

## Development and validation of an individual alternative splicing prognostic signature in gastric cancer

Shenghan Lou<sup>1,\*</sup>, Jian Zhang<sup>2,\*</sup>, Zhao Zhai<sup>1</sup>, Xin Yin<sup>1</sup>, Yimin Wang<sup>1</sup>, Tianyi Fang<sup>1</sup>, Yingwei Xue<sup>1</sup>

<sup>1</sup>Department of Gastroenterological Surgery, Harbin Medical University Cancer Hospital, Harbin 150081, Heilongjiang Province, China

<sup>2</sup>Department of Thoracic Surgery, Harbin Medical University Cancer Hospital, Harbin 150081, Heilongjiang Province, China

\*Equal contribution

**Correspondence to:** Yingwei Xue; email: [xueyingwei@hrbmu.edu.cn](mailto:xueyingwei@hrbmu.edu.cn)

**Keywords:** gastric cancer, alternative splicing, prognostic indicator

**Received:** October 15, 2020

**Accepted:** December 23, 2020

**Published:** February 17, 2021

**Copyright:** © 2021 Lou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

Gastric cancer (GC) is a heterogeneous disease with different clinical manifestations and prognoses. Alternative splicing (AS) is a determinant of gene expression and contributes to protein diversity from a rather limited gene transcript in metazoans. AS events are associated with different aspects of cancer biology, including cell proliferation, apoptosis, invasion, etc. Here, we present a comprehensive analysis of the prognostic AS profile in GC. GC-specific AS (GCAS) events were analyzed, and overall survival-associated GCAS (OS-GCAS) events were verified among the genome-wide AS events identified in The Cancer Genome Atlas (TCGA) database. In total, 1,287 GCAS events of 837 genes and 173 OS-GCAS events of 130 genes were identified. The parental genes of OS-GCAS events were significantly enriched in the development of GC. Protein-protein interaction (PPI) and OS-GCAS-associated splicing factor (SF) interaction networks were constructed. Multivariate Cox regression analysis with least absolute shrinkage and selection operator (LASSO) penalty was performed to establish a prognostic risk formula, representing 23 OS-GCAS events. The low-risk group had better OS than the high-risk group and lower immune and stromal scores. Cox proportional hazard regression was applied to generate an AS-clinical integrated prognostic model with a considerable area under the curve (AUC) value in both the training and validation datasets. Our study provides a profile of OS-GCAS events and an AS-clinical nomogram to predict the prognosis of GC.

### INTRODUCTION

Gastric cancer (GC) is the fifth most frequently diagnosed cancer and the third leading cause of cancer-related mortality worldwide, with an incidence of approximately 1,000,000 cases, accounting for approximately 800,000 deaths every year [1, 2]. Owing to the lack of early symptoms, most GC patients are diagnosed at an advanced stage [3–5]. Despite effective treatment, relapse and metastasis are common for advanced GC, which causes a quite low survival rate [3, 4, 6]. Currently, the tumor-node-metastasis (TNM)

system is still the most common assessment system in predicting GC prognosis. However, the prognosis of GC patients within the same TNM stage is usually different, suggesting that the TNM system is not sufficiently satisfactory to predict GC prognosis [7–9]. Thus, a novel effective signature for predicting the prognosis of GC is urgently needed.

Alternative splicing (AS) is one of the most critical posttranscriptional regulatory processes and modifies more than 90% of all human genes [10]. The process of AS significantly contributed to protein diversity by

generating different mRNA isoforms from a single gene [11]. Splicing abnormalities may cause a series of consequences from changing the stability to adding or deleting structural domains, modifying the interactive relationship of proteins, which significantly alters the abundance and complexity of the organism's protein-protein interactions (PPIs) [12, 13]. Abnormal AS events participate in several tumorigenic processes, such as proliferation, apoptosis, hypoxia, angiogenesis, immune escape and metastasis [14, 15]. More importantly, growing evidence has demonstrated that AS has clinical potential in cancer therapy [12, 16, 17]. Thus, cancer-specific AS events might serve not only as prognostic signatures but also as potential therapeutic targets.

Despite the indispensable role of AS in oncogenesis, systematic analyses of the clinical significance of AS events in GC and the potential regulatory mechanism are still lacking. Thus, our study aimed to conduct systematic profiling of genome-wide AS events and their prognostic value associated with GC. First, we identified GC-specific AS (GCAS) events and investigated the relationship between GCAS events and overall survival (OS). The potential biological function and underlying molecular regulatory network of these OS-GCAS events were further explored. Second, we constructed an AS prognostic signature, which could stratify risk for patients with GC. We also uncovered the distinct clinical, molecular and immune features between high- and low-risk patients. Finally, by integrating the AS signature with clinical characteristics, an AS-clinical nomogram with high performance was established for clinical application.

## RESULTS

### Overview of AS events in the TCGA GC cohort

We preliminarily detected 119,697 AS events from 14,433 genes, which accounted for approximately 70% of protein-coding genes [18]. A large proportion of AS events could only be detected in a few samples, and specific splicing isoforms were barely detected (PSI value < 0.05). Then, we implemented a series of filters (percentage of samples with PSI values  $\geq 75$  and average PSI value  $\geq 0.05$ ) to obtain a reliable set of AS events in GC. Consequently, a total of 37,139 AS events were obtained from 10,380 genes, which were used for further analysis. Among these AS events, they were comprised of “15,816 ES events from 6,462 genes”, “6,633 AP events from 4,016 genes”, “5,988 AT events from 3,665 genes”, “3,285 AA events from 2,404 genes”, “2,691 AD events from 1,986 genes”, “2,564 RI events from 1,780 genes” and “162 ME events in 156 genes” (Supplementary Figure 1A). Among these seven

types of splicing patterns, ES was the most frequent splicing pattern, followed by AP and AT, while ME was the least frequent splicing pattern. Of note, the number of AS events far exceeded the total number of genes, which indicated that one gene might undergo multiple splicing patterns. As shown in Supplementary Figure 1B, most genes contained two or more AS events, and several genes could have up to five different splicing modes.

### Identification of GC-associated AS events

To identify GC-associated AS events, we compared the PSI values between 33 paired tumor and normal tissues. According to the defined threshold, a total of 1,287 GCAS events were identified from 837 genes, which included 738 upregulated AS events from 625 genes and 549 downregulated AS events from 509 genes (Figure 1A and Supplementary Table 1). The proportion of AS splicing modes between the GCAS profiles and the entire AS profile was inconsistent (Figure 1B). A larger proportion of ES events (42.6%) were detected among all AS profiles than among GCAS events (32.6%). Moreover, AP events accounted for 17.9% of events among all AS profiles and accounted for 35.1% of GCAS events. Forty percent of the splicing-parent genes possessed more than one GCAS event. However, the proportion of splicing-parent genes derived from more than two GCAS events was very low (Supplementary Figure 1C).

### Identification of OS-GCAS events

To investigate the relationship between GCAS and OS in GC patients, we performed a univariate Cox regression analysis of 1,287 GCAS events in 354 patients. As shown in Figure 1C, 173 OS-GCAS events were identified from 130 genes (Supplementary Table 2). The proportion of AS splicing modes between the OS-GCAS and GCAS profiles was also inconsistent (Supplementary Figure 1D). The AP pattern (69 cases) contained the most OS-GCAS events, followed by the ES pattern (55 cases) and AT pattern (29 cases), while the AA pattern (2 cases) contained the fewest OS-GCAS events. Although 33.3% of splicing-parent genes had two or more OS-GCAS events, almost all the genes had only one splicing pattern (Supplementary Figure 1E). Of note, among the 130 splicing-parent genes, 38 genes contained paired survival-related GCAS events. For example, the AP of IL11RA in exon site 1 (IL11RA\_86209\_AP) was a favorable prognostic indicator, whereas the AP of IL11RA in exon site 2 (IL11RA\_86208\_AP) was a poor prognostic indicator (Figure 1D). Interestingly, the splicing pattern of all the paired survival-related GCAS events was either AP or AT, and the expression of the paired survival-related

GCAS events was opposite between tumor and normal tissues (Figure 1E).

### Enrichment and interaction analysis of OS-GCAS events

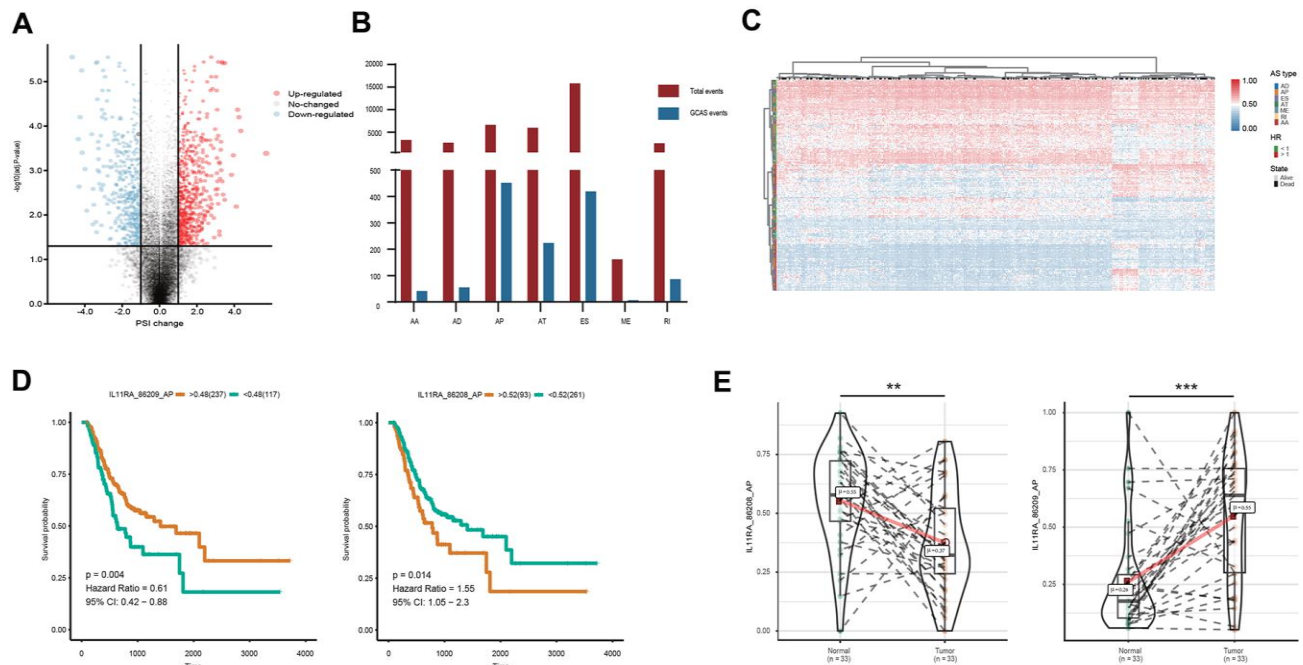
All 130 splicing-parent genes derived from 173 OS-GCAS events were assessed by functional and pathway enrichment analysis to explore the underlying mechanisms of the OS-GCAS events. The results revealed that genes were enriched in GO categories closely related to GC development, such as “microtubule polymerization”, “centrosome localization” and “cytokinesis” (Figure 2 and Supplementary Table 3). In addition, some canonical pathways involved in GC metastasis and recurrence were also enriched, such as the “Notch signaling pathway”, “adherens junction” and “bacterial invasion of epithelial cells” (Figure 2 and Supplementary Table 3).

Since AS events could inevitably affect protein characteristics, it is necessary to investigate these OS-GCAS events at the protein level [19]. A protein interaction network of OS-GCAS events was constructed not only to provide an overview of the interactions under normal conditions but also to reveal the potential influence of OS-GCAS events on the entire

network. After removing the isolated nodes, 79 genes were mapped in the protein interaction network, and these parental genes were significantly correlated with each other (Supplementary Figure 2A). The top three hub genes were SORBS1, BPTF and SEPT2, which were determined by the maximal clique centrality method. The MCODE algorithm identified three individual modules from the whole protein interaction network (Supplementary Figure 2B–2D).

### The network of OS-GCAS events and SFs

SFs are critical regulators of AS events that bind to pre-mRNAs and influence exon selection and splicing site choice [20]. It has been determined that dysregulated AS events within the tumor microenvironment may be regulated by a limited number of SFs [21]. Thus, correlation analysis was performed to explore the potential interaction networks of OS-GCAS events and SFs. As shown in Figure 3A, 103 of 173 OS-GCAS events were significantly associated with 41 experimentally validated SFs in the regulatory network ( $|R| \geq 0.4$  and  $P$ -value  $< 0.05$ ). Among the networks, most SFs were significantly correlated with more than one OS-GCAS event, while a single OS-GCAS could be regulated by up to 21 different SFs (Figure 3A). Moreover, we found that some SFs played opposite roles



**Figure 1. Profiling of AS events identified in the TCGA GC cohort.** (A) GCAS events between GC and adjacent normal tissues were visualized in a volcano plot. (B) The number of GCAS events and total AS events were depicted according to the seven AS types. (C) Heat map for the PSI values of 173 OS-GCAS events identified in 354 GC patients. (D) Kaplan-Meier curves for the paired survival-related GCAS events (IL11RA\_86208\_AP and IL11RA\_86209\_AP). (E) Violin plots for the PSI values of the paired survival-related GCAS events (IL11RA\_86208\_AP and IL11RA\_86209\_AP) between GC and adjacent normal tissues.

in the regulation of different AS events. For example, the expression of SRSF9 was negatively correlated with ATP2B4\_9450\_ES but positively correlated with APOLD1\_20517\_AT (Figure 3B).

### Construction and evaluation of the AS prognostic signature

To facilitate the application of AS events in the clinical monitoring of the prognosis of GC patients, we applied a LASSO penalty with a multivariate Cox regression analysis to the 173 OS-GCAS events in the training set of 261 patients. A total of 23 features were identified with nonzero coefficients (Figure 4A, 4B). These 23 LASSO-selected features were used to build the AS prognostic signature. The coefficients and PSI values of these 23 OS-GCAS events were used together to calculate the risk scores for both the training and validation datasets (Figure 4C).

This 23-AS event prognostic signature showed excellent performance in both datasets (Figure 4D). The C-index values of the signature were 0.725 and 0.762 in the training and validation datasets, respectively. The signature significantly stratified patients into low- and high-risk groups in the training set and validation set (Figure 4E, 4F). Significant RMS time ratios were also observed in the two datasets (Table 1). Moreover, the 23-AS event prognostic signature remained an

independent prognostic factor in multivariate analyses after adjusting for clinical and pathologic factors, such as age, TNM stage and microsatellite instability (MSI) status (Table 2).

Furthermore, we performed sensitivity analyses according to age, sex and TNM stage to evaluate the robustness of the 23-AS event prognostic signature in different clinical subgroups. The prognostic signature could stratify patients into significantly different prognostic groups in all the subgroups, which indicated that this prognostic signature might predict OS independently of clinical characteristics in GC (Supplementary Figure 3).

### Clinical, molecular and immune features underlying the 23-AS event prognostic signature

To explore the relationship between patient characteristics and the 23-AS event prognostic signatures, patients in the entire TCGA dataset were stratified into high- and low-risk groups according to the optimal cut-off of 0.046 (Figure 5A). Among the 327 patients, 137 patients were assigned to the high-risk group, and 190 patients were assigned to the low-risk group. The low-risk group had a significantly more favorable prognosis of OS than the high-risk group (Figure 5B). We found that almost all 23 LASSO-selected features were significantly differentially

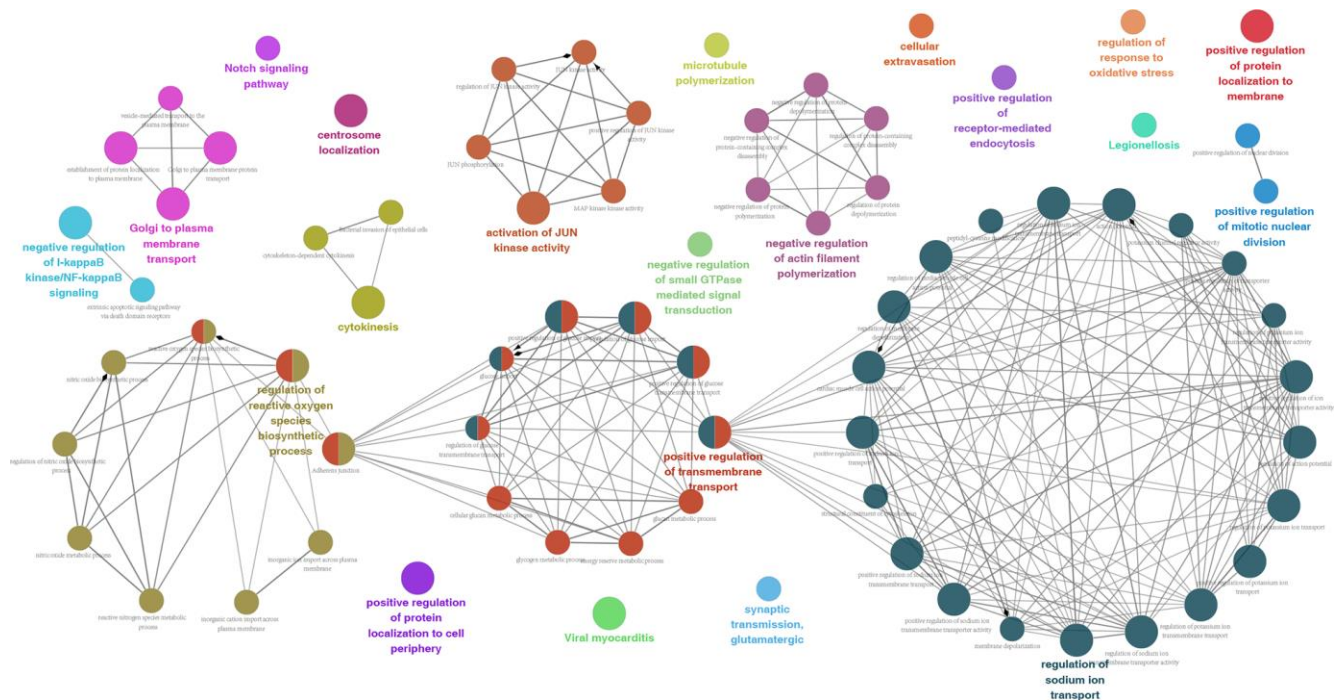
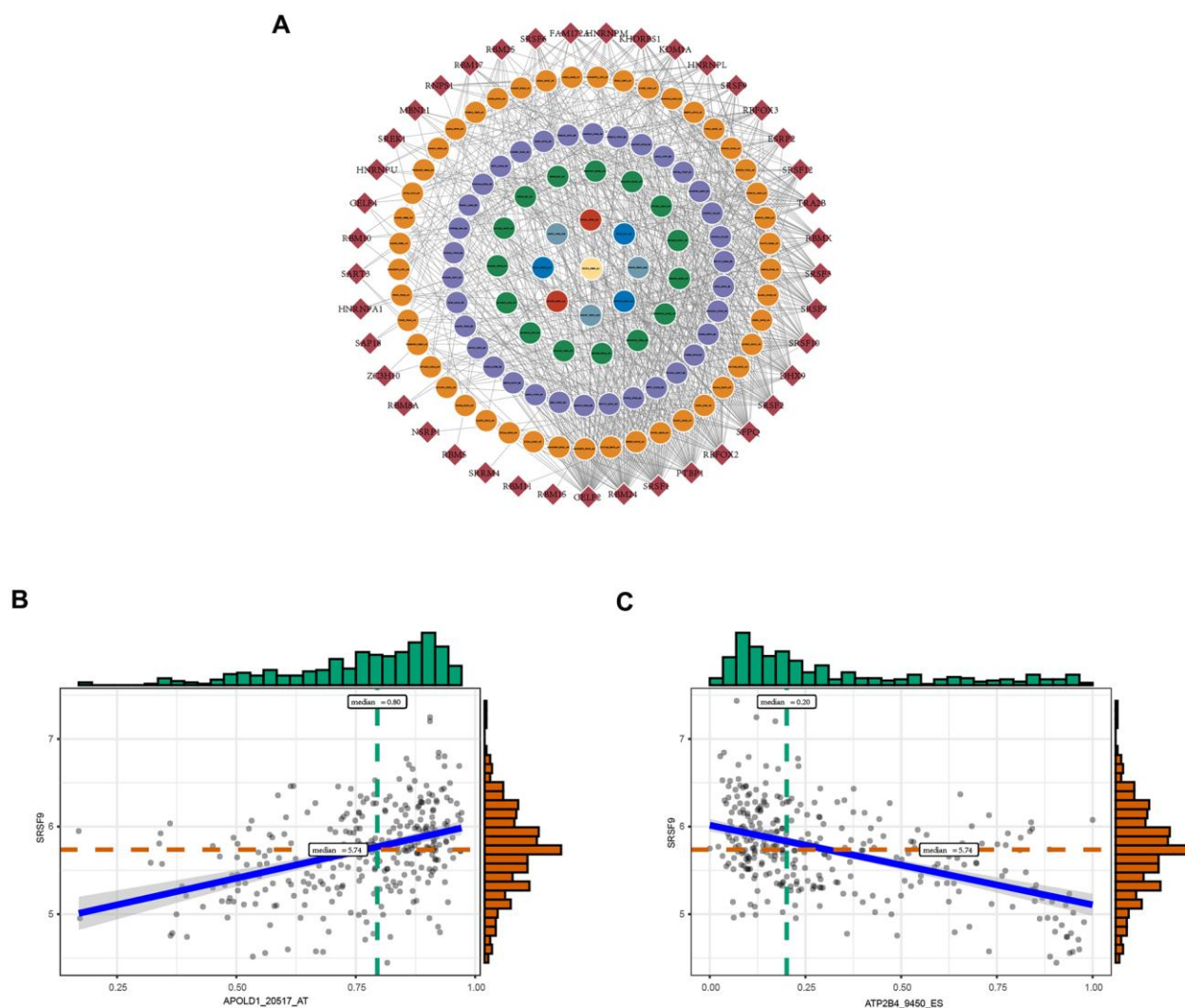


Figure 2. Potential biological functions of 173 OS-GCAS events in GC.

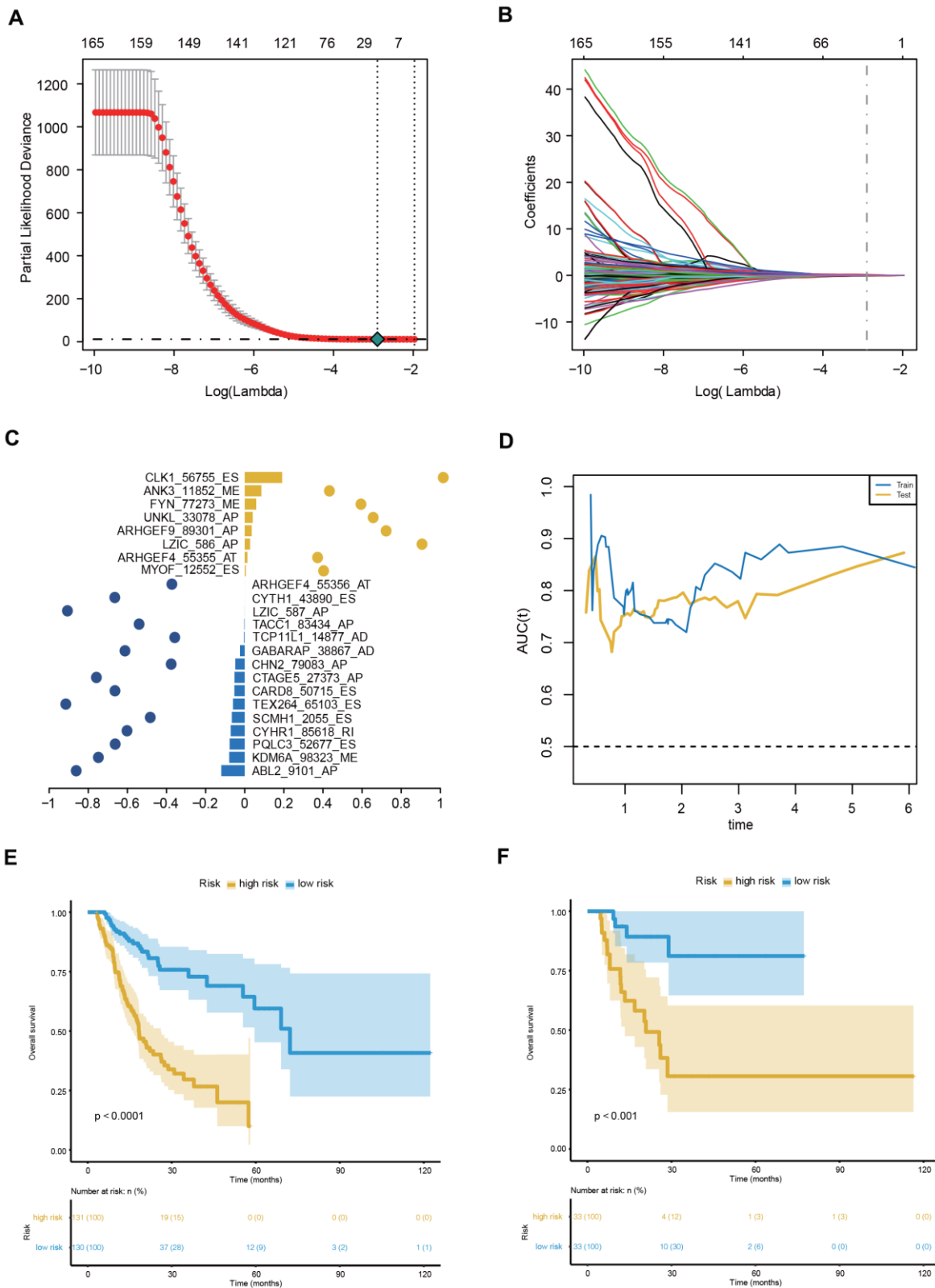
expressed between the two risk groups (Figure 5C, 5D). Moreover, the distribution of MSI status, TNM stage and molecular subtype were significantly different between the high- and low-risk groups (Figure 5C and Table 3).

We then performed GSEA to verify the difference in biological function and pathway between the high- and low-risk groups. The results revealed that some well-known pathways related to GC development, such as “apoptosis”, “hypoxia” and “epithelial-mesenchymal transition”, were significantly enriched in the high-risk group (Figure 5E). Intriguingly, immune-related pathways, such as “inflammatory response”, “interferon-alpha response” and “interferon-gamma response”, were also significantly enriched in the high-risk group (Figure 5F).

Thus, we compared the tumor immune micro-environment between the high- and low-risk groups to comprehensively characterize their different immune features. Both the immune and stromal scores were significantly higher in the high-risk group (Figure 6A). Correspondingly, a lower tumor purity was also found in the high-risk group (Figure 6A). In detail, our studies of immune cell infiltration revealed that many types of immune cells were not randomly distributed across the two groups (Figure 6B, 6C). We found a significantly higher proportion of T cells, monocytes, myeloid dendritic cells, endothelial cells and fibroblasts in the high-risk group (Figure 6B, 6C). However, the proportions of the other five types of immune cells were comparable between the two groups (Figure 6B, 6C). These results indicated the activation of stromal



**Figure 3. The regulatory splicing correlation network in GC.** (A) The correlation of OS-GCAS events with SFs is shown in network plots. The circular node represents the OS-GCAS event. The diamond node represents the SF. (B) Representative positive correlations between OS-GCAS events and SFs are shown in scatter plots. (C) Representative negative correlations between OS-GCAS events and SFs are shown in scatter plots.



**Figure 4. Construction and evaluation of the 23-AS event prognostic signature for GC patients. (A, B)** LASSO regression analysis of OS- GCAS events. **(C)** The 23 OS-GCAS events included in the signature. Corresponding coefficients from multivariate Cox regression using LASSO and log<sub>10</sub>(HRs) are depicted by horizontal bars and dots, respectively. **(D)** Time-dependent ROC curves for the 23-AS event signature in the training and validation datasets. **(E, F)** Kaplan-Meier curves with difference detection by log-rank test for patients from the training and validation datasets.

**Table 1. Restricted mean survival (RMS) time ratio between the two risk groups in different datasets.**

Dataset	N <sub>high-risk</sub>	N <sub>low-risk</sub>	RMS <sub>HRisk</sub>	RMS <sub>LRisk</sub>	RMS ratio	P-value
			(95% CI) <sup>a</sup>	(95% CI) <sup>a</sup>	(95% CI) <sup>b</sup>	
Training dataset	130	131	25.8 (21.8-29.9)	46.4 (42.5-50.3)	0.56 (0.47-0.67)	<0.001
Validation dataset	33	33	34.3 (22.1-46.5)	66.1 (55.9-76.3)	0.52 (0.35-0.77)	<0.001

<sup>a</sup>RMS time: months.

<sup>b</sup>RMS ratio = RMS<sub>HRisk</sub>/RMS<sub>LRisk</sub>.

**Table 2. Multivariate Cox proportional hazard regression in the training and validation datasets.**

	Training dataset (n = 261)						Validation dataset (n = 66)					
	Univariate regression			Multiple regression			Univariate regression			Multiple regression		
	HR	95% CI	P-value	HR	95% CI	P-value	HR	95% CI	P-value	HR	95% CI	P-value
Age	<b>1.02</b>	<b>(1.003, 1.037)</b>	<b>0.022</b>	<b>1.025</b>	<b>(1.006, 1.043)</b>	<b>0.008</b>	1.005	(0.958, 1.055)	0.825			
Gender												
Female	Reference						Reference					
Male	1.186	(0.810, 1.736)	0.38				2.39	(0.808, 7.069)	0.115			
MSI status												
MSS	Reference						Reference			Reference		
MSI-H	0.685	(0.421, 1.114)	0.127				0.709	(0.200, 2.519)	0.595	2.28	(0.553, 9.409)	0.254
MSI-L	1.145	(0.681, 1.925)	0.611				<b>4.436</b>	<b>(1.594, 12.343)</b>	<b>0.004</b>	2.786	(0.973, 7.976)	0.056
TNM stage												
Stage I	Reference			Reference			Reference			Reference		
Stage II	1.845	(0.843, 4.036)	0.125	1.543	(0.697, 3.413)	0.284	1.018	(0.170, 6.099)	0.985	1.214	(0.196, 7.500)	0.835
Stage III	<b>2.902</b>	<b>(1.389, 6.062)</b>	<b>0.005</b>	<b>2.222</b>	<b>(1.055, 4.682)</b>	<b>0.036</b>	3.927	(0.883, 17.464)	0.072	2.919	(0.621, 13.713)	0.175
Stage IV	<b>4.128</b>	<b>(1.793, 9.501)</b>	<b>0.001</b>	<b>3.737</b>	<b>(1.598, 8.739)</b>	<b>0.002</b>	<b>7.888</b>	<b>(1.312, 47.411)</b>	<b>0.024</b>	4.385	(0.612, 31.440)	0.141
Molecular subtype												
CIN	Reference						Reference					
EBV	0.718	(0.347, 1.487)	0.372				1.483	(0.319, 6.901)	0.616			
GS	0.968	(0.573, 1.635)	0.903				2.453	(0.912, 6.599)	0.075			
HM-SNV	0.974	(0.239, 3.974)	0.971				2.277	(0.280, 18.492)	0.441			
MSI	0.618	(0.374, 1.021)	0.06				0.783	(0.210, 2.927)	0.717			
Neoplasm grade												
G1	Reference									NA		
G2	0.56	(0.173, 1.809)	0.333				Reference					
G3	0.754	(0.238, 2.394)	0.632				2.583	(0.949, 7.029)	0.063			
Risk score	<b>8.182</b>	<b>(5.357, 12.496)</b>	<b>&lt;0.001</b>	<b>7.765</b>	<b>(4.834, 12.475)</b>	<b>&lt;0.001</b>	<b>19.3</b>	<b>(6.015, 61.925)</b>	<b>&lt;0.001</b>	<b>13.545</b>	<b>(3.295, 55.687)</b>	<b>&lt;0.001</b>

and immune components in the tumor immune microenvironment of high-risk patients together with the activation of oncogenic pathways based on the proposed signatures, which likely contributed at least partially to the poor prognosis of these patients.

**Integrating the 23-AS event prognostic signature with clinical characteristics**

In addition to the 23-AS event prognostic signature, age and TNM stage were also determined as independent prognostic factors in the training dataset, which suggested their complementary value (Table 2). To further improve the prognostic accuracy, we integrated the 23-AS event prognostic signature with the other two independent prognostic factors, age and TNM stage,

using the coefficients generated from the multivariate Cox regression model in the training dataset and derived an AS-clinical prognostic model as follows: Risk score = (1.948 × RSASS) + (0.348 × stage) + (0.022 × age). This integrated model was further validated in the validation dataset. Significant improvement in the assessment of survival was achieved with the AS-clinical model relative to the 23-AS event signature only (Figure 7A, 7B).

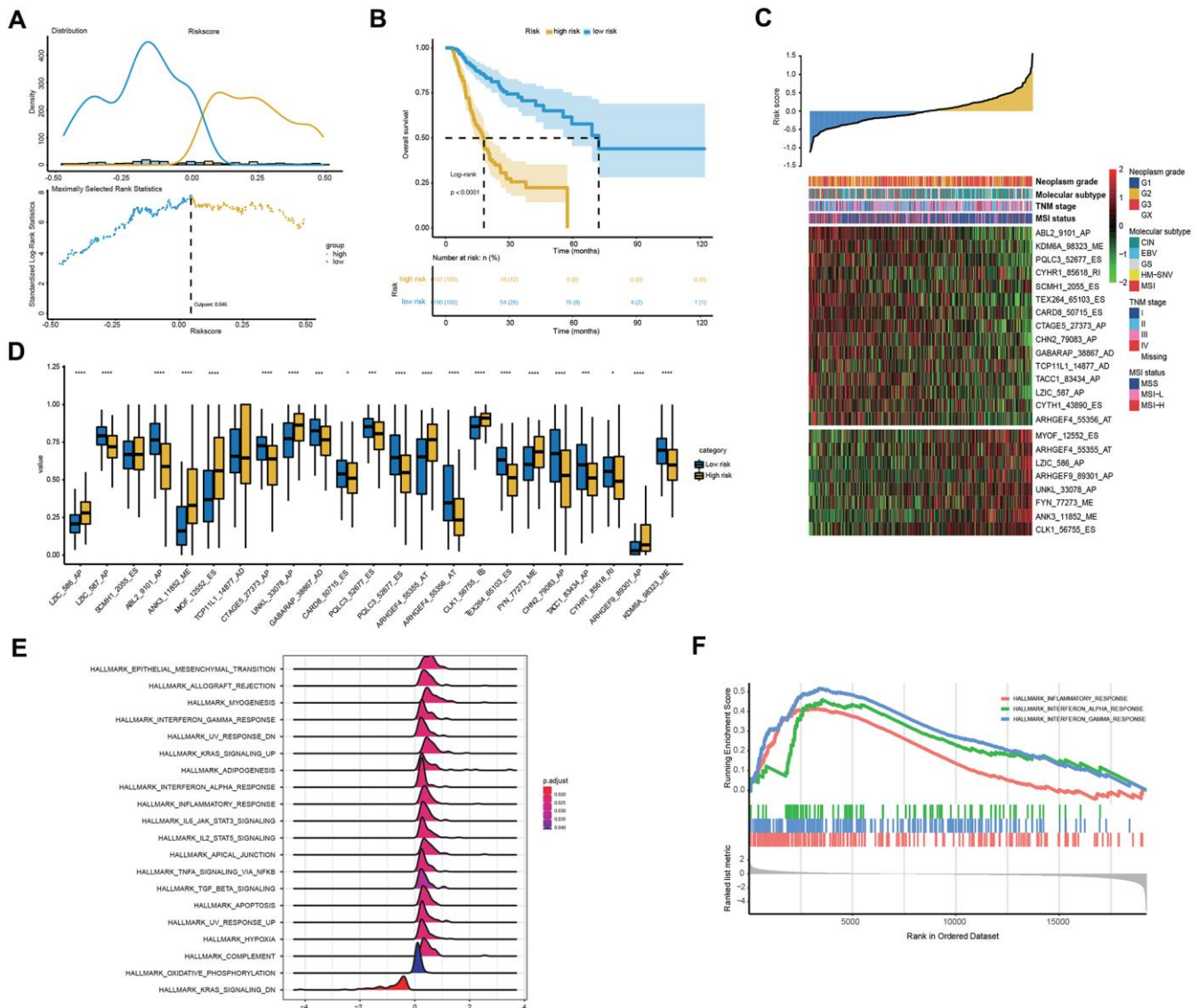
Moreover, a nomogram was established for model visualization and clinical application (Figure 7C). The ROC curve confirmed that the nomogram showed great performance in predicting the prognosis of GC (Figure 7D). The calibration curve also presented an optimal prediction for 1-year, 3-year and 5-year OS compared with the actual observations (Figure 7E). All these

findings indicated that the nomogram built in our study might contribute significantly to the prediction of prognosis for patients with GC.

## DISCUSSION

AS is a critical posttranslational modification process that generates multiple mRNA and protein isoforms with distinct structural, regulatory and functional properties [11, 12]. It has been determined that abnormal AS events contribute to numerous diseases, including several types of cancers [14, 15]. In

particular, accumulating evidence has proven that the specific dysregulation of AS events plays critical roles in GC initiation, progression and metastasis [22]. For instance, CD44, MUTYH, HOXA10 and MRPL33 splice variants participate in the carcinogenesis, proliferation, metastasis and drug resistance of GC [23–27]. However, previous studies mainly focused on monogenic isoforms with limited sample sizes, which lacked a general view of all AS events. Hence, we systematically profiled AS events in a large-scale GC cohort to characterize the role of AS events in the tumorigenesis and prognosis of GC.

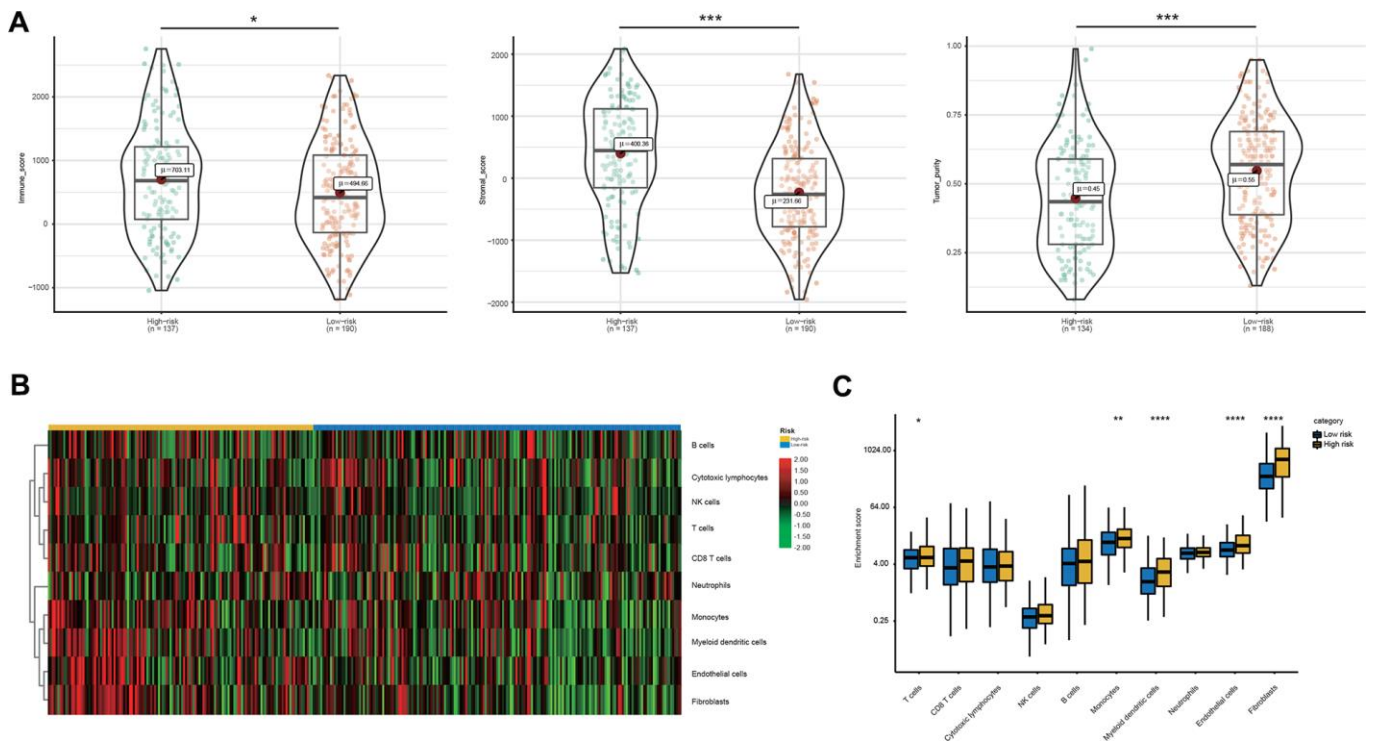


**Figure 5. Clinical and molecular features underlying the 23-AS event signature.** (A) The optimal cut-off of the 23-AS event signature in the entire TCGA dataset. (B) Kaplan-Meier curve with difference detection by log-rank test for patients in the entire TCGA dataset. (C) Heat map for the expression patterns of the 23 OS-GCAS events for the entire TCGA sample set sorted by the risk score in ascending order. (D) The differential expression of the 23 OS-GCAS events between the high- and low-risk groups. (E, F) GSEA of the 50 hallmark gene sets between the high- and low-risk groups.

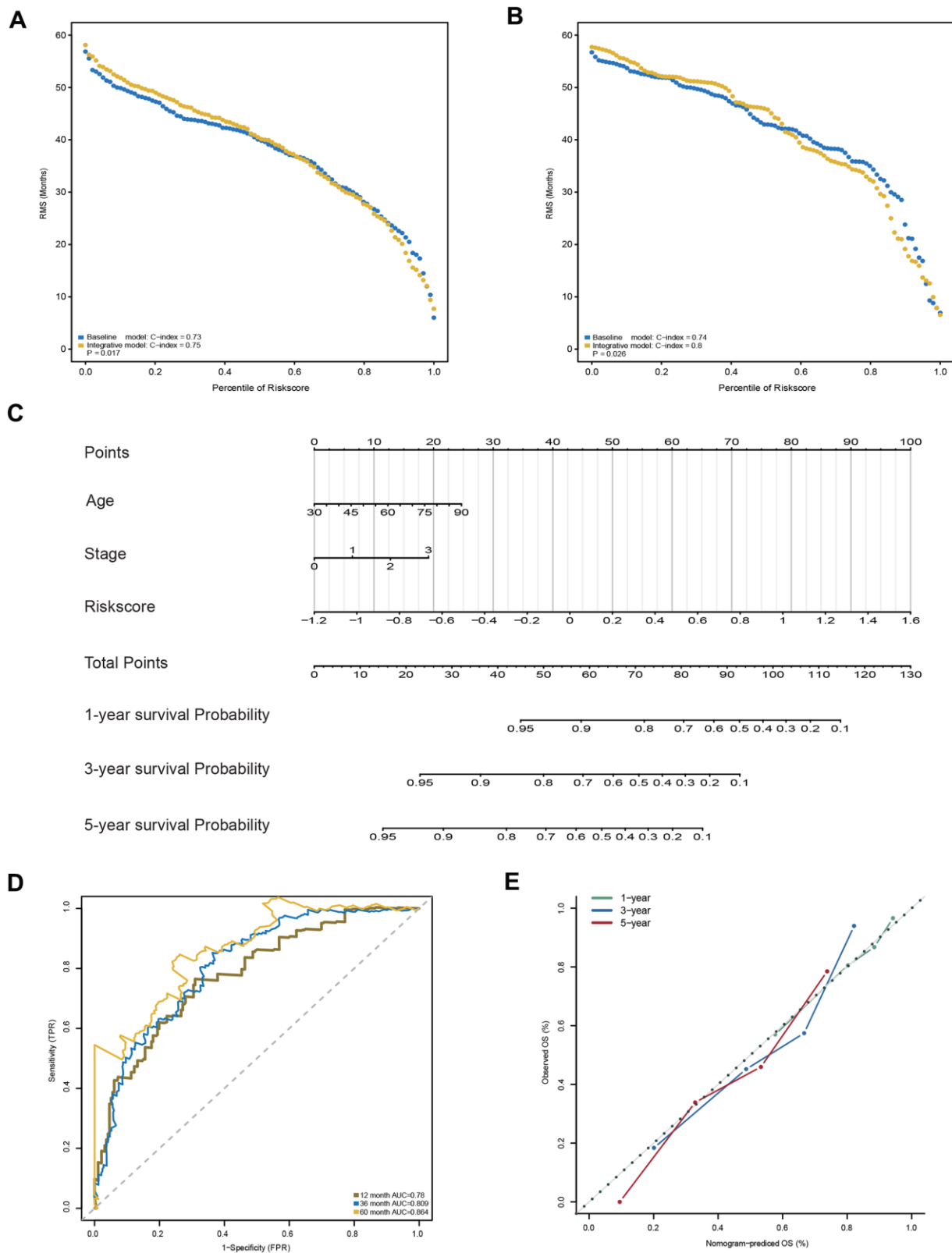


**Table 3. Differences in patient characteristics between the high- and low-risk groups in the entire TCGA dataset.**

	Total sample	High-risk group	Low-risk group	P-value
<b>MSI status</b>				<i>&lt; 0.001</i>
MSS	221	107 (48.42%)	114 (51.58%)	
MSI-H	62	11 (17.74%)	51 (82.26%)	
MSI-L	44	19 (43.18%)	25 (56.82%)	
<b>TNM stage</b>				<i>0.006</i>
Stage I	44	10 (22.73%)	34 (77.27%)	
Stage II	108	41 (37.96%)	67 (62.04%)	
Stage III	140	62 (44.29%)	78 (55.71%)	
Stage IV	29	18 (62.07%)	11 (37.93%)	
<b>Molecular subtype</b>				<i>&lt; 0.001</i>
CIN	194	92 (47.42%)	102 (52.58%)	
EBV	24	8 (33.33%)	16 (66.67%)	
GS	42	23 (54.76%)	19 (45.24%)	
HM-SNV	7	3 (42.86%)	4 (57.14%)	
MSI	60	11 (18.33%)	49 (81.67%)	
<b>Neoplasm grade</b>				<i>0.11</i>
G1	6	4 (66.67%)	2 (33.33%)	
G2	124	44 (35.48%)	80 (64.52%)	
G3	190	86 (45.26%)	104 (54.74%)	



**Figure 6. Immune microenvironment features underlying the 23-AS event signature. (A)** Violin plots for the immune scores, stromal scores, and tumor purity between the high- and low-risk groups. **(B)** Heat map for immune cell infiltration between the high- and low-risk groups. **(C)** The differential expression of immune and stromal cells between the high- and low-risk groups.



**Figure 7. Construction and evaluation of the AS-clinical nomogram for GC patients.** (A, B) RMS curves for the 23-AS event prognostic signature and the AS-clinical signature in the training and validation datasets. *P*-values represent the difference between the two signatures in terms of the C-index. (C) Nomogram prediction of 1-year, 3-year, and 5-year OS. For stage, 0 means TNM stage I, 1 means TNM stage II, 2 means TNM stage III, and 3 means TNM stage IV. (D) Time-dependent ROC curves for the nomogram at different time points. (E) Calibration curves of observed and predicted probabilities for the nomogram.

In the current study, a total of 37,139 AS events from 10,380 genes were detected after the rigorous filter, which indicated that AS is a common process in GC. By comparing the tumor and paired normal tissues, 1,287 AS events from 837 genes were detected as GCAS events. As expected, all these experimentally validated splice variants were also identified by our procedure, suggesting that the GCAS events identified in our study are ubiquitous in GC. Interestingly, we found that GC shared some common cancer-specific AS events with colorectal and head-neck squamous cell cancers, which also illustrated the role of shared AS events in cancer tumorigenesis and development [28, 29].

Among the 1,287 detected GCAS events, 173 OS-GCAS events from 130 genes were identified by univariate Cox regression analysis. In our study, a higher proportion of AP and AT events was found in the prognostic AS profile, even though ES events are the predominant components in the entire AS profile and the cancer-specific AS profile (Supplementary Figure 1A, 1D). Previous studies also confirmed that AP and AT splicing types often confer splicing isoform-specific localization and control the survival and migration of cancer cells [30, 31]. Moreover, for OS-GCAS events of AP and AT splicing types, splicing isoforms at different splice sites had distinct expression patterns and prognostic values. All the results suggested that AP and AT splicing types played important roles in the development and progression of GC.

The potential biological function and underlying molecular regulatory network of these OS-GCAS events were further explored. We found that their splicing-parent genes were significantly enriched in GC initiation and maintenance by GO term and KEGG pathway enrichment analysis. In the PPI network, we found that 79 proteins interacted closely with each other. The complexity of the network indicated that the prognosis of GC was not driven by a single AS-relevant protein; it was a process regulated by the whole integrated system.

Given the influence of SFs on the process of RNA splicing, we performed an integrated analysis of SFs and OS-GCAS events to explore the underlying mechanism of the splicing pathway involved in GC patient survival [32]. The splicing correlation network showed distinguished interactions between SFs and OS-GCAS events. Of note, we found that a given SF might play dual roles in the positive and negative regulation of different AS events and that the same OS-GCAS event could be synergistically or antagonistically regulated by different SFs, which suggested that multiple SFs could affect the survival of GC patients by synergistically regulating the AS events of genes. The relationships

between AS events and SFs might be suitable to consider as a dynamic interaction network instead of a simple “one-to-one” pattern.

To explore the prognostic significance of AS events, we constructed a robust signature for GC. The 23-AS event prognostic signature showed excellent performance in both the training and validation datasets. The 23-AS event prognostic signature precisely stratified risk. Patients in the high-risk group had a poorer prognosis than those in the low-risk group. Subgroup analysis also indicated that the prognostic signature was stable in different situations. All the results suggest that the 23-AS event prognostic signature could provide patients with new predictive information and facilitate patient assessment.

Of note, we found that the distribution of MSI status, TNM stage and molecular subtype was significantly different between the high- and low-risk groups, which indicated different biological functions between the high- and low-risk groups. Consistent with the results, GSEA revealed that several GC-associated pathways were significantly enriched in the high-risk group. Intriguingly, immune-related pathways were also significantly enriched, which suggested an enhanced immune phenotype in the high-risk group.

Correspondingly, we further estimated the population abundance of tissue-infiltrating immune and stromal cell populations to characterize the tumor immune microenvironment of the high- and low-risk groups. Generally, significantly higher immune and stromal scores were found in the high-risk group, which was characterized by a low tumor purity. Consistent with our study, previous studies have determined that low tumor purity is associated with poor prognosis and an enhanced immune phenotype [33–35]. Specifically, significantly higher proportions of T cells, monocytes, myeloid dendritic cells, endothelial cells and fibroblasts were identified in the high-risk group. The higher proportion of T cells in the high-risk group might be an essential factor for the immunotherapy response [36].

To further improve the prognostic accuracy, we integrated the 23-AS event prognostic signature with clinicopathologic factors. The prediction performance of the integrative AS-clinical prognostic model was superior to that of the single 23-AS event prognostic signature. Finally, a nomogram was established for model visualization and clinical application. The nomogram also showed great performance in predicting the prognosis of GC. Thus, we propose that the AS-clinical nomogram could serve as an individualized, single-sample estimate of survival of GC and might be readily translated to clinical practice.

In summary, our study comprehensively investigated the prognostic value of genome-wide AS events in GC and explored their potential biological function and the underlying splicing pathways of these 173 OS-GCAS events. More importantly, we constructed a 23-AS event prognostic signature to classify the risk of GC patients and identified differential clinical, molecular and immune features underlying the 23-AS event prognostic signature. In addition, combining data concerning age, TNM stage and the 23-AS event prognostic signature, we constructed an AS-clinical nomogram to predict the survival of GC patients. These results provide fundamental information for understanding the roles of the AS process and indicate the potential clinical implications of AS events in GC.

## MATERIALS AND METHODS

### Data acquisition and curation process

Patients who met the following criteria were included in The Cancer Genome Atlas (TCGA) GC cohort: (1) histologically confirmed primary GC; (2) alternative RNA splicing data available; (3) detailed clinicopathological and follow-up information available, including sex, age, TNM stage, neoplasm grade, microsatellite status, molecular subtype, and OS status; and (4) an OS time of over 90 days; this latter criterion was applied to, avoid immortal time bias. The alternative RNA splicing data of the TCGA GC cohort were downloaded from the TCGASpliceSeq dataset [37]. Corresponding clinical information and RNA-seq data were downloaded from the Genomic Data Commons data portal using the "GDCRNATools" package [38].

Splicing events in the TCGASpliceSeq dataset were divided into seven categories: exon skip (ES), retained intron (RI), alternate promoter (AP), alternate terminator (AT), alternate donor site (AD), alternate acceptor site (AA), and mutually exclusive exons (ME). Each splicing event was quantified by the percent spliced in (PSI) value, which ranges from 0 to 1 and represents the ratio of normalized read counts to indicate the inclusion of a transcript element over the total normalized reads for that event [39]. To generate a reliable set of AS events, we implemented a series of stringent filters, which included "percentage of samples with a PSI value  $\geq$  75%" and "average PSI value  $\geq$  0.05" [28, 29]. The missing PSI values were further filled in using the k nearest neighbors (KNN) algorithm [40].

### Identification of GCAS events and OS-GCAS events

To identify GCAS events, we applied the Wilcoxon matched-pairs signed-rank test to compare the PSI values between tumor tissues and matched adjacent

normal tissues. The *P*-value was adjusted by the Benjamini-Hochberg (BH) method. GCAS events were defined as a median PSI value that varied more than 0.1 between tumor tissues and matched adjacent normal tissues with a BH-adjusted *P*-value  $<$  0.05 [30, 41].

To determine the survival-associated GCAS events, we performed univariate Cox proportional hazards regression analysis to estimate the PSI values of GCAS events with OS. GCAS events with a *P*-value  $<$  0.05 in the univariate Cox regression analysis were selected as OS-GCAS events.

### Functional enrichment analysis and protein interaction network

To investigate the potential functions of OS-GCAS events, we subjected their parent genes to enrichment analyses of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the ClueGO plug-in of Cytoscape [42]. For the functional enrichment analysis, a BH-adjusted *P*-value  $<$  0.05 was considered statistically significant.

To observe the PPIs among the OS-GCAS events, we mapped their corresponding genes to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [43]. A minimum required interaction score of 0.4 was used to identify the protein interaction results in STRING, which were further visualized by Cytoscape [44]. In addition, hub genes and specific modules of the protein interaction network were identified by the CytoHubba and Molecular Complex Detection (MCODE) plug-ins of Cytoscape [45, 46]. The maximal clique centrality (MCC) algorithm of CytoHubba was used to predict the top three hub genes in the PPI network. The MCODE options were set as degree cutoff = 2, K-core = 2, and node score cutoff = 0.

### Splicing factors (SFs) and the splicing correlation network

A total of 78 genes that participated in the process of alternative RNA splicing (GO:0000380) were obtained from the Molecular Signatures Database (MSigDB) [47]. The read count values of these SFs were extracted from the RNA-seq data, normalized by the trimmed means of M (TMM) method and further transformed by "voom" [48, 49]. Differentially expressed SFs were identified through the "limma" package, and a BH-adjusted *P*-value  $<$  0.05 was used as the significance threshold [50]. Spearman's correlation analysis was conducted to explore the underlying correlation between the expression of the SFs and the PSI values of the OS-GCAS events. The absolute value of the correlation

coefficient  $\geq 0.4$  with a  $P$ -value  $< 0.05$  was considered statistically significant. The splicing correlation network of SFs and the OS-GCAS events was visualized by Cytoscape [44].

### Identification of the AS prognostic signature

To enhance the robustness of the AS prognostic signature, the entire TCGA cohort of 327 samples was first randomly distributed into two datasets (7:3), namely, the training and validation datasets. Next, to avoid overfitting in the multivariate model, the least absolute shrinkage and selection operator (LASSO) penalty was applied in the training dataset to build an optimal prognostic signature with the minimum number of OS-GCAS events. Ten-fold cross-validation was conducted to determine the optimal value of the penalty parameter  $\lambda$ , which gives the minimum partial likelihood deviance. Finally, a set of OS-GCAS events and their nonzero coefficients were identified, which was used to build the AS prognostic signature. The time-dependent receiver operator characteristic (ROC) curve and concordance index (C-index) were applied to assess the performance of the prognosis model.

The risk score for the AS signature was calculated for each sample via a linear combination of the selected features, weighted by the corresponding coefficients. Patients were divided into high- and low-risk groups using the cohort-specific median risk score as the cut-off for each dataset. Kaplan-Meier survival analysis was performed to compare patient prognosis between the high- and low-risk groups. Furthermore, the restricted mean survival (RMS) time was also used to compare the prognostic differences between the two groups [51]. The higher the RMS value was, the greater the prognostic difference. A  $P$ -value  $< 0.05$  was considered statistically significant.

### Association between different risk groups and clinical, molecular, and immune features

Patients in the entire TCGA cohort were divided into high-risk and low-risk groups according to the optimal cut-off value, which was determined by the maximally selected rank statistics test. For the difference between the two groups, clinical features (age, sex, TNM stage, neoplasm grade, and survival status), molecular features (microsatellite status and molecular subtype), and immune features (immune score, stromal score, tumor purity, and immune cell infiltration) were analyzed.

Tumor purity was measured by “ABSOLUTE” [52]; immune and stromal scores were calculated by “ESTIMATE” [53]; and immune cell infiltration was

estimated with “MCPcounter” [54]. Kaplan-Meier curve with log-rank test was used for survival data; Mann-Whitney test was used for continuous data; chi-square test was used for categorical data. A  $P$ -value  $< 0.05$  was considered statistically significant. In addition, to further explore the difference in biological functions and pathways between high-risk and low-risk groups, gene set enrichment analysis (GSEA) was performed using the “hallmark gene sets” downloaded from MSigDB [47, 55]. A BH-adjusted  $P$ -value  $< 0.05$  was considered statistically significant.

### Development and verification of an AS-clinical nomogram

Based on the results derived from multivariate analyses, we integrated age, TNM stage, and AS prognostic signature to generate an AS-clinical prognostic model by applying Cox proportional hazard regression in the training dataset. The AS-clinical prognostic model was then applied to the validation dataset for further validation. Next, the prognostic value of the AS-clinical model was compared with that of the AS signature in terms of C-index, and the results were revealed by the RMS curve [18]. RMS represents the life expectancy at five years for patients with different risk scores. Finally, based on the AS-clinical prognostic model, a nomogram was developed to estimate the individual survival probability of patients. The ROC and calibration curves were used to assess the discrimination and calibration ability of the AS-clinical nomogram.

### AUTHOR CONTRIBUTIONS

Conception and design: Yingwei Xue and Shenghan Lou. Data acquisition and assembly: Shenghan Lou and Jian Zhang. Data analyses and interpretation: Shenghan Lou, Jian Zhang, Zhao Zhai, Xin Yin, Yimin Wang and Tianyi Fang. Financial support: Yingwei Xue. All authors contributed to the manuscript. All authors read and approved the final manuscript.

### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

### FUNDING

This work was supported by funding from Project No. 10 of Harbin Medical University Cancer Hospital (Grant Number 102017–03) and the fundamental research funds for the provincial universities.

## REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394–424.  
<https://doi.org/10.3322/caac.21492> PMID:30207593
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87–108.  
<https://doi.org/10.3322/caac.21262> PMID:25651787
3. Gomceli I, Demiriz B, Tez M. Gastric carcinogenesis. *World J Gastroenterol.* 2012; 18:5164–70.  
<https://doi.org/10.3748/wjg.v18.i37.5164> PMID:23066309
4. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin.* 2013; 63:11–30.  
<https://doi.org/10.3322/caac.21166> PMID:23335087
5. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. *Lancet.* 2016; 388:2654–64.  
[https://doi.org/10.1016/S0140-6736\(16\)30354-3](https://doi.org/10.1016/S0140-6736(16)30354-3) PMID:27156933
6. Dikken JL, van de Velde CJ, Coit DG, Shah MA, Verheij M, Cats A. Treatment of resectable gastric cancer. *Therap Adv Gastroenterol.* 2012; 5:49–69.  
<https://doi.org/10.1177/1756283X11410771> PMID:22282708
7. Shi H, Jiang Y, Cao H, Zhu H, Chen B, Ji W. Nomogram based on systemic immune-inflammation index to predict overall survival in gastric cancer patients. *Dis Markers.* 2018; 2018:1787424.  
<https://doi.org/10.1155/2018/1787424> PMID:30627220
8. Hu L, Li HL, Li WF, Chen JM, Yang JT, Gu JJ, Xin L. Clinical significance of expression of proliferating cell nuclear antigen and E-cadherin in gastric carcinoma. *World J Gastroenterol.* 2017; 23:3721–29.  
<https://doi.org/10.3748/wjg.v23.i20.3721> PMID:28611525
9. Zhu X, Tian X, Yu C, Shen C, Yan T, Hong J, Wang Z, Fang JY, Chen H. A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Mol Cancer.* 2016; 15:60.  
<https://doi.org/10.1186/s12943-016-0544-0> PMID:27647437
10. Feng H, Qin Z, Zhang X. Opportunities and methods for studying alternative splicing in cancer with RNA-Seq. *Cancer Lett.* 2013; 340:179–91.  
<https://doi.org/10.1016/j.canlet.2012.11.010> PMID:23196057
11. Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol.* 2014; 15:108–21.  
<https://doi.org/10.1038/nrm3742> PMID:24452469
12. Lee SC, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. *Nat Med.* 2016; 22:976–86.  
<https://doi.org/10.1038/nm.4165> PMID:27603132
13. Blencowe BJ. The relationship between alternative splicing and proteomic complexity. *Trends Biochem Sci.* 2017; 42:407–08.  
<https://doi.org/10.1016/j.tibs.2017.04.001> PMID:28483376
14. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene.* 2014; 33:5311–18.  
<https://doi.org/10.1038/onc.2013.533> PMID:24336324
15. Climente-González H, Porta-Pardo E, Godzik A, Eyra E. The functional impact of alternative splicing in cancer. *Cell Rep.* 2017; 20:2215–26.  
<https://doi.org/10.1016/j.celrep.2017.08.012> PMID:28854369
16. 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.  
<https://doi.org/10.1016/j.semcdb.2017.09.018> PMID:28919308
17. Martinez-Montiel N, Rosas-Murrieta NH, Anaya Ruiz M, Monjaraz-Guzman E, Martinez-Contreras R. Alternative splicing as a target for cancer treatment. *Int J Mol Sci.* 2018; 19:545.  
<https://doi.org/10.3390/ijms19020545> PMID:29439487
18. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, Tress ML. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet.* 2014; 23:5866–78.  
<https://doi.org/10.1093/hmg/ddu309> PMID:24939910
19. Neverov AD, Artamonova II, Nurtdinov RN, Frishman D, Gelfand MS, Mironov AA. Alternative splicing and protein function. *BMC Bioinformatics.* 2005; 6:266.  
<https://doi.org/10.1186/1471-2105-6-266> PMID:16274476
20. 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; 35:2413–27.  
<https://doi.org/10.1038/onc.2015.318> PMID:26300000
21. Brosseau JP, Lucier JF, Nwilati H, Thibault P, Garneau D, Gendron D, Durand M, Couture S, Lapointe E, Prinos P, Klinck R, Perreault JP, Chabot B, Abou-Elela S. Tumor

- microenvironment-associated modifications of alternative splicing. *RNA*. 2014; 20:189–201.  
<https://doi.org/10.1261/rna.042168.113>  
PMID:[24335142](https://pubmed.ncbi.nlm.nih.gov/24335142/)
22. Li Y, Yuan Y. Alternative RNA splicing and gastric cancer. *Mutat Res*. 2017; 773:263–73.  
<https://doi.org/10.1016/j.mrrev.2016.07.011>  
PMID:[28927534](https://pubmed.ncbi.nlm.nih.gov/28927534/)
  23. Chen XY, Wang ZC, Li H, Cheng XX, Sun Y, Wang XW, Wu ML, Liu J. Nuclear translocations of beta-catenin and TCF4 in gastric cancers correlate with lymph node metastasis but probably not with CD44 expression. *Hum Pathol*. 2005; 36:1294–301.  
<https://doi.org/10.1016/j.humpath.2005.09.003>  
PMID:[16311123](https://pubmed.ncbi.nlm.nih.gov/16311123/)
  24. Kobayashi K, Shida A, Yamada H, Ishibashi Y, Nakayama R, Toriumi Y, Mitsumori N, Kashiwagi H, Yanaga K. Frequent splicing aberration of the base excision repair gene hMYH in human gastric cancer. *Anticancer Res*. 2008; 28:215–21.  
PMID:[18383848](https://pubmed.ncbi.nlm.nih.gov/18383848/)
  25. Zhang Y, Yuan Z, Jiang Y, Shen R, Gu M, Xu W, Gu X. Inhibition of splicing factor 3b subunit 1 (SF3B1) reduced cell proliferation, induced apoptosis and resulted in cell cycle arrest by regulating homeobox A10 (HOXA10) splicing in AGS and MKN28 human gastric cancer cells. *Med Sci Monit*. 2020; 26:e919460.  
<https://doi.org/10.12659/MSM.919460>  
PMID:[31927557](https://pubmed.ncbi.nlm.nih.gov/31927557/)
  26. Li J, Feng D, Gao C, Zhang Y, Xu J, Wu M, Zhan X. Isoforms S and L of MRPL33 from alternative splicing have isoform-specific roles in the chemoresponse to epirubicin in gastric cancer cells via the PI3K/AKT signaling pathway. *Int J Oncol*. 2019; 54:1591–600.  
<https://doi.org/10.3892/ijo.2019.4728> PMID:[30816492](https://pubmed.ncbi.nlm.nih.gov/30816492/)
  27. Peng WZ, Liu JX, Li CF, Ma R, Jie JZ. hnRNP promotes gastric tumorigenesis through regulating CD44E alternative splicing. *Cancer Cell Int*. 2019; 19:335.  
<https://doi.org/10.1186/s12935-019-1020-x>  
PMID:[31857793](https://pubmed.ncbi.nlm.nih.gov/31857793/)
  28. Li ZX, Zheng ZQ, Wei ZH, Zhang LL, Li F, Lin L, Liu RQ, Huang XD, Lv JW, Chen FP, He XJ, Guan JL, Kou J, et al. Comprehensive characterization of the alternative splicing landscape in head and neck squamous cell carcinoma reveals novel events associated with tumorigenesis and the immune microenvironment. *Theranostics*. 2019; 9:7648–65.  
<https://doi.org/10.7150/thno.36585> PMID:[31695792](https://pubmed.ncbi.nlm.nih.gov/31695792/)
  29. 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; 36:183–95.  
<https://doi.org/10.1016/j.ebiom.2018.09.021>  
PMID:[30243491](https://pubmed.ncbi.nlm.nih.gov/30243491/)
  30. Zhang Y, Yan L, Zeng J, Zhou H, Liu H, Yu G, Yao W, Chen K, Ye Z, Xu H. Pan-cancer analysis of clinical relevance of alternative splicing events in 31 human cancers. *Oncogene*. 2019; 38:6678–95.  
<https://doi.org/10.1038/s41388-019-0910-7>  
PMID:[31391553](https://pubmed.ncbi.nlm.nih.gov/31391553/)
  31. Taliaferro JM, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, Wang ET, Graveley BR, Gertler FB, Swanson MS, Burge CB. Distal alternative last exons localize mRNAs to neural projections. *Mol Cell*. 2016; 61:821–33.  
<https://doi.org/10.1016/j.molcel.2016.01.020>  
PMID:[26907613](https://pubmed.ncbi.nlm.nih.gov/26907613/)
  32. Zhu LY, Zhu YR, Dai DJ, Wang X, Jin HC. Epigenetic regulation of alternative splicing. *Am J Cancer Res*. 2018; 8:2346–58.  
PMID:[30662796](https://pubmed.ncbi.nlm.nih.gov/30662796/)
  33. Zhang C, Cheng W, Ren X, Wang Z, Liu X, Li G, Han S, Jiang T, Wu A. Tumor purity as an underlying key factor in glioma. *Clin Cancer Res*. 2017; 23:6279–91.  
<https://doi.org/10.1158/1078-0432.CCR-16-2598>  
PMID:[28754819](https://pubmed.ncbi.nlm.nih.gov/28754819/)
  34. Mao Y, Feng Q, Zheng P, Yang L, Liu T, Xu Y, Zhu D, Chang W, Ji M, Ren L, Wei Y, He G, Xu J. Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. *Cancer Manag Res*. 2018; 10:3569–77.  
<https://doi.org/10.2147/CMAR.S171855>  
PMID:[30271205](https://pubmed.ncbi.nlm.nih.gov/30271205/)
  35. Rhee JK, Jung YC, Kim KR, Yoo J, Kim J, Lee YJ, Ko YH, Lee HH, Cho BC, Kim TM. Impact of tumor purity on immune gene expression and clustering analyses across multiple cancer types. *Cancer Immunol Res*. 2018; 6:87–97.  
<https://doi.org/10.1158/2326-6066.CIR-17-0201>  
PMID:[29141981](https://pubmed.ncbi.nlm.nih.gov/29141981/)
  36. Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol*. 2020; 20:651–68.  
<https://doi.org/10.1038/s41577-020-0306-5>  
PMID:[32433532](https://pubmed.ncbi.nlm.nih.gov/32433532/)
  37. Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, Melott J, Weinstein J. TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res*. 2016; 44:D1018–22.  
<https://doi.org/10.1093/nar/gkv1288> PMID:[26602693](https://pubmed.ncbi.nlm.nih.gov/26602693/)
  38. Li R, Qu H, Wang S, Wei J, Zhang L, Ma R, Lu J, Zhu J, Zhong WD, Jia Z. GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. *Bioinformatics*. 2018; 34:2515–17.

- <https://doi.org/10.1093/bioinformatics/bty124>  
PMID:[29509844](https://pubmed.ncbi.nlm.nih.gov/29509844/)
39. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456:470–76.  
<https://doi.org/10.1038/nature07509>  
PMID:[18978772](https://pubmed.ncbi.nlm.nih.gov/18978772/)
40. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak*. 2016 (Suppl 3); 16:74.  
<https://doi.org/10.1186/s12911-016-0318-z>  
PMID:[27454392](https://pubmed.ncbi.nlm.nih.gov/27454392/)
41. Sebestyén E, Singh B, Miñana B, Pagès A, Mateo F, Pujana MA, Valcárcel J, Eyras E. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res*. 2016; 26:732–44.  
<https://doi.org/10.1101/gr.199935.115>  
PMID:[27197215](https://pubmed.ncbi.nlm.nih.gov/27197215/)
42. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009; 25:1091–93.  
<https://doi.org/10.1093/bioinformatics/btp101>  
PMID:[19237447](https://pubmed.ncbi.nlm.nih.gov/19237447/)
43. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019; 47:D607–13.  
<https://doi.org/10.1093/nar/gky1131> PMID:[30476243](https://pubmed.ncbi.nlm.nih.gov/30476243/)
44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–504.  
<https://doi.org/10.1101/gr.1239303>  
PMID:[14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
45. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014 (Suppl 4); 8:S11.  
<https://doi.org/10.1186/1752-0509-8-S4-S11>  
PMID:[25521941](https://pubmed.ncbi.nlm.nih.gov/25521941/)
46. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003; 4:2.  
<https://doi.org/10.1186/1471-2105-4-2>  
PMID:[12525261](https://pubmed.ncbi.nlm.nih.gov/12525261/)
47. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015; 1:417–25.  
<https://doi.org/10.1016/j.cels.2015.12.004>  
PMID:[26771021](https://pubmed.ncbi.nlm.nih.gov/26771021/)
48. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26:139–40.  
<https://doi.org/10.1093/bioinformatics/btp616>  
PMID:[19910308](https://pubmed.ncbi.nlm.nih.gov/19910308/)
49. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014; 15:R29.  
<https://doi.org/10.1186/gb-2014-15-2-r29>  
PMID:[24485249](https://pubmed.ncbi.nlm.nih.gov/24485249/)
50. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015; 43:e47.  
<https://doi.org/10.1093/nar/gkv007> PMID:[25605792](https://pubmed.ncbi.nlm.nih.gov/25605792/)
51. Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, Schrag D, Takeuchi M, Uyama Y, Zhao L, Skali H, Solomon S, Jacobus S, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol*. 2014; 32:2380–85.  
<https://doi.org/10.1200/JCO.2014.55.2208>  
PMID:[24982461](https://pubmed.ncbi.nlm.nih.gov/24982461/)
52. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhi R, Pellman D, Levine DA, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012; 30:413–21.  
<https://doi.org/10.1038/nbt.2203> PMID:[22544022](https://pubmed.ncbi.nlm.nih.gov/22544022/)
53. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013; 4:2612.  
<https://doi.org/10.1038/ncomms3612> PMID:[24113773](https://pubmed.ncbi.nlm.nih.gov/24113773/)
54. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016; 17:218.  
<https://doi.org/10.1186/s13059-016-1070-5>  
PMID:[27765066](https://pubmed.ncbi.nlm.nih.gov/27765066/)



55. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting

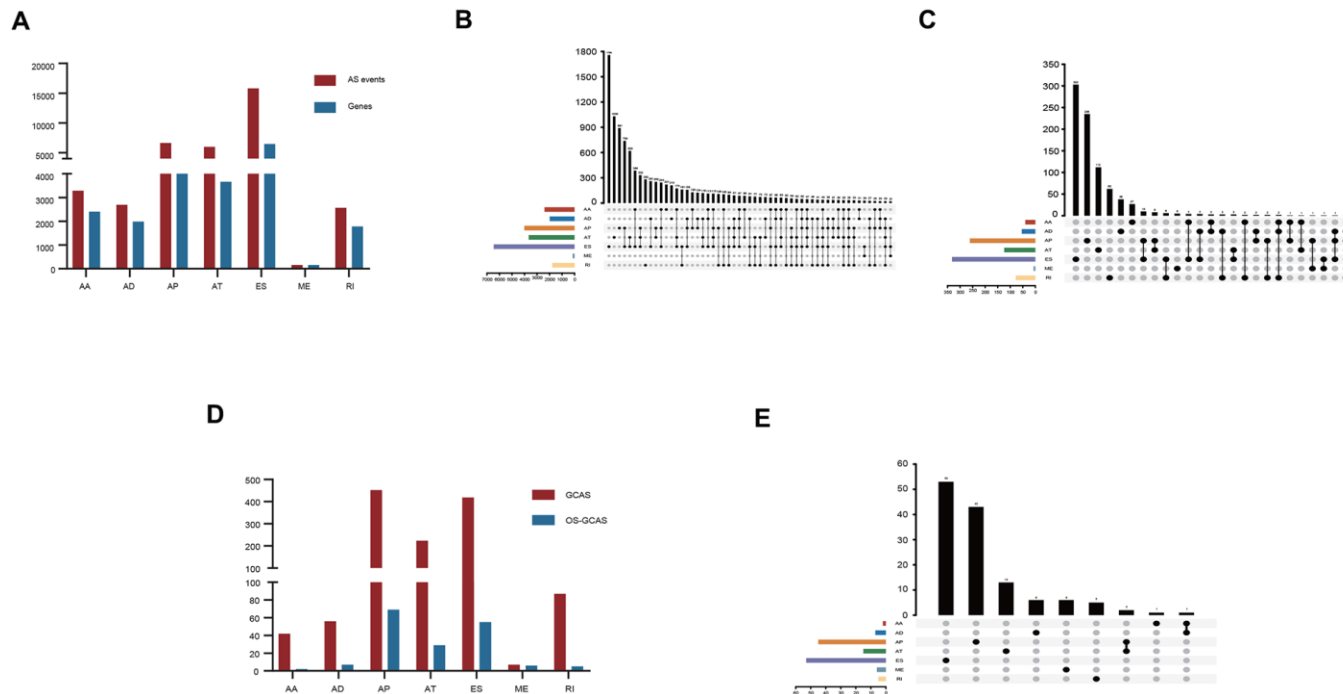
genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–50.

<https://doi.org/10.1073/pnas.0506580102>

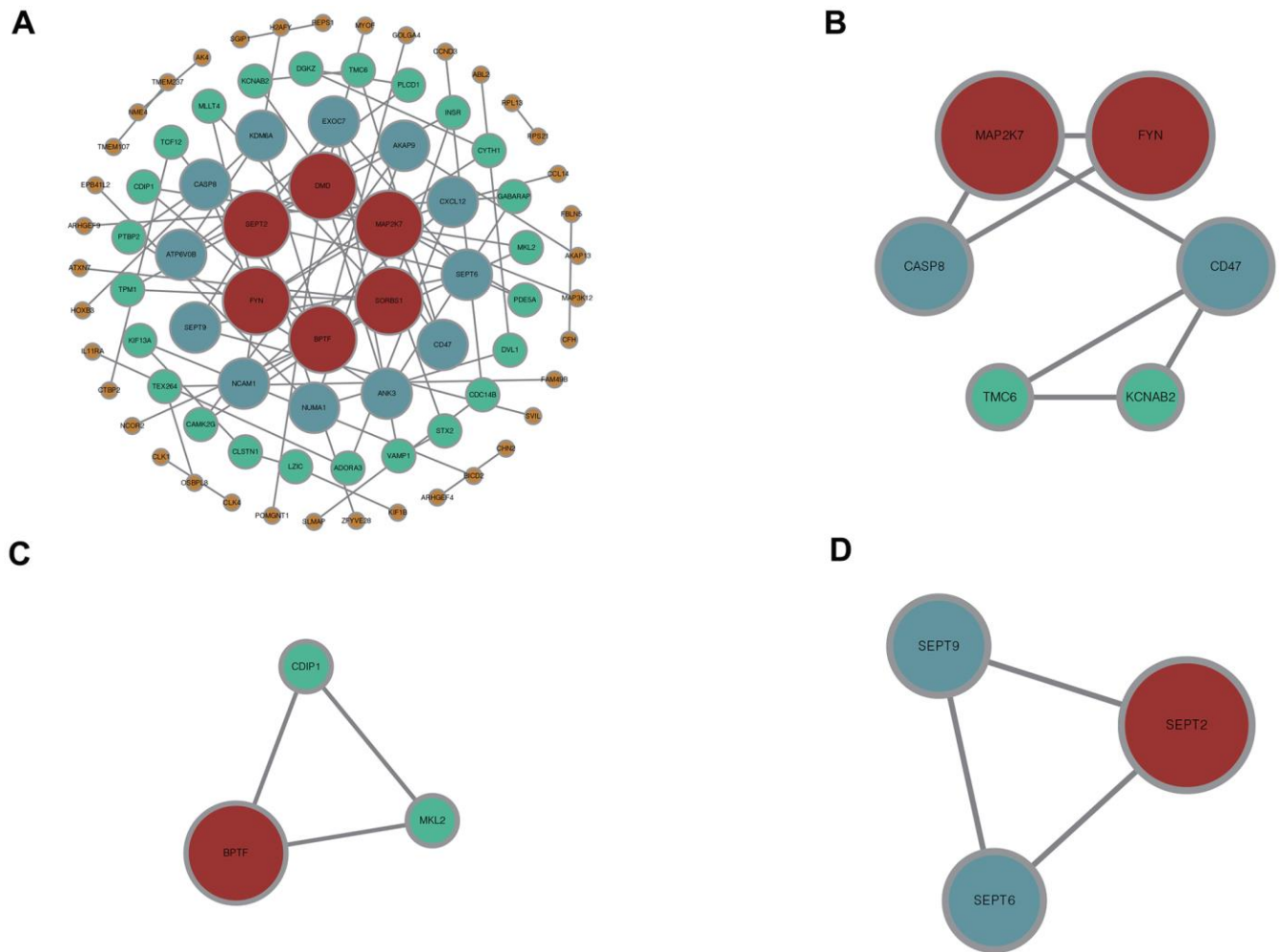
PMID:[16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)

# SUPPLEMENTARY MATERIALS

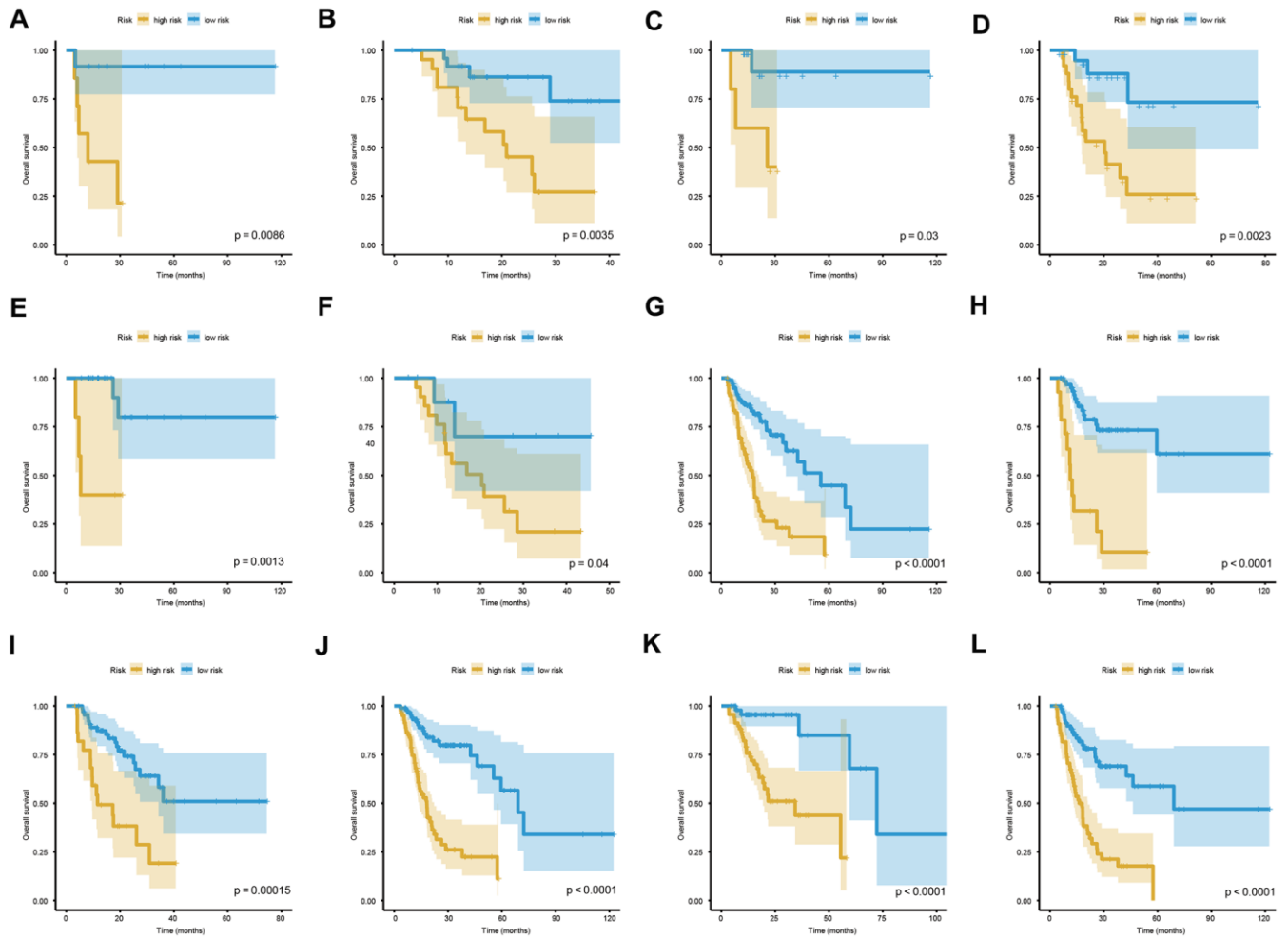
## Supplementary Figures



**Supplementary Figure 1. Overview of AS event profiling in the TCGA cohort.** (A) The numbers of AS events and corresponding genes are depicted according to the seven AS types. (B) UpSet plot for the intersection of the seven types of detected AS events. (C) UpSet plot for the intersection of the seven types of GCAS events. (D) The numbers of GCAS events and OS-GCAS events are depicted according to the seven AS types. (E) UpSet plot for the intersection of the seven types of OS-GCAS events.



**Supplementary Figure 2. PPI analysis of the 173 OS-GCAS events.** (A) The PPI network for the interactome of 130 parental genes for 173 OS-GCAS events. (B–D) Three individual modules from the whole PPI network.



**Supplementary Figure 3. Estimation of the prognostic value of the 23-AS event signature in diverse subgroups classified by different clinical characteristics.** The dataset was classified into diverse subgroups according to age (< 60 years or ≥ 60 years), sex (female or male), and TNM stage (I + II or III + IV). (A–F) shows the validation dataset. (G–L) shows the training dataset.

## **Supplementary Tables**

Please browse Full Text version to see the data of Supplementary Tables 1–3.

**Supplementary Table 1. Identification of GCAS events.**

**Supplementary Table 2. Identification of OS-GCAS events.**

**Supplementary Table 3. Functional enrichment analysis of the parent genes of OS-GCAS events.**