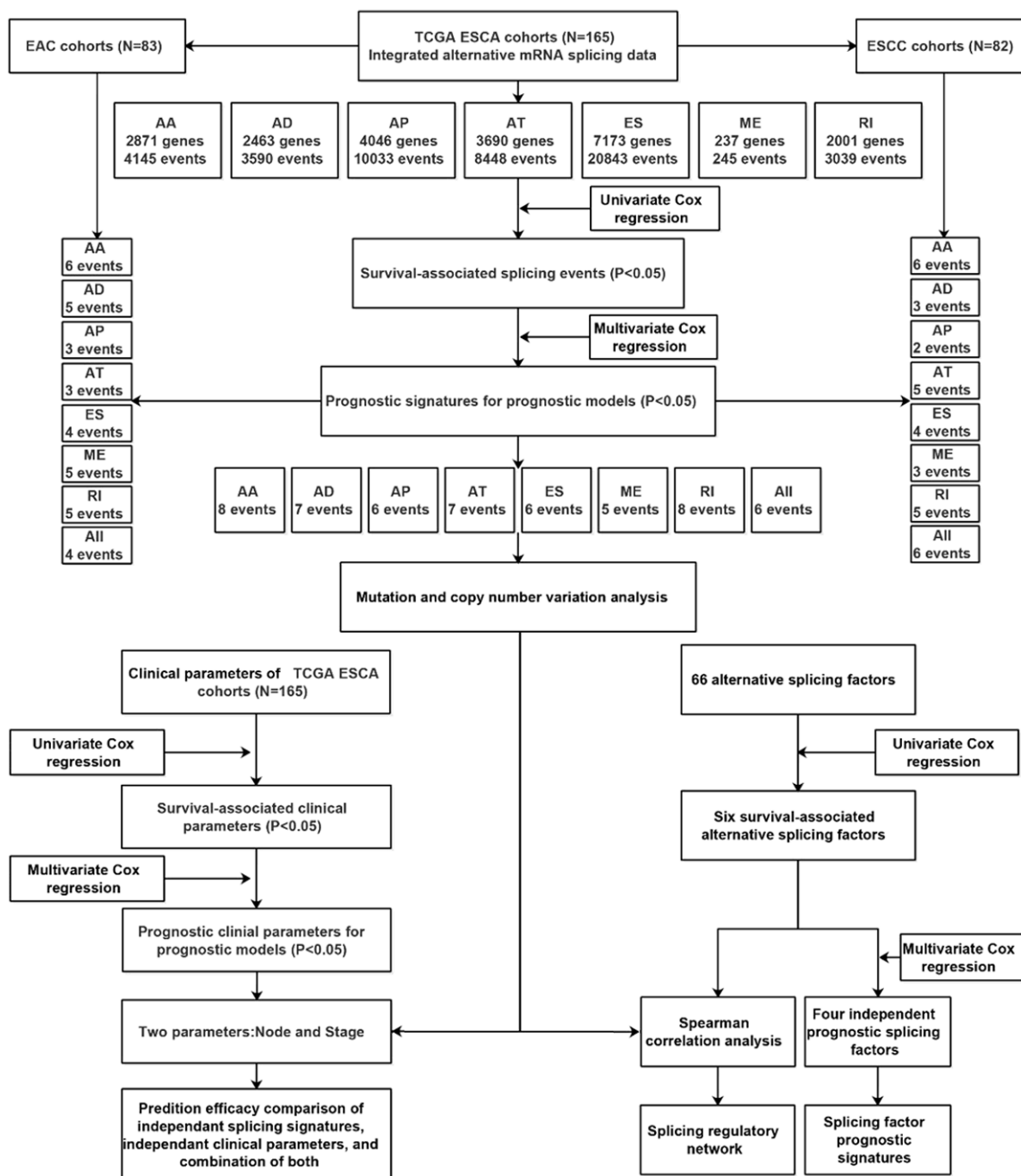


Figure 1. Flowchart of the article's design.

$$\text{Risk score} = \sum_{i=1}^n \text{PSI}_i * \beta_i$$

Likewise, the prognostic model of survival-associated clinical parameters for ESCA was also constructed. In order to investigate the influence of clinical parameters in the prognostic models of AS in ESCA, we also constructed combined prognostic models of splicing signatures and clinical parameters. The combined

risk score was calculated using the following formula [35]:

$$\text{Combined risk score} = \sum_{i=1}^i X_i * \beta_i + \sum_{j=1}^j Z_j * \gamma_j$$

Where, X_i and Z_j represents the splicing events and clinical parameters in ESCA, respectively. And β_i and γ_j represents the estimated coefficients of splicing events and clinical parameters.

ters, respectively. We divided the patients into two groups based on median cut-off value of the risk score when performing the Kaplan-Meier survival prediction. To compare the predictive efficacy of each AS event with respect to the survival of ESCA patients, the survival receiver operator characteristic (ROC) package in R was used to construct the ROC curve. Then, the area under the curve (AUC) was calculated to evaluate the predictive ability of each model.

Construction of gene network and correlation analysis

Upset plot, an innovative method to visualize interactive sets [36], was used to display the intersections among the seven types of prognosis-associated AS events. The gene symbols of the most significant AS events ($P < 0.005$) in ESCA and its histological subtypes were used to construct the interactive gene network. The Reactome FI plugin of Cytoscape 3.60 [37] was used to identify the key hub genes among the survival-associated AS genes.

The information of alternative splicing factors was obtained from SpliceAid2 (<http://www.introni.it/splicing.html>) [38]. The level 3 data of splicing factors were downloaded from the TCGA database. The expression data were normalized using the fragments per kilobase of transcript per million mapped reads method. Univariate and multivariate Cox regressions were then conducted to identify survival-associated splicing factors. Student's *t*-tests were used to compare the expression of survival-associated splicing factors between ESCA and adjacent normal tissues. In order to understand the correlation between the splicing factor genes and AS events, a correlation analysis between the expression of the survival-associated splicing factor genes and PSI values of survival-associated AS events was performed. Cytoscape ver. 3.60 was used to visualize the correlation network.

Gene amplification, mutation, and copy number variation analysis

The cBioPortal for Cancer Genomics (<http://cbioportal.org>), an intuitive Web interface, provides visualization, analysis, and downloading of multidimensional cancer genomics datasets [39]. The possible gene amplification, mutation, and copy number variation of the prognostic signatures in ESCA may help illuminate how

splicing events can influence survival. As a result, we utilized 184 ESCA samples with sequencing and copy number alternation data to investigate the possible alternation of the eight sets of gene symbols of the prognostic signatures in ESCA. Segment mean of zero indicated neutral or normal copy number. A heat map was also used to visualize the copy number alteration.

Results

Integrated AS events in the ESCA cohort

A total of 165 ESCA patients including 83 EAC and 82 ESCC patients were enrolled in order to analyze the AS events from the TCGA database. In total, 50,342 AS events containing 10,765 gene symbols, indicating that different AS events may occur in a single gene, were collected from the TCGA database. Specifically, we obtained 4,146 AAs from 2,872 genes, 3,591 ADs from 2,464 genes, 10,034 APs from 4,047 genes, 8,449 ATs from 3,691 genes, 20,844 ESs from 7,174 genes, 246 MEs from 238 genes, and 3,039 RIs from 2,002 genes. ES was the most frequent AS event, while ME was the least frequent AS event.

Survival-associated AS events in cohorts for ESCA and its histological subtypes

As shown in **Figure 2**, we identified 3,276, 4,850 and 2,199 survival-associated AS events from 2,157, 2,943 and 1,501 genes in the ESCA, ECA, and ESCC cohorts through univariate Cox regression, respectively ($P < 0.05$). As we observed that one gene may have more than one AS event, we presented the intersecting sets of the seven types of AS events using Upset plots. The Upset plots of ESCA, ECA, and ESCC are displayed in **Figure 3**. ES was responsible for the majority of the events, followed by AP. In ESCA and EAC, a single gene such as CIRBP, could have up to five types of AS events, while in ESCC a single gene could only have up to three types of AS events. In order to identify the most important genes among the survival-associated AS events, we selected the gene symbols from the most significant events ($P < 0.005$) to construct the gene interactive network for ESCA and its histological subtypes. As shown in **Figures 4** and **S1**, the gene network revealed important hub genes such as GRB2, UBC, and KIF2A in ESCA, EAC, and ESCC based on network topology [40].

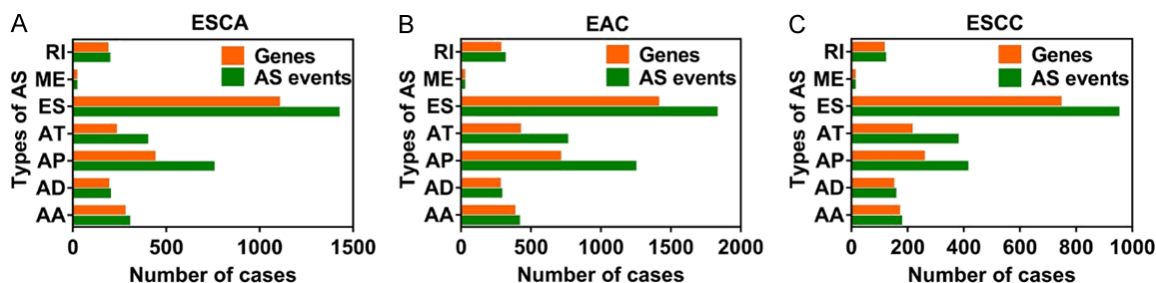


Figure 2. Bar chart for survival-associated AS events and gene symbols. The horizontal axis represents the number of AS events, while the vertical axis represents the type AS events. A. The number of AS events and gene symbols for ESCA. B. The number of AS events and gene symbols for EAC. C. AS events and gene symbols for ESCC.

Prognostic predictor for ESCA, EAC, and ESCC patients

To predict the prognostic markers for clinical ESCA patients, we selected the most significant AS events for each type of event and the seven combined events using multivariate Cox regression. For the ESCA cohort, the top 15 significant AS events were selected and the significant AS events ($P < 0.05$) from the multivariate Cox regression were used to build a prognostic model (Table 1). Similarly, in order to investigate the different roles of AS events in the histological subtypes of ESCA, we selected the top 10 significant AS events in the ECA and ESCC cohorts using multivariate Cox regressions, respectively. A prognostic model was built for each type of splicing as well as a combined one using significantly independent survival-associated AS events (Tables 2 and 3). The cohorts were then divided into low- and high-risk groups using the median risk score value as a cut-off. Detailed prognostic signature information of ESCA groups are visualized in Figure 5. Regarding EAC and ESCC groups, detailed prognostic signature information was presented in Figures S2 and S3. As displayed in Figures 6 and S4, the role of AS events in predicting survival showed consistency in ESCA, EAC, and ESCC groups. The low-risk group for each type of event as well as the combined events predicted a better outcome for the patients. In order to compare the predictive efficacy of each prognostic model, we performed a ROC curve analysis, and then calculated the AUC. As displayed in Figures 7 and S5, the AUCs for some types of events in EAC, ESCC and ESCA cohorts showed promising predictive efficacy. In ESCA patients, the prognostic model of AA showed the highest predictive efficacy (AUC = 0.83) followed by ES (AUC = 0.803). In the EAC and

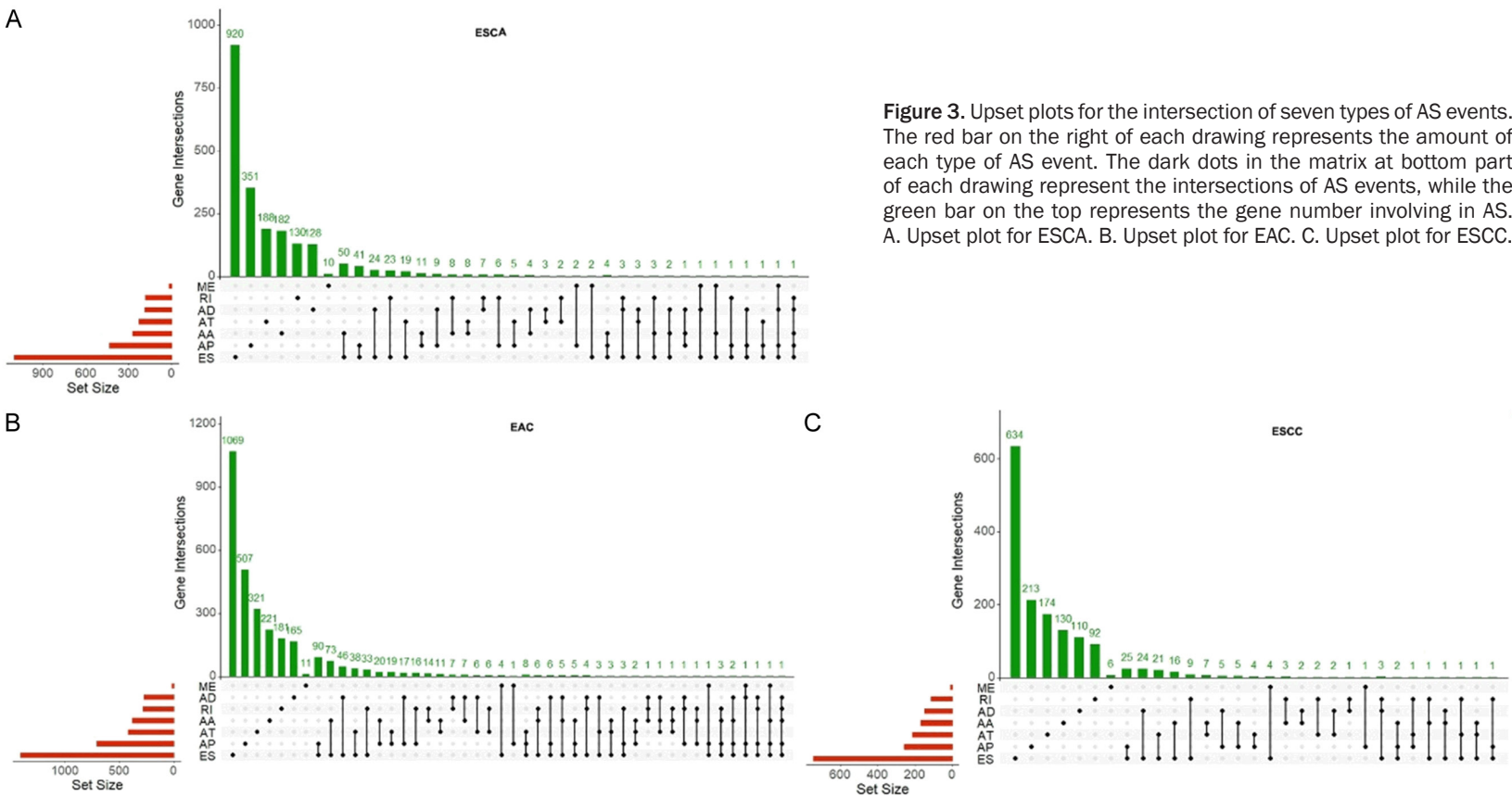
ESCC groups, the prognostic signatures of AD (AUC = 0.99) and AT (AUC = 0.974) dominated the predictive ability with both followed by AA (AUC = 0.871, 0.874), respectively. It is worth noting that the prognostic model of seven combined splicing did not show the highest predictive efficacy, as it had AUCs of only 0.799, 0.832, and 0.81 in ESCA, EAC, and ESCC, respectively. Compared to ESCC, the EAC predictors showed better predictive efficacy except for the AA and AT predictive models.

Combined prognostic models in ESCA

The prognostic value of related clinical parameters in ESCA was analyzed using univariate and multivariate Cox regression analysis. As shown in Table 4, pathological N stage, pathological M stage, pathological stage, pathological grade, and residual tumor were significant prognostic factors in univariate Cox regression, while pathological N stage and pathological stage were independent prognostic parameters in multivariate Cox regression. Prognostic model of pathological N stage and pathological stage showed that high-risk score was associated with poor survival of ESCA patients. It showed a moderate predictive efficacy at an AUC of 0.714 (Figure S6). In Figure 7, The combined prognostic model of pathological N stage, pathological stage, and each type of splicing show better predictive efficacy than independent splicing model or clinical parameter model. ESCA patients with high-risk score in combined models had a worse survival (Figure S7).

Correlation among AS factors and events

AS factors are RNA-binding proteins which regulate splicing sites. Due to that fact, we intended to identify important AS factors, which may



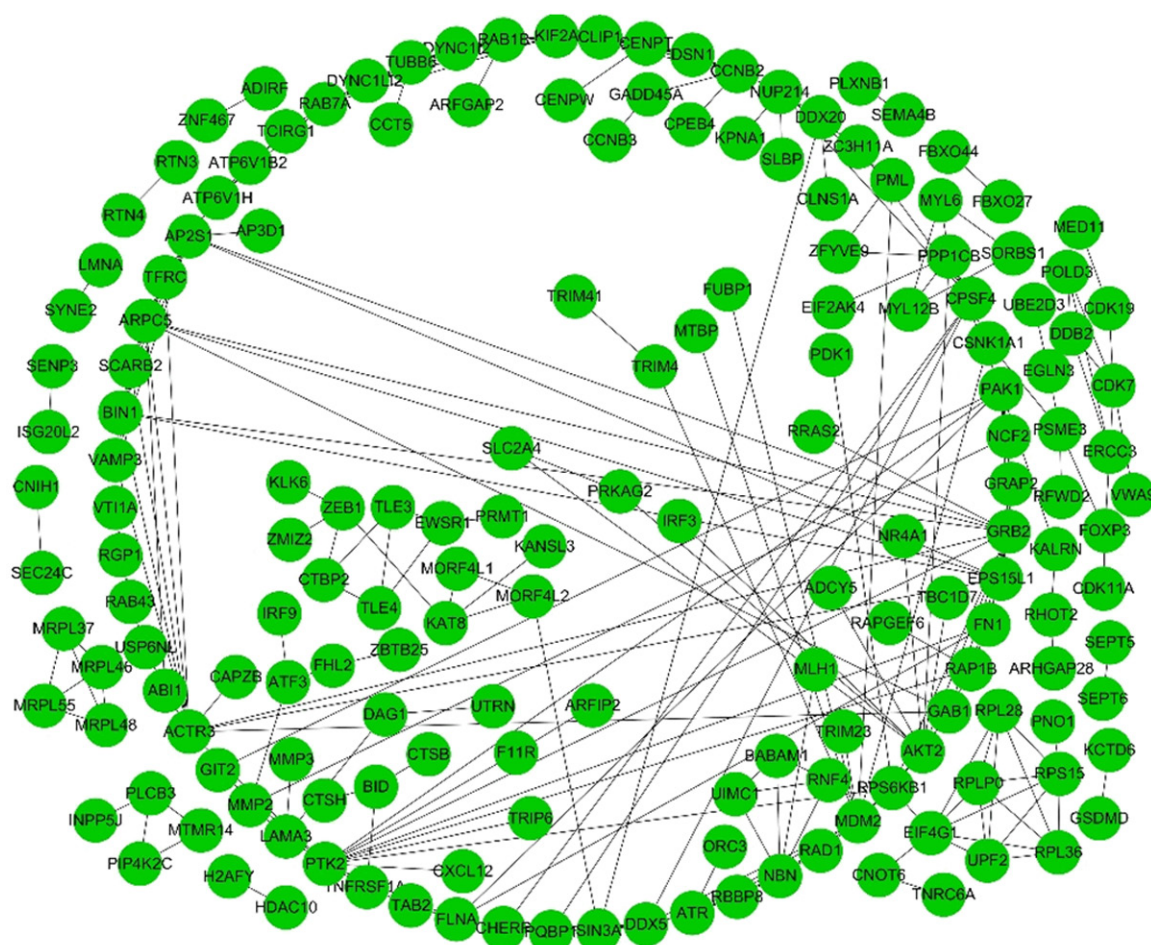


Figure 4. Gene network for gene symbols of the most significant AS events in ESCA. The network was constructed using the Reactome FI plugin of Cytoscape ver. 3.60. Each node represents a gene symbol and the edges among nodes indicate their interactions.

affect the significant survival-associated AS events in ESCA patients. A total of 66 AS factors were identified, and the expression profiles of these genes were downloaded from the TCGA database. First, univariate Cox regression was performed to identify the survival-associated AS factors. Six AS factors-FMR1, ELAVL4, PCBP2, RBMX, RBFOX2, and hnRNPH2-were identified in total. Four AS factors-FMR1, PCBP2, RBMX, and hnRNPH2-were found to be significantly up-regulated in ESCA patients (**Figure 8A-D**). Spearman correlation analysis was then performed to explore the correlation among the six significant AS factors and the significantly independent predictors of AS events. As shown in **Figure 9A**, six AS factors were significantly involved in 25 survival-associated AS events, which included 10 favorable events and 15 unfavorable events in ESCA patients. In

EAC patients, six AS factors were significantly correlated with 9 favorable events and 12 unfavorable events (**Figure 9B**). And in ESCC patients, only five AS factors involving in 9 favorable events and 3 unfavorable events were revealed (**Figure 9C**). Interestingly, we found that the six AS factors were almost always negatively correlated with favorable AS events and positively correlated with unfavorable AS events for most splicing. Unfavorable events and positive correlations in the EAC group tended to be more common than in the ESCC group. Top correlations in ESCA, EAC, and ESCC were displayed as linear correlation plots (**Figure S8**). In order to identify the independent prognostic predictors of AS factors, a further multivariate Cox regression was performed, and four AS factors-ELAVL4, FMR1, PCBP2, and RBFOX2-were identified. A prognostic model of ELAVL4,

Table 1. Details of splicing events used for constructing prognostic models for esophageal carcinoma

Type	AS Events	β	HR	Lower	Upper	P value
AA	ACHE_AA_81030	-0.051	0.951	0.920	0.982	0.002
	ALAS1_AA_65182	-0.097	0.907	0.855	0.963	0.001
	ATP6V1B2_AA_82900	-1.129	0.323	0.149	0.702	0.004
	ATR_AA_67118	-0.886	0.412	0.243	0.701	0.001
	CCDC93_AA_55091	-6.226	0.002	0.000	0.037	0.000
	CPNE1_AA_59187	0.566	1.762	1.055	2.943	0.031
	FAM135A_AA_76637	0.055	1.057	1.025	1.089	0.000
	VEZT_AA_23760	-0.078	0.925	0.880	0.973	0.002
AD	CCT5_AD_71559	-29.561	0.000	0.000	0.000	0.002
	COX6C_AD_84682	0.033	1.033	1.003	1.064	0.030
	EGLN3_AD_27150	-0.091	0.913	0.838	0.995	0.038
	GLT8D1_AD_96400	0.118	1.125	1.044	1.213	0.002
	PPP1CB_AD_53075	-0.337	0.714	0.566	0.901	0.004
	PQBP1_AD_89028	0.025	1.025	1.009	1.041	0.002
	RABAC1_AD_50104	0.050	1.052	1.007	1.099	0.024
	CDK19_AP_77240	-0.252	0.777	0.677	0.893	0.000
AP	CHRD12_AP_17776	0.126	1.134	1.076	1.196	0.000
	IAH1_AP_52629	0.100	1.105	1.053	1.159	0.000
	MAP7_AP_77911	0.396	1.486	1.179	1.874	0.001
	TLE3_AP_31406	0.150	1.161	1.055	1.278	0.002
	TMBIM1_AP_57467	0.241	1.273	1.021	1.588	0.032
	BRSK1_AT_52059	0.084	1.088	1.041	1.137	0.000
AT	CRISPLD2_AT_37866	-0.282	0.754	0.670	0.849	0.000
	ECM2_AT_86878	2.767	15.905	2.906	87.042	0.001
	EYS_AT_76613	0.176	1.193	1.092	1.303	0.000
	L2HGDH_AT_27460	0.095	1.099	1.034	1.168	0.002
	NAE1_AT_36849	28.445	2.257E+12	1.578E+07	3.227E+17	0.000
	TRIM4_AT_80863	0.046	1.047	1.027	1.068	0.000
	C19orf82_ES_47381	0.029	1.029	1.012	1.046	0.001
	CLNS1A_ES_17959	-0.781	0.458	0.285	0.737	0.001
ES	CPSF4_ES_80636	-1.079	0.340	0.157	0.734	0.006
	DYNC1I2_ES_55951	-0.354	0.702	0.526	0.936	0.016
	RAB1B_ES_16997	-2.679	0.069	0.018	0.265	0.000
	TMPRSS4_ES_18957	-0.110	0.896	0.815	0.985	0.023
	ANKRD12_ME_44596	0.067	1.069	1.008	1.134	0.025
	EPB41L2_ME_210763	0.856	2.354	1.040	5.330	0.040
	KLHL2_ME_71038	-0.030	0.970	0.952	0.989	0.002
	P4HA1_ME_12122	-0.047	0.954	0.911	0.999	0.043
ME	PHF21A_ME_157593	0.084	1.087	1.024	1.154	0.006
	C11orf48_RI_16389	-0.901	0.406	0.238	0.692	0.001
	CIRBP_RI_46441	0.094	1.098	1.014	1.189	0.021
	CXCL12_RI_11346	-0.257	0.774	0.610	0.981	0.034
	EP400NL_RI_25245	-0.925	0.397	0.175	0.901	0.027
	FAM136A_RI_53889	0.034	1.035	1.009	1.061	0.007
	GPR85_RI_81465	-0.788	0.455	0.274	0.754	0.002
	PTPN7_RI_9400	-0.233	0.792	0.641	0.979	0.031
RI	TRIM23_RI_72236	-0.613	0.542	0.343	0.855	0.008
	IAH1_AP_52629	0.114	1.121	1.065	1.180	0.000
	C19orf82_ES_47381	0.019	1.019	1.002	1.037	0.033
	ATR_AA_67118	-1.325	0.266	0.088	0.803	0.019
	CLNS1A_ES_17959	-1.010	0.364	0.213	0.624	0.000
	C11orf48_RI_16389	-1.744	0.175	0.088	0.348	0.000
	PNKP_ES_51105	-0.380	0.684	0.502	0.932	0.016

Table 2. Details of splicing events used for constructing prognostic models for esophageal adenocarcinoma

Type	AS Events	β	HR	Lower	Upper	P value
AA	AKT1_AA_29565	-0.995	0.370	0.163	0.839	0.017
	CDV3_AA_66842	0.091	1.095	1.032	1.161	0.003
	CENPW_AA_77442	-3.853	0.021	0.001	0.662	0.028
	OPTN_AA_10778	0.109	1.115	1.029	1.209	0.008
	UBL4A_AA_90615	0.182	1.199	1.102	1.305	0.000
	VEZT_AA_23760	-0.252	0.777	0.698	0.866	0.000
AD	INTS10_AD_82887	0.169	1.184	1.063	1.319	0.002
	MLH1_AD_63935	-0.196	0.822	0.722	0.937	0.003
	POLD3_AD_17775	0.136	1.146	1.036	1.267	0.008
	VDAC3_AD_83726	9.555	1.411E+04	141.453	1.408E+06	0.000
	ZC3H11A_AD_9462	-0.060	0.942	0.897	0.989	0.016
AP	GFOD1_AP_75386	0.088	1.092	1.020	1.168	0.011
	SDCBP2_AP_58485	0.202	1.224	1.108	1.352	0.000
	UGP2_AP_53743	1.653	5.221	2.207	12.355	0.000
AT	ACP6_AT_7384	0.049	1.051	1.017	1.086	0.003
	MAP1LC3B_AT_37942	-0.915	0.400	0.233	0.687	0.001
	TRIM4_AT_80864	-0.042	0.959	0.937	0.981	0.000
ES	CCNB2_ES_30929	-28.169	0.000	0.000	0.000	0.000
	CNIH1_ES_27583	1.408	4.088	1.975	8.461	0.000
	PIP4K2C_ES_22654	-0.759	0.468	0.325	0.674	0.000
	RAB1B_ES_16997	-2.027	0.132	0.028	0.617	0.010
	KLHL2_ME_71038	-0.020	0.980	0.962	0.999	0.040
ME	MAU2_ME_48628	0.310	1.363	1.042	1.783	0.024
	MTHFSD_ME_102413	-0.092	0.912	0.855	0.972	0.005
	P4HA1_ME_12122	-0.044	0.957	0.917	0.998	0.038
	PHF21A_ME_157593	0.037	1.038	1.003	1.075	0.035
	ALDOA_RI_36039	1.643	5.173	1.655	16.169	0.005
RI	ITGAX_RI_36251	-2.576	0.076	0.020	0.287	0.000
	LMNA_RI_8182	8.107	3.318E+03	12.518	8.795E+05	0.004
	PARP10_RI_85521	0.192	1.211	1.058	1.386	0.005
	UBC_RI_25166	-0.263	0.769	0.650	0.910	0.002
	CENPW_ES_77441	-0.496	0.609	0.441	0.841	0.003
All	EAPP_AD_27154	0.271	1.311	1.146	1.498	0.000
	CCNB2_ES_30929	-22.301	0.000	0.000	0.000	0.000
	PIP4K2C_ES_22654	-0.796	0.451	0.315	0.645	0.000

FMR1, PCBP2, and RBF0X2 was constructed, and the low-risk group indicated better survival of ESCA patients (**Figure 8E**).

Genomic alteration of prognostic signatures in ESCA

As shown in **Figure S9**, gene amplification and mutation existed in almost all the gene symbols in the eight prognostic signatures. Amplification

was the most common alteration, while deep deletion could also be seen in some gene symbols. Expression alternation could very clearly be observed in the heat map for most of the gene amplifications and mutations. Several genes, such as ACHE, ATR, CCT5, EYS, TRIM4, and CPSF4, had alteration rates over 10%. **Figure S10** visualizes the copy number variations in the ESCA samples. A segment mean value of zero was regarded as neutral or normal copy number. We observed that most genes showed copy losses or amplifications. These results might help with understanding the potential mechanisms of how AS events were significantly correlated the patients' outcomes.

Discussion

AS is a mechanism for expanding the genome's capacity in the process of pre-mRNA maturation, which provides a basis for generating a diverse array of proteins from a single gene [41]. Accumulating evidence has shown that AS dysfunction is an important factor contributing to human diseases, particularly cancer. The aberrant regulation of AS in various tumors has been the source of a great deal of research recently. For example, AS has been reported to act as a molecular switch in regulating the metabolism of various cancers. This has included the regulation of the metabolic mTOR pathway, the c-Myc-SRSF1-mTOR axis, and meta-

culating evidence has shown that AS dysfunction is an important factor contributing to human diseases, particularly cancer. The aberrant regulation of AS in various tumors has been the source of a great deal of research recently. For example, AS has been reported to act as a molecular switch in regulating the metabolism of various cancers. This has included the regulation of the metabolic mTOR pathway, the c-Myc-SRSF1-mTOR axis, and meta-

Table 3. Details of splicing events used for constructing prognostic models for esophageal squamous cell carcinoma

Type	AS Events	β	HR	Lower	Upper	P value
AA	CHORDC1_AA_18270	-0.828	0.437	0.268	0.711	0.001
	MED29_AA_49816	-2.333	0.097	0.014	0.656	0.017
	NAT6_AA_64990	-0.087	0.917	0.879	0.956	0.000
	SEMA4D_AA_86808	-5.289	0.005	0.000	0.972	0.049
	STX16_AA_59967	-2.375	0.093	0.025	0.345	0.000
AD	ZNF707_AA_85482	-0.051	0.951	0.915	0.988	0.011
	LYSMD2_AD_30622	-1.381	0.251	0.096	0.660	0.005
	TIMM10_AD_15857	-0.748	0.473	0.283	0.792	0.004
AP	YBEY_AD_60921	0.120	1.128	1.056	1.204	0.000
	CHRD12_AP_17777	-0.243	0.784	0.658	0.934	0.006
AT	IGF2BP2_AP_68029	-0.774	0.461	0.277	0.767	0.003
	AGAP3_AT_82351	0.083	1.086	1.031	1.145	0.002
	BRSK1_AT_52059	0.121	1.129	1.060	1.202	0.000
	MRPL37_AT_3138	-0.159	0.853	0.747	0.974	0.019
ES	RTFDC1_AT_59879	-1.258	0.284	0.081	0.996	0.049
	ZNF814_AT_52354	0.080	1.083	1.043	1.125	0.000
	CASC5_ES_30004	-0.493	0.611	0.424	0.880	0.008
	DMKN_ES_49154	-0.431	0.650	0.435	0.970	0.035
ME	HMHA1_ES_46380	-0.917	0.400	0.222	0.719	0.002
	MSI2_ES_94619	-0.089	0.914	0.855	0.978	0.009
	CALCOCO2_ME_42227	-0.729	0.482	0.244	0.951	0.035
	OPN3_ME_204971	0.162	1.176	1.055	1.311	0.003
RI	SRGAP1_ME_93242	-0.084	0.920	0.878	0.964	0.000
	C11orf48_RI_16389	-2.401	0.091	0.028	0.294	0.000
	HES6_RI_58208	-0.052	0.950	0.918	0.983	0.003
	SCAF8_RI_78226	-2.832	0.059	0.010	0.349	0.002
All	SERAC1_RI_78257	-0.106	0.899	0.834	0.970	0.006
	ZNF329_RI_52413	-0.189	0.828	0.703	0.976	0.024
	IGF2BP2_AP_68029	-0.778	0.459	0.264	0.800	0.006
	C11orf48_RI_16389	-2.179	0.113	0.035	0.369	0.000
	ZNF185_ES_90404	-0.126	0.882	0.824	0.944	0.000
	YBEY_AD_60921	0.074	1.077	1.007	1.151	0.030
	FMNL3_ES_21603	-0.051	0.951	0.909	0.994	0.027
	ZHX3_RI_59397	-2.350	0.095	0.026	0.348	0.000

bolic enzymes, such as the glucose transporter 1 and hexokinase [42]. AS was also found to regulate some apoptotic genes, such as FAS, caspase 9, and BCL-X, resulting in cell survival [43]. Additionally, AS may significantly alter the coding region of drug targets, leading to increased drug resistance in some cancer therapies. Several classic key examples have been reported, such as BIM splicing and tyrosine-kinase inhibitor-resistant chronic myeloid leukemia, folypolyglutamate synthetase splicing

and antifolate resistant acute lymphoblastic leukemia, CD19 splicing and CART-19 immunotherapy resistant B-ALL, BRAF splicing and vemurafenib resistant melanoma, androgen receptor splicing and drug resistance in castration-resistant prostate cancer, and BRCA1 splicing and PARPi resistant breast cancer [44].

Over the last few years, an increasing number of AS events have been implicated in the progression of many types of cancers. In ovarian cancer, the AS of CD44 was shown to be regulated by ESRP1, enhancing the invasion and migration of ovarian cancer [45]. The regulation of cortactin exon 11 by PTBP1 was found by Wang et al. to promote the invasion of colorectal cancer [46]. In renal cancer, the AS of EZH2 by SF-3B3 was implicated in enhancing tumorigenic potential [47]. Similarly, the regulation of the AS of Bcl-x by BC200 was found to contribute to the pathogenesis of breast cancer [48]. Moreover, Johnson et al. found that the regulation of CPEB2 mRNA splicing was a key driving force in triple negative breast cancer metastasis [49]. Regarding ESCA, He et al. found that the splice variant of TCFC was significantly correlated with the tumorigenesis of ESCC [50]. Kahkhaie et al. also found that specific MUC1 splice variants

were involved in the progression of ESCA [51]. Moreover, in Wu's study, the AS of RUNX1 regulated by lincRNA-uc002yug.2 was suggested to be involved in the prognosis of ESCA [52]. Recently, Lin et al have investigated survival-associated splicing signatures in gastrointestinal pan-adenocarcinoma, and identified survival-associated AS events in EAC [26]. And Mao et al have also comprehensively investigated the prognostic splicing events in EAC and ESCC patients [32]. However, Mao's study mainly

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

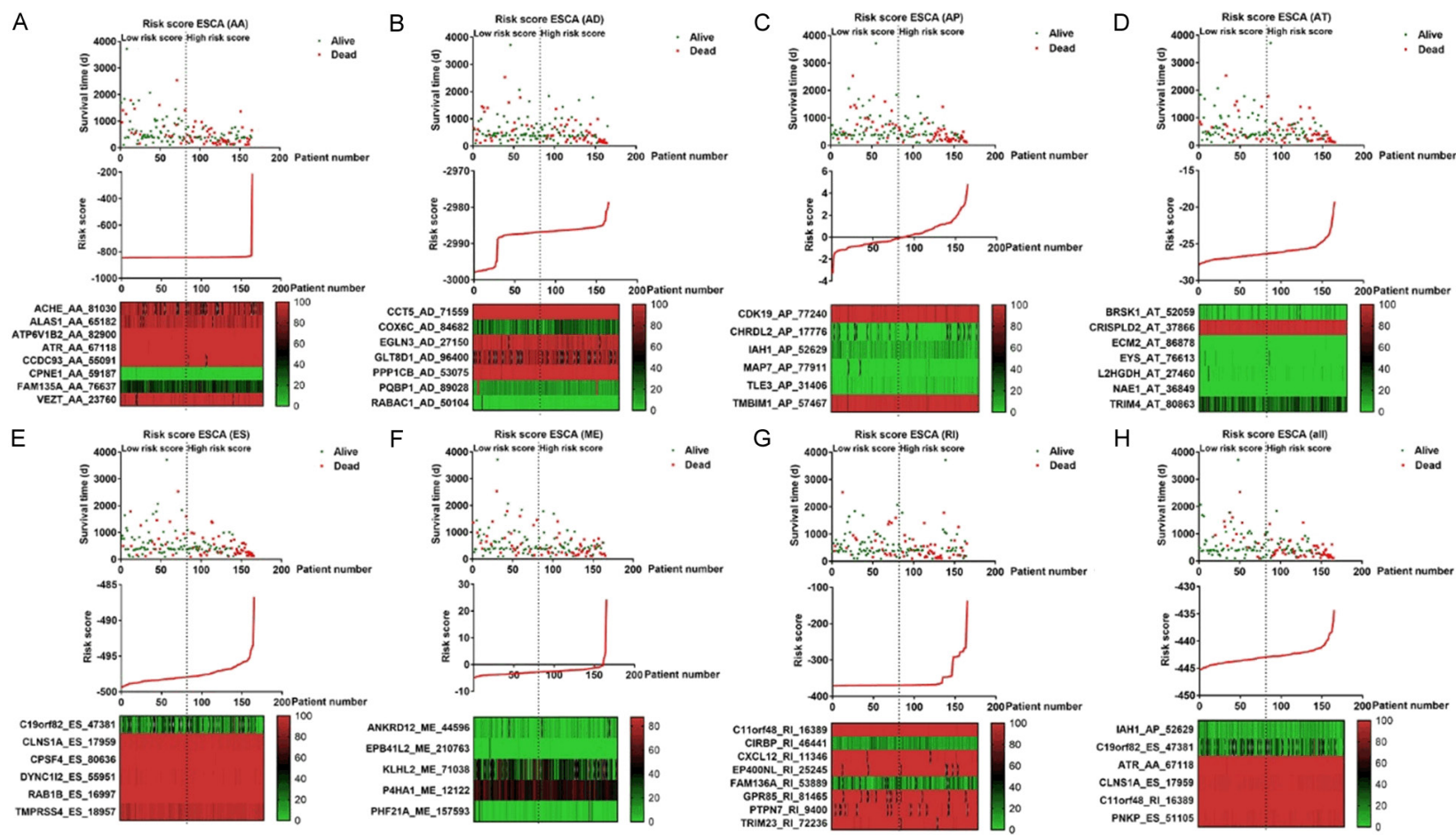


Figure 5. Construction of risk score for ESCA patients using prognostic signatures in multivariate Cox regression analysis. ESCA patients were divided into low-risk and high-risk cohorts using the median risk score value as a cut-off. The top part displays and sorts the patients' survival data based on risk score, the middle part shows the risk score's distribution curve, and the bottom part presents the heat map of the PSIs value of each prognostic signature. A-H. Risk scores were built with survival-associated splicing events for splicing AA, AD, AP, AT, ES, ME, RI and seven combined events in ESCA.

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

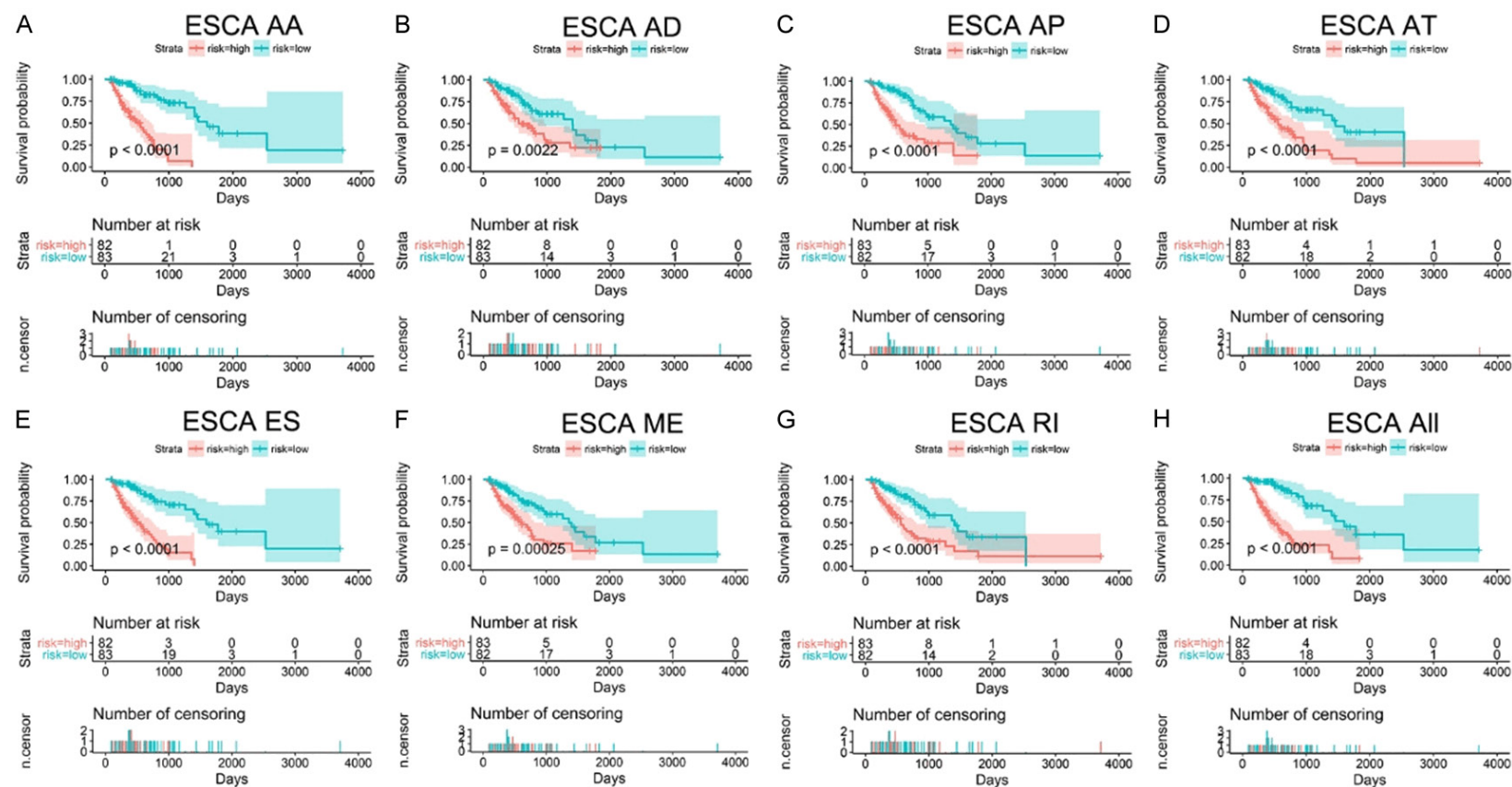


Figure 6. Prognostic models of AS events for ESCA patients. The top part of each drawing displays the survival curve of ESCA patients. Blue curves represent the low-risk group, while red curves represent the high-risk group. The middle part displays the number of high/low-risk patients at a certain time point, while the bottom part displays the relationship between the number of censoring and time. A-H. Kaplan-Meier curves of prognostic models were built with AA, AD, AP, AT, ES, ME, RI, and seven combined events in ESCA.

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

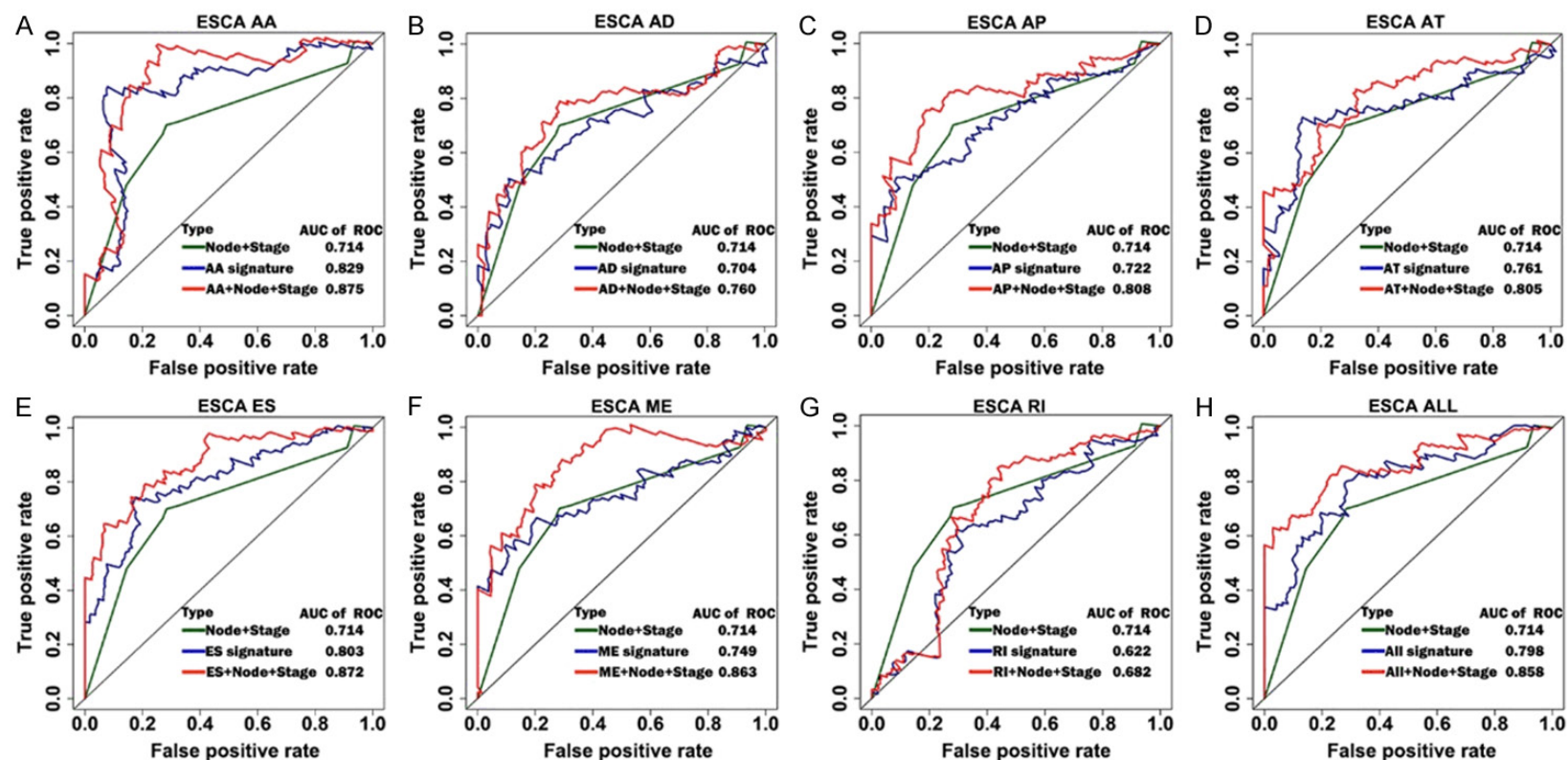


Figure 7. ROC curves for comparing prognostic models of independent AS events, independent clinical parameters, and the combination of both in ESCA. The horizontal axis represents false positive rate, while the vertical axis represents true positive rate. A-H. ROC curves of combined prognostic signatures were built with clinical parameters and each type of AS events in ESCA.

Table 4. Prognostic value of clinical parameters of esophageal carcinoma patients

Variables	Patient N = 168	Univariate Cox regression			Multivariate Cox regression		
		B	HR (95% CI)	P value	B	HR (95% CI)	P value
Age	168	0.072	1.074 (0.662-1.744)	0.771			
< 65	98						
≥ 65	70						
Gender	168	-0.639	0.528 (0.228-1.222)	0.136			
Male	141						
Female	27						
BMI	168	-0.078	0.925 (0.57-1.5)	0.752			
< 25	98						
≥ 25	70						
Pathological T stage	165	0.346	1.413 (0.855-2.335)	0.177			
T1+T2	72						
T3+T4	93						
Pathological N stage	154	1.253	3.501 (1.881-6.517)	< 0.001	1.049	2.855 (1.035-7.872)	0.043
N0	68						
N1-N3	86						
Pathological M stage	148	1.209	3.351 (1.766-6.361)	< 0.001	0.295	1.342 (0.438-4.117)	0.477
M0	133						
M1	15						
Pathological Stage	163	1.083	2.954 (1.779-4.907)	< 0.001	0.858	2.358 (1.007-5.526)	0.048
Stage I+II	95						
Stage III+IV	68						
Pathological Grade	133	0.615	1.849 (1.081-3.164)	0.025	0.014	0.986 (0.491-1.982)	0.993
Grade 1+2	86						
Grade 3	47						
Residual tumor	139	0.968	2.632 (1.331-5.205)	0.005	0.722	2.059 (0.846-5.012)	0.087
R0	125						
R1+R2	14						
Alcohol history	167	-0.319	0.727 (0.443-1.192)	0.207			
Yes	48						
No	119						
Smoke history	148	0.092	1.096 (0.746-1.611)	0.639			
Yes	96						
No	52						
Barretts esophagus	133	0.138	1.148 (0.643-2.048)	0.641			
Yes	27						
No	106						
Reflux history	143	0.192	1.212 (0.699-2.101)	0.494			
Yes	54						
No	89						

focused on the prognostic value of splicing events in EAC and ESCC, respectively. The influence of clinicopathological parameters in the survival of ESCA patients is no doubt a nonnegligible factor, which has not been investigated in Mao's study. And regarding the potential

molecular mechanism on how the splicing influence the survival of ESCA patient, little investigation has been achieved in Mao's study. Therefore, more exploration on combination of splicing and clinicopathological parameters, as well as potential mechanism of the survival-

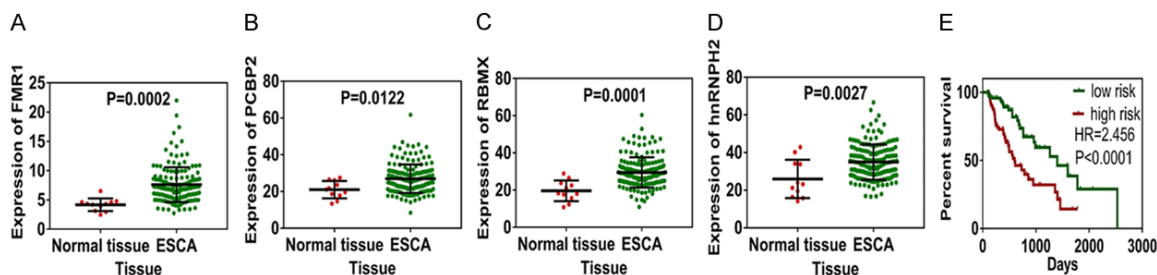


Figure 8. Expression of survival-associated splicing factors and construction of prognostic model. A-D. Expression of FMR1, PCBP2, RBMX, and hnRNPH2 was significantly up-regulated in ESCA patients. E. Prognostic model in ESCA was constructed by survival-associated splicing factors. Green curve represents the low-risk group, while red curve represents the high-risk group.

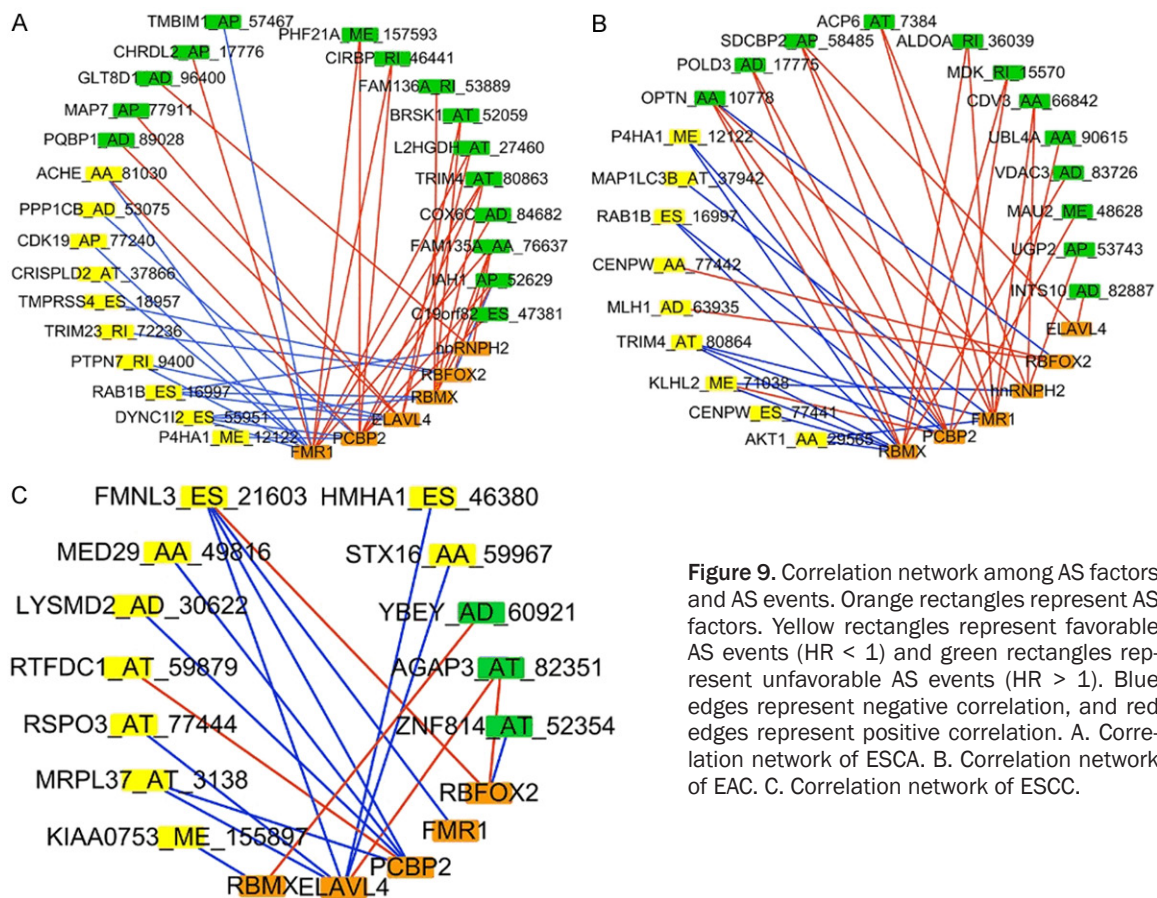


Figure 9. Correlation network among AS factors and AS events. Orange rectangles represent AS factors. Yellow rectangles represent favorable AS events (HR < 1) and green rectangles represent unfavorable AS events (HR > 1). Blue edges represent negative correlation, and red edges represent positive correlation. A. Correlation network of ESCA. B. Correlation network of EAC. C. Correlation network of ESCC.

associated splicing events in ESCA are needed. As a result, in our current study, we took clinicopathological parameters into consideration when constructing prognostic models. Regarding splicing factors, we also constructed prognostic signature, which further helps understand the role of splicing factors in ESCA. More importantly, we investigated the genomic alteration of the survival-associated splicing events, including gene amplification, mutation

and copy number variation, which could be the molecular mechanism of how splicing influence patient's survival.

In this study, we analyzed survival-associated AS events using the records of 165 ESCA patients consisting of 83 EAC and 82 ESCC patients from the TCGA database. A total of 3,276 AS events in ESCA, 4,850 events in EAC, and 2,199 events in ESCC were found to be sig-

nificantly associated with survival. Gene networks of ESCA, EAC, and ESCA uncovered significant hubs for GRB2, UBC, and KIF2A, respectively. Prognostic models were built for each type of AS event and combined AS events in ESCA and its histological subtypes. Most of the models showed satisfactory predictive efficacy for the survival of patients. The predictive models of the EAC group tended to show better predictive efficacy than those of the ESCC group. However, the seven combined events prognostic models in ESCA and its subtypes did not show as high of predictive efficacy as expected. A similar situation was also observed in another study, which investigated the prognostic role of AS events in prostate adenocarcinoma [27]. This might suggest that some survival-associated genes in ESCA and its histological subtypes may undergo the same type of splicing and exert more powerful effects on survival. Due to the limitation of sample sizes, another independent cohort is urgently required to validate these findings in the future. These predictors may greatly help predict the outcome of ESCA patients, providing important information for clinical practice. More importantly, we also analyzed the prognostic value of clinicopathological parameters, which showed pathological T stage and pathological stage were independent survival-associated factors. We constructed novel combined prognostic models using these two survival-associated clinicopathological parameters and the eight sets of AS signatures. Interestingly, the combined prognostic models had better performance in predicting survival of ESCA than independent splicing models or clinical parameter model. These indicated that the prognostic significance of clinicopathological parameters should not be ignored when constructing prognostic models. It would be more effective and practical to use the combined prognostic models in clinical application.

AS factors have been previously found to exert regulatory effects on AS events, which play vital roles in oncogenesis and tumor progression. Therefore, we wondered what AS factors may be significantly correlated with survival-associated AS events in ESCA patients. We collected the expression profiles of 66 AS factors from the TCGA database and selected survival-associated AS factors. A total of six survival-associated AS factors-FMR1, ELAVL4, PCBP2, RBMX, RBFOX2, and hnRNPH2-were found, among

which, four-ELAVL4, FMR1, PCBP2, and RBFOX2-were determined to be independent predictors. The low-risk score of the four predictors indicated significantly better survival of ESCA patients (hazard ratio [HR] = 2.456). As we know, AS factors exert functions by regulating AS events. Therefore, it would very helpful to illuminate what AS events the AS factors may influence. By virtue of Spearman's correlation analysis, we found that the majority of the favorable AS events were negatively correlated with AS factors, while the unfavorable ones showed the opposite correlation. In fact, unfavorable events and positive correlations in EAC tended to be more common than that in ESCC. The regulatory network may help us to understand the regulatory mechanisms of AS events in ESCA and its histological subtypes. However, whether the high expression of AS factors influenced the survival of ESCA patients by down-regulating AS events still requires functional experiments be performed. Some genomic evidence of the effects of AS factors in cancer has included mutation, gene amplifications, and copy number variation [53]. So, as a result, we decided to investigate the mutation and copy number variation of the eight sets of gene symbols in ESCA prognostic signatures. Interestingly, we discovered that most of the genes undergo mutations including amplifications and deep deletion. In addition, copy number losses or amplifications were also common in the gene symbols. These abnormalities might be the molecular basis of how splicing affects survival.

Several limitations need to be addressed in the current study. The results here were based on 165 ESCA patients (83 EAC and 82 ESCC), which is a limited sample size. Another independent cohort would greatly strengthen the findings, which we hope to achieve in future research. Moreover, the data in the current study was based on an online database and thus remains at a bioinformatics level. It is necessary to adopt experimental methods to verify the findings. Furthermore, we only investigated the regulatory relationship between AS events and classic splicing factors in the current study. However, other RNA-binding proteins could also regulate AS events, which still requires further investigation in the future.

In general, we provided an overview of survival-associated AS events and built valuable prog-

nostic models for predicting the outcomes of ESCA, EAC, and ESCC patients. A number of promising therapeutic targets in ESCA patients, which have great potential and need to be validated in the future, were provided. Importantly, after combining the clinicopathological parameters and splicing signatures, we constructed more effective prognostic models in predicting the survival of ESCA patients. Moreover, a regulatory network of AS factors and events was constructed, which hopefully will illuminate the regulatory mechanisms of AS in the initiation and progression of ESCA and its histological subtypes. Additionally, the findings in the mutation and copy number variation of prognostic signatures also provided more molecular perspectives in the mechanisms of splicing as related to survival. In sum, the findings in the current study may provide a basis for spliceosomes in ESCA and its subtypes, and the methods used in this study could provide novel perspectives in other fields of tumor study to help shed light on future oncology research.

Acknowledgements

This study was funded by the Guangxi Medical University Training Program for Distinguished Young Scholars (2017). The funding provider was not involved in the study design, data collection or analysis, preparation of the manuscript, or the decision to publish. And the authors would like to thank the National Cancer Institute for access to the TCGA database and for its valuable data.

Disclosure of conflict of interest

None.

Address correspondence to: Gang Chen, Department of Pathology, First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning 530021, Guangxi Zhuang Autonomous Region, P. R. China. E-mail: chengang@gxmu.edu.cn

References

- [1] Hong L, Han Y, Zhang H and Fan D. Prognostic markers in esophageal cancer: from basic research to clinical use. *Expert Rev Gastroenterol Hepatol* 2015; 9: 887-889.
- [2] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2017. *CA Cancer J Clin* 2017; 67: 7-30.
- [3] Kataoka K, Nakamura K, Mizusawa J, Kato K, Eba J, Katayama H, Shibata T and Fukuda H.

Surrogacy of progression-free survival (PFS) for overall survival (OS) in esophageal cancer trials with preoperative therapy: Literature-based meta-analysis. *Eur J Surg Oncol* 2017; 43: 1956-1961.

- [4] Tang H, Wu Y, Qin Y, Wang H, Wang L, Guan X, Luo S and Wang Q. Reduction of AZGP1 predicts poor prognosis in esophageal squamous cell carcinoma patients in Northern China. *Onco Targets Ther* 2017; 10: 85-94.
- [5] Tan HZ, Wu ZY, Wu JY, Long L, Jiao JW, Peng YH, Xu YW, Li SS, Wang W, Zhang JJ, Li EM and Xu LY. Single nucleotide polymorphism rs1304-2395 in the SLC52A3 gene as a biomarker for regional lymph node metastasis and relapse-free survival of esophageal squamous cell carcinoma patients. *BMC Cancer* 2016; 16: 560.
- [6] Wang L, Yu X, Li J, Zhang Z, Hou J and Li F. Prognostic significance of p53 expression in patients with esophageal cancer: a meta-analysis. *BMC Cancer* 2016; 16: 373.
- [7] Vrana D, Matzenauer M, Aujesky R, Vrba R, Neoral C, Melichar B and Soucek P. Potential predictive role of microRNAs in the neoadjuvant treatment of esophageal cancer. *Anti-cancer Res* 2017; 37: 403-412.
- [8] Hemmatzadeh M, Mohammadi H, Karimi M, Musavishenas MH and Baradaran B. Differential role of microRNAs in the pathogenesis and treatment of esophageal cancer. *Biomed Pharmacother* 2016; 82: 509-519.
- [9] Hong L, Han Y, Zhang H, Zhao Q, Wu K and Fan D. Prognosis-related microRNAs in esophageal cancer. *Expert Opin Biol Ther* 2014; 14: 483-489.
- [10] Lv H, He Z, Wang H, Du T and Pang Z. Differential expression of miR-21 and miR-75 in esophageal carcinoma patients and its clinical implication. *Am J Transl Res* 2016; 8: 3288-3298.
- [11] Zhou DD, Liu XF, Lu CW, Pant OP and Liu XD. Long non-coding RNA PVT1: emerging biomarker in digestive system cancer. *Cell Prolif* 2017; 50: e12398.
- [12] Sugihara H, Ishimoto T, Miyake K, Izumi D, Baba Y, Yoshida N, Watanabe M and Baba H. Noncoding RNA expression aberration is associated with cancer progression and is a potential biomarker in esophageal squamous cell carcinoma. *Int J Mol Sci* 2015; 16: 27824-27834.
- [13] Hu HB, Jie HY and Zheng XX. Three circulating lncRNA predict early progress of esophageal squamous cell carcinoma. *Cell Physiol Biochem* 2016; 40: 117-125.
- [14] Yao J, Huang JX, Lin M, Wu ZD, Yu H, Wang PC, Ye J, Chen P, Wu J and Zhao GJ. Microarray expression profile analysis of aberrant long non-coding RNAs in esophageal squamous cell carcinoma. *Int J Oncol* 2016; 48: 2543-2557.

- [15] Visser E, Franken IA, Brosens LA, Ruurda JP and van Hillegersberg R. Prognostic gene expression profiling in esophageal cancer: a systematic review. *Oncotarget* 2017; 8: 5566-5577.
- [16] Tan C, Qian X, Guan Z, Yang B, Ge Y, Wang F and Cai J. Potential biomarkers for esophageal cancer. *Springerplus* 2016; 5: 467.
- [17] Oltean S and Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene* 2014; 33: 5311-5318.
- [18] Fiszbein A and Kornblihtt AR. Alternative splicing switches: Important players in cell differentiation. *Bioessays* 2017; 39: 1600157.
- [19] Chen J and Weiss WA. Alternative splicing in cancer: implications for biology and therapy. *Oncogene* 2015; 34: 1-14.
- [20] Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M and Stamm S. Function of alternative splicing. *Gene* 2013; 514: 1-30.
- [21] Lin JC. Therapeutic applications of targeted alternative splicing to cancer treatment. *Int J Mol Sci* 2017; 19: 75.
- [22] Song X, Zeng Z, Wei H and Wang Z. Alternative splicing in cancers: from aberrant regulation to new therapeutics. *Semin Cell Dev Biol* 2018; 75: 13-22.
- [23] Shi Y, Chen Z, Gao J, Wu S, Gao H and Feng G. Transcriptome-wide analysis of alternative mRNA splicing signature in the diagnosis and prognosis of stomach adenocarcinoma. *Oncol Rep* 2018; 40: 2014-2022.
- [24] Xiong Y, Deng Y, Wang K, Zhou H, Zheng X, Si L and 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.
- [25] Liu J, Li H, Shen S, Sun L, Yuan Y and Xing C. Alternative splicing events implicated in carcinogenesis and prognosis of colorectal cancer. *J Cancer* 2018; 9: 1754-1764.
- [26] Lin P, He RQ, Ma FC, Liang L, He Y, Yang H, Dang YW and Chen G. Systematic analysis of survival-associated alternative splicing signatures in gastrointestinal pan-adenocarcinomas. *EBioMedicine* 2018; 34: 46-60.
- [27] Huang ZG, He RQ and Mo ZN. Prognostic value and potential function of splicing events in prostate adenocarcinoma. *Int J Oncol* 2018; 53: 2473-2487.
- [28] He RQ, Zhou XG, Yi QY, Deng CW, Gao JM, Chen G and Wang QY. Prognostic signature of alternative splicing events in bladder urothelial carcinoma based on spliceseq data from 317 cases. *Cell Physiol Biochem* 2018; 48: 1355-1368.
- [29] Suo C, Hrydziuszko O, Lee D, Pramana S, Saputra D, Joshi H, Calza S and Pawitan Y. Integration of somatic mutation, expression and functional data reveals potential driver genes predictive of breast cancer survival. *Bioinformatics* 2015; 31: 2607-2613.
- [30] Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z and He J. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer Lett* 2017; 393: 40-51.
- [31] Zhu J, Chen Z and Yong L. Systematic profiling of alternative splicing signature reveals prognostic predictor for ovarian cancer. *Gynecol Oncol* 2018; 148: 368-374.
- [32] Mao S, Li Y, Lu Z, Che Y, Sun S, Huang J, Lei Y, Wang X, Liu C, Zheng S, Zang R, Li N, Li J, Sun N and He J. Survival-associated alternative splicing signatures in esophageal carcinoma. *Carcinogenesis* 2018; 40: 121-130.
- [33] Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, Melott J and Weinstein J. TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res* 2016; 44: D1018-1022.
- [34] Ryan MC, Cleland J, Kim R, Wong WC and Weinstein JN. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics* 2012; 28: 2385-2387.
- [35] Siriwardhana C, Khadka VS, Chen JJ and Deng Y. Development of a miRNA-seq based prognostic signature in lung adenocarcinoma. *BMC Cancer* 2019; 19: 34.
- [36] Lex A, Gehlenborg N, Strobel H, Vuilleumot R and Pfister H. UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 2014; 20: 1983-1992.
- [37] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498-2504.
- [38] Piva F, Giulietti M, Burini AB and Principato G. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum Mutat* 2012; 33: 81-85.
- [39] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C and Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; 6: pl1.
- [40] Albert R and Barabási AL. Statistical mechanics of complex networks. *Reviews of Modern Physics* 2002; 74: 47-97.
- [41] Le KQ, Prabhakar BS, Hong WJ and Li LC. Alternative splicing as a biomarker and potential target for drug discovery. *Acta Pharmacol Sin* 2015; 36: 1212-1218.
- [42] Kozlovski I, Siegfried Z, Amar-Schwartz A and Karni R. The role of RNA alternative splicing in

- regulating cancer metabolism. *Hum Genet* 2017; 136: 1113-1127.
- [43] Paronetto MP, Passacantilli I and Sette C. Alternative splicing and cell survival: from tissue homeostasis to disease. *Cell Death Differ* 2016; 23: 1919-1929.
- [44] Siegfried Z and Karni R. The role of alternative splicing in cancer drug resistance. *Curr Opin Genet Dev* 2017; 48: 16-21.
- [45] Chen L, Yao Y, Sun L, Zhou J, Miao M, Luo S, Deng G, Li J, Wang J and Tang J. Snail driving alternative splicing of CD44 by ESRP1 enhances invasion and migration in epithelial ovarian cancer. *Cell Physiol Biochem* 2017; 43: 2489-2504.
- [46] Wang ZN, Liu D, Yin B, Ju WY, Qiu HZ, Xiao Y, Chen YJ, Peng XZ and Lu CM. High expression of PTBP1 promote invasion of colorectal cancer by alternative splicing of cortactin. *Oncotarget* 2017; 8: 36185-36202.
- [47] Chen K, Xiao H, Zeng J, Yu G, Zhou H, Huang C, Yao W, Xiao W, Hu J, Guan W, Wu L, Huang J, Huang Q, Xu H and Ye Z. Alternative splicing of EZH2 pre-mRNA by SF3B3 contributes to the tumorigenic potential of renal cancer. *Clin Cancer Res* 2017; 23: 3428-3441.
- [48] Singh R, Gupta SC, Peng WX, Zhou N, Pochampally R, Atfi A, Watabe K, Lu Z and Mo YY. Regulation of alternative splicing of Bcl-x by BC200 contributes to breast cancer pathogenesis. *Cell Death Dis* 2016; 7: e2262.
- [49] Johnson RM, Vu NT, Griffin BP, Gentry AE, Archer KJ, Chalfant CE and Park MA. The alternative splicing of cytoplasmic polyadenylation element binding protein 2 drives anoikis resistance and the metastasis of triple negative breast cancer. *J Biol Chem* 2015; 290: 25717-25727.
- [50] He G, Guan X, Chen X, Wang Y, Luo C and Zhang B. Expression and splice variant analysis of human TCF4 transcription factor in esophageal cancer. *J Cancer* 2015; 6: 333-341.
- [51] Kakhkhaie KR, Moaven O, Abbaszadegan MR, Montazer M and Gholamin M. Specific MUC1 splice variants are correlated with tumor progression in esophageal cancer. *World J Surg* 2014; 38: 2052-2057.
- [52] Wu H, Zheng J, Deng J, Zhang L, Li N, Li W, Li F, Lu J and Zhou Y. LincRNA-uc002yug.2 involves in alternative splicing of RUNX1 and serves as a predictor for esophageal cancer and prognosis. *Oncogene* 2015; 34: 4723-4734.
- [53] Shilo A, Siegfried Z and Karni R. The role of splicing factors in deregulation of alternative splicing during oncogenesis and tumor progression. *Mol Cell Oncol* 2015; 2: e970955.

1

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

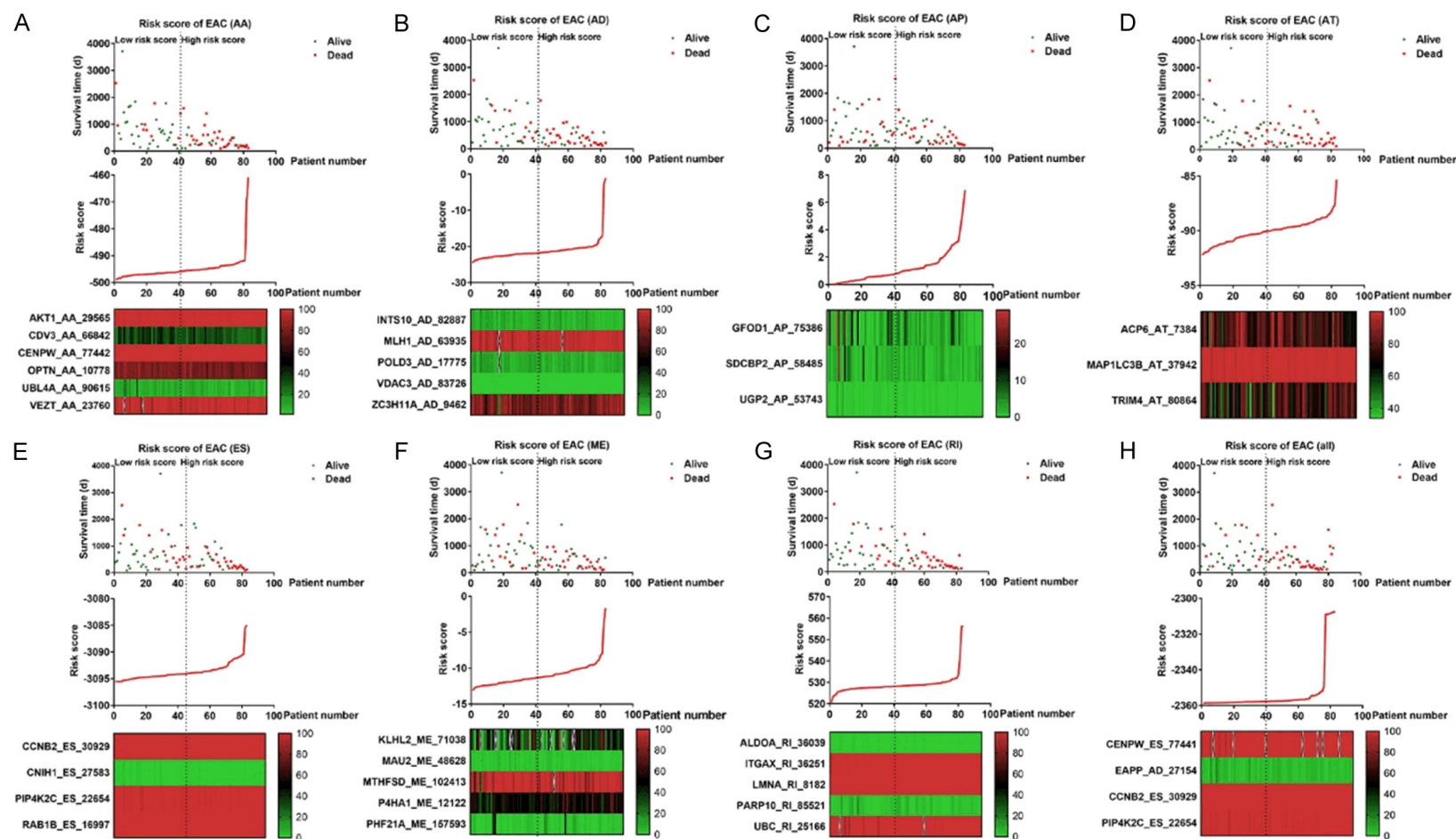


Figure S2. Construction of risk score for EAC patients using prognostic signatures in multivariate Cox regression analysis. EAC patients were divided into low-risk and high-risk cohorts using the median risk score value as a cut-off. The top part displays and sorts the patients' survival data based on risk score, the middle part shows the risk score's distribution curve, and the bottom part presents the heat map of the PSIs value of each prognostic signature. A-H. Risk scores were built with survival-associated splicing events for each type of splicing and seven combined splicing in EAC.

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

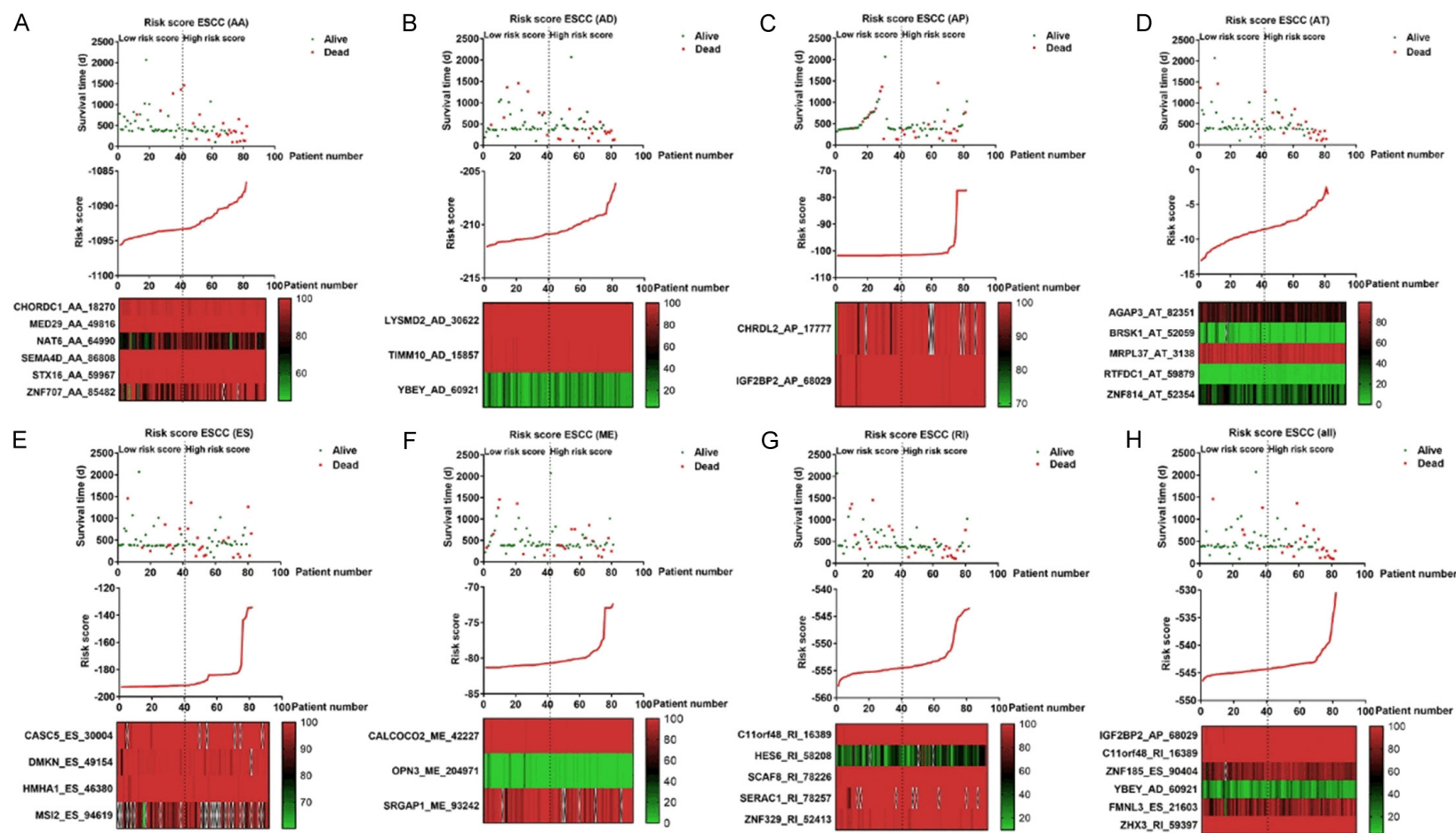
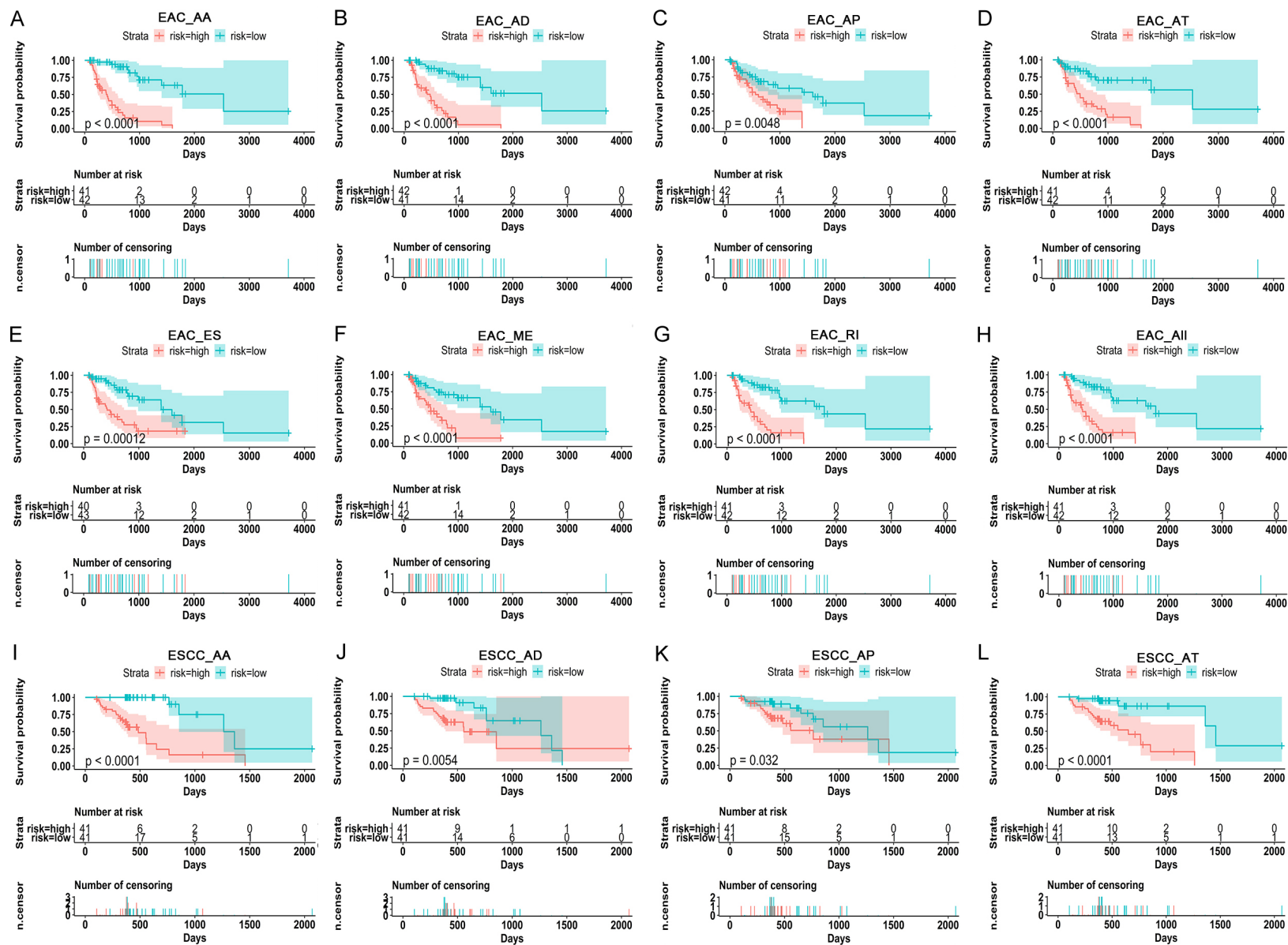


Figure S3. Construction of risk score for ESCC patients using prognostic signatures in multivariate Cox regression analysis. ESCC patients were divided into low-risk and high-risk cohorts using the median risk score value as a cut-off. The top part displays and sorts the patients' survival data based on risk score, the middle part shows the risk score's distribution curve, and the bottom part presents the heatmap of the PSIs value of each prognostic signature. A-H. Risk scores were built with survival-associated splicing events for each type of splicing and seven combined splicing in ESCC.

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma



Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

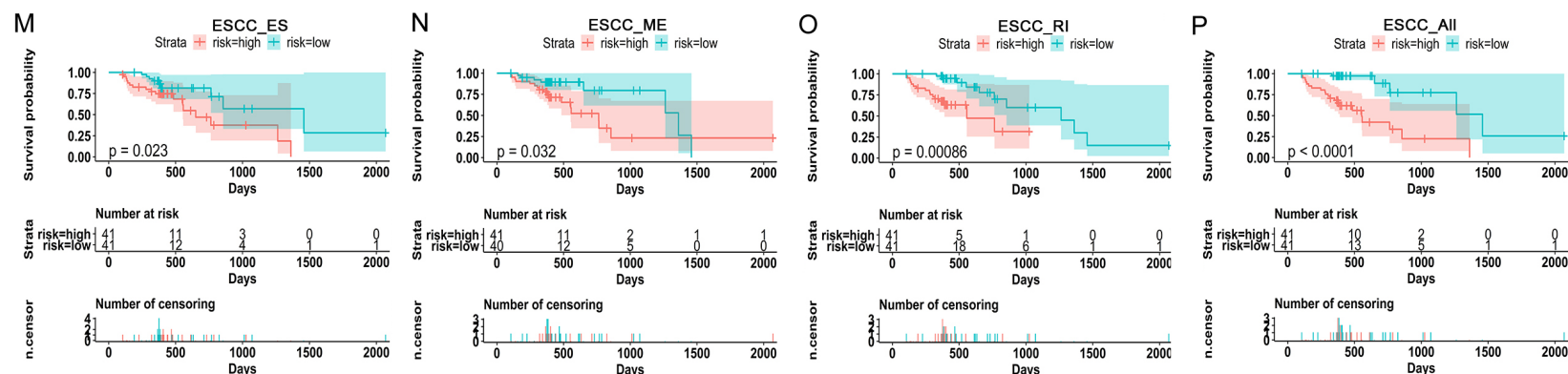


Figure S4. Prognostic models of AS events for EAC and ESCC patients. The top part of each drawing displays the survival curve of ESCA patients. Blue curves represent the low-risk group, while red curves represent the high-risk group. The middle part displays the number of high/low-risk patients at a certain time point, while the bottom part displays the relationship between the number of censoring and time. A-P. survival curves of prognostic signatures were built with a single type of AS event and seven combined events in EAC and ESCC.

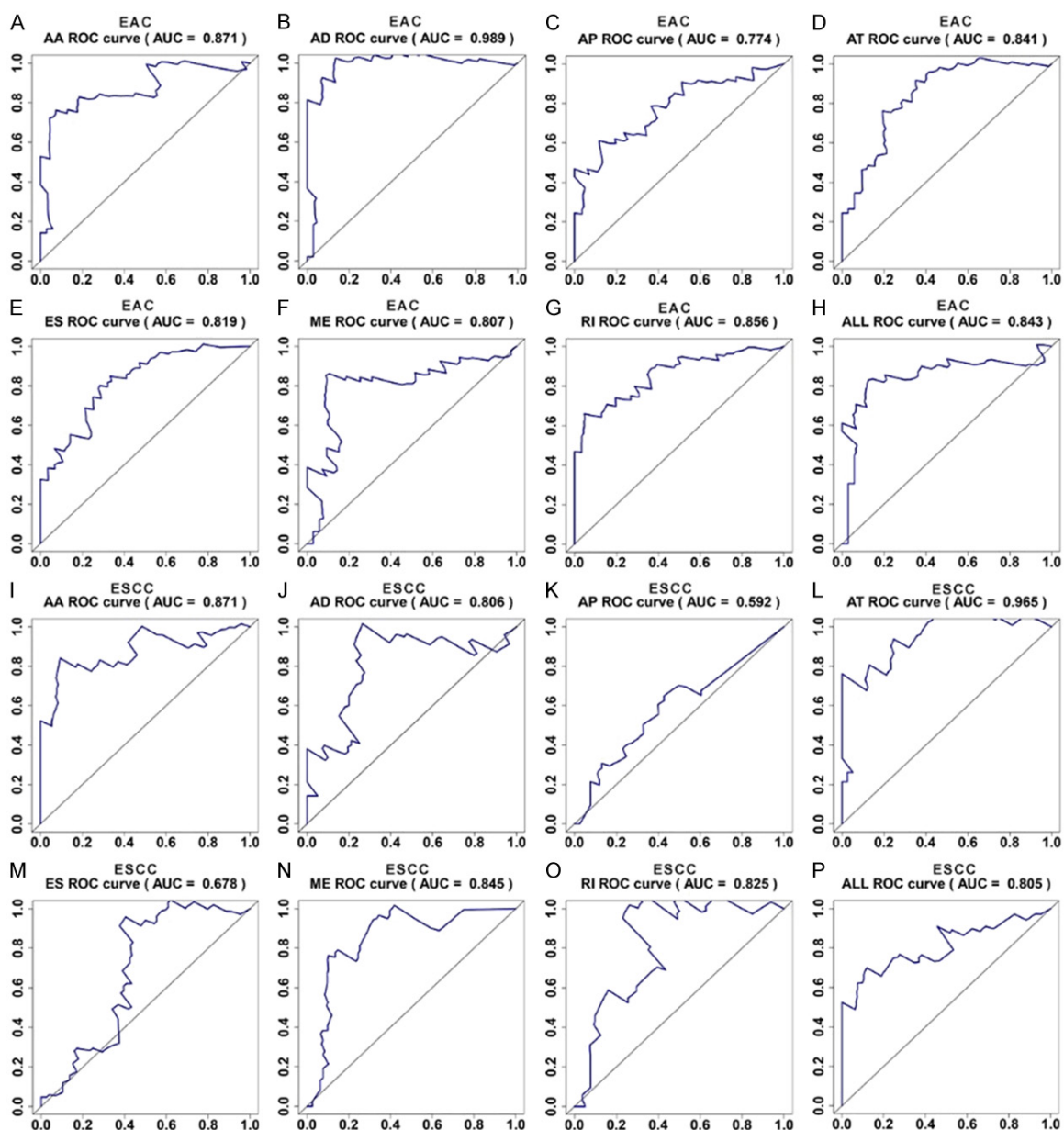


Figure S5. ROC curves for comparing prognostic models of AS events in EAC and ESCC. The horizontal axis represents false positive rate, while the vertical axis represents true positive rate. A-P. ROC curves of prognostic signatures were built with a single type of AS event and seven combined events in EAC and ESCC.

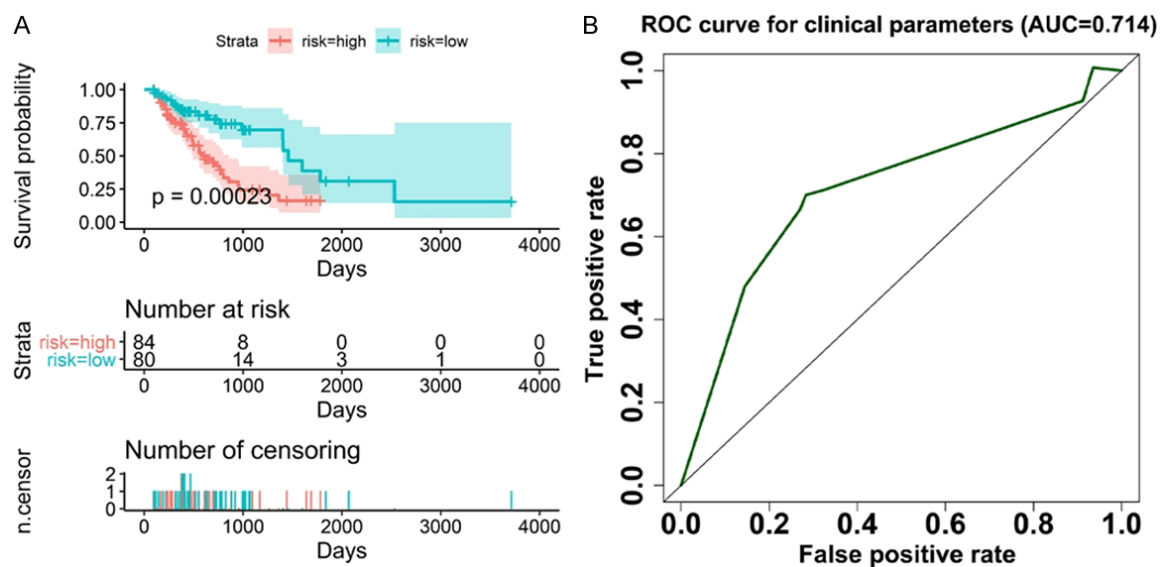


Figure S6. The survival curve and ROC curve of the prognostic models of clinicopathological parameters in ESCA. A. The top part of the drawing displays the survival curve of ESCA patients. Blue curves represent the low-risk group, while red curves represent the high-risk group. The middle part displays the number of high/low-risk patients at a certain time point, while the bottom part displays the relationship between the number of censoring and time. B. The horizontal axis represents false positive rate, while the vertical axis represents true positive rate.

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

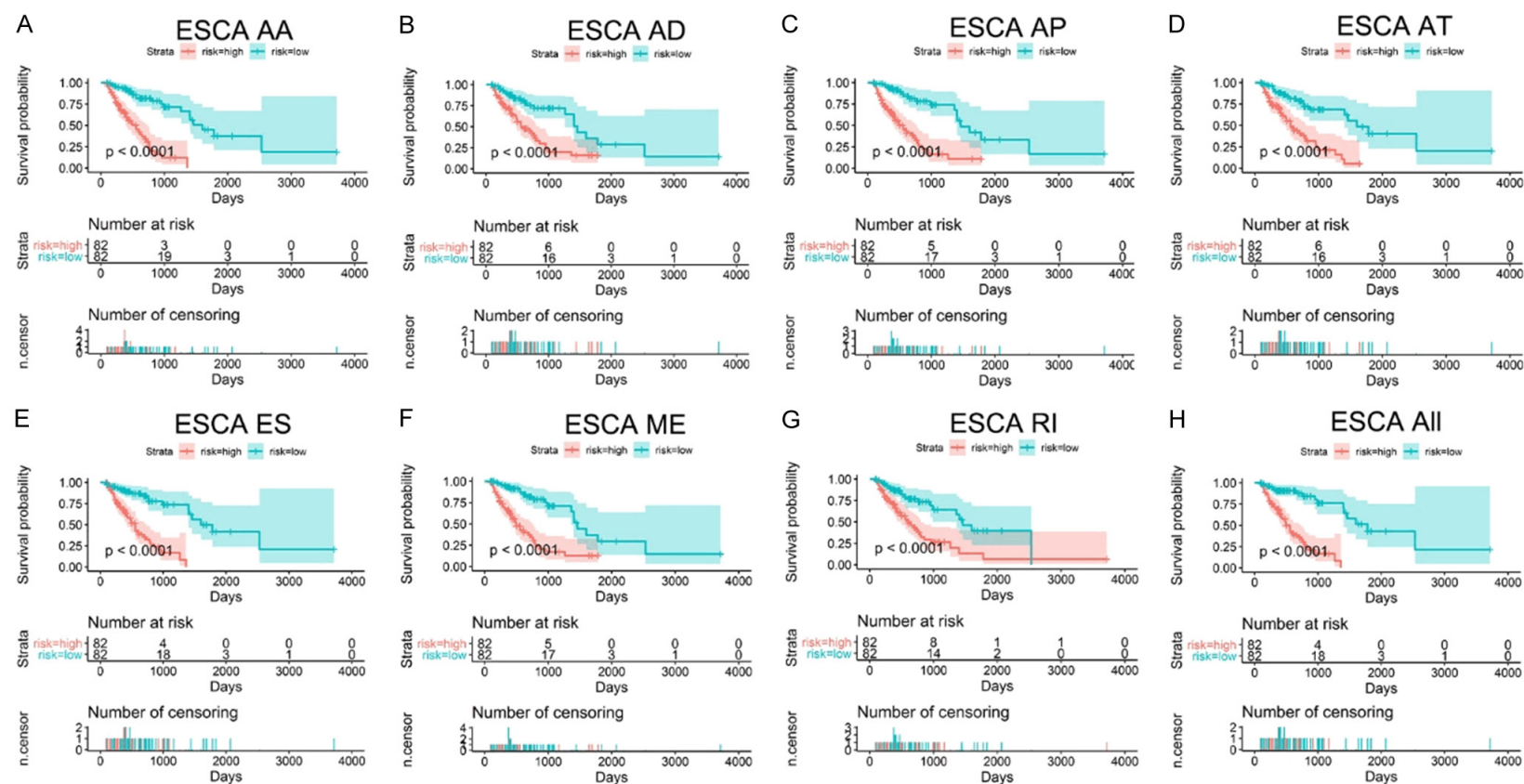


Figure S7. Combined prognostic models of AS events and clinical parameters for ESCA patients. The top part of each drawing displays the survival curve of ESCA patients. Blue curves represent the low-risk group, while red curves represent the high-risk group. The middle part displays the number of high/low-risk patients at a certain time point, while the bottom part displays the relationship between the number of censoring and time. A-H. Kaplan-Meier curves of prognostic models were built with clinical parameters and each type of AS events in ESCA.

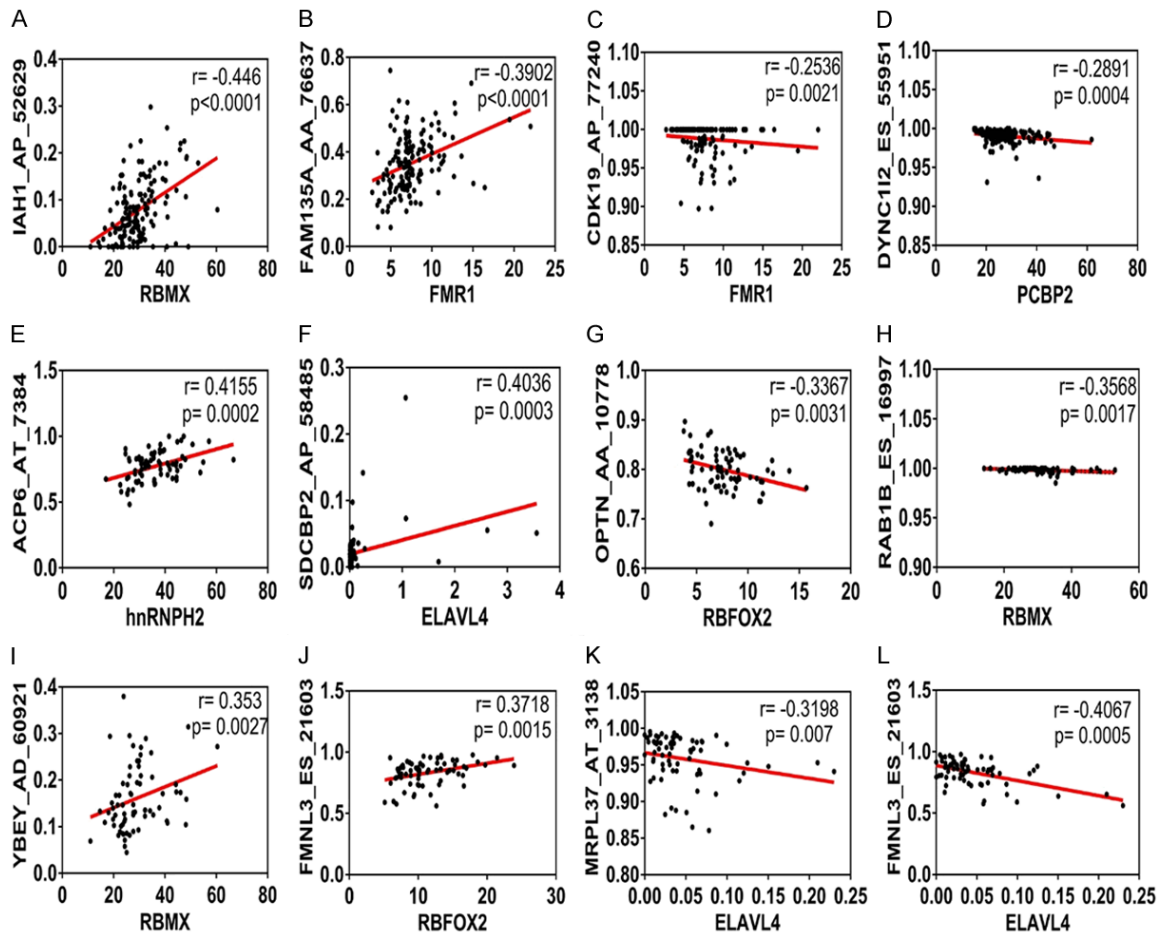


Figure S8. Linear correlation plots between survival-associated splicing factors and AS events. The horizontal axis represents the expression of splicing factor, while the vertical axis represents the percent spliced in value of AS event. A-D. Most significant linear correlation plots in ESCA. E-H. Most significant linear correlation plots in EAC. I-L. Most significant linear correlation plots in ESCC.

Prognostic alternative splicing signatures and regulatory network in esophageal carcinoma

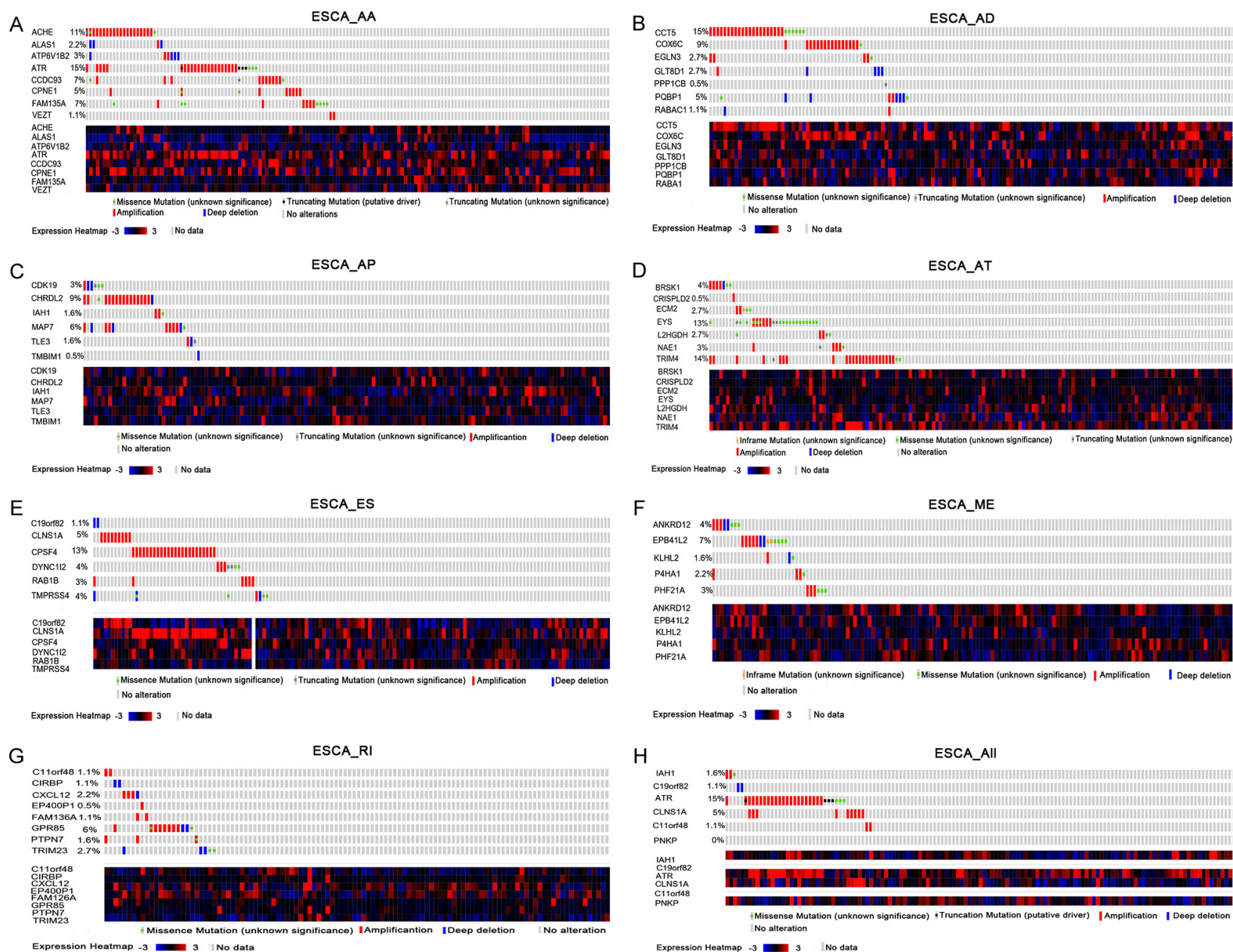


Figure S9. Genomic alteration of each set of gene symbols in the prognostic signatures in ESCA. The top part displays the mutation rate and types, and the bottom part presents the heat map showing the expression of each gene symbol. A-H. Mutation plots and heat map for splicing AA, AD, AP, AT, ES, ME, RI and seven combined events.

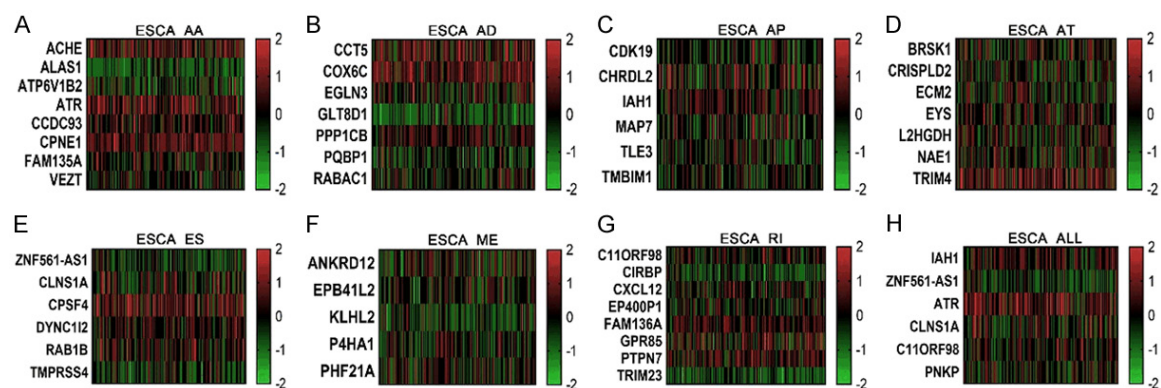


Figure S10. Analysis of copy number variation of each set of gene symbols in the prognostic signatures in ESCA. A-H. Heat map displaying the segment mean value of copy number for splicing AA, AD, AP, AT, ES, ME, RI, and seven combined events.