

## ORIGINAL RESEARCH

# Development of a nomogram for prognostic prediction of lower-grade glioma based on alternative splicing signatures

Yaning Wang  | Zihao Wang | Binghao Zhao | Wenlin Chen | Yu Wang  | Wenbin Ma

Departments of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

**Correspondence**

Yu Wang and Wenbin Ma, Departments of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. Email: ywang@pumch.cn; mawb2001@hotmail.com

**Funding information**

The Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences, Grant/Award Number: 2016-I2M-2-001; Tsinghua University-Peking Union Medical College Hospital Initiative Scientific Research Program, Grant/Award Number: 2019ZLH101; The Beijing Municipal Natural Science Foundation, Grant/Award Number: 19JCZDJC64200(Z) and 7202150

**Abstract**

**Background:** The prognosis of lower-grade glioma (LGG) differs from that of other grades gliomas. Although lots of studies on the prognostic biomarkers of LGG have been reported, few have significant clinical impact. Alternative splicing (AS) events can affect cell function by splicing precursor mRNA. Therefore, a prognostic model for LGG based on AS events are important to establish.

**Methods:** RNA sequencing, clinical, and AS event data of 510 LGG patients from the TCGA database were downloaded. Univariate Cox regression analysis was used to screen out prognostic-related AS events and LASSO regression and multivariate Cox regression were used to establish prognostic risk scores for patients in the training set ( $n = 340$ ). After validation, a nomogram model was established based on the AS signature and clinical information, which was able to predict 1-, 3-, and 5-year survival rates. Finally, considering the regulatory effect of splicing factors (SFs) on AS events, an AS-SF regulatory network was analyzed.

**Results:** The most common AS event was exon skipping and the least was mutually exclusive exons. All the seven AS events were related to the prognosis of LGG patients, regardless of whether they were separated or considered as a whole event (integrated AS event), and the integrated AS event had the most significant correlation. After further inclusion of clinical indicators, eight factors were screened out: age, new event, KPS, WHO grade, treatment, integrated AS signature, IDH1 and TP53 mutation status, and a nomogram model was established. The study also constructed an AS-SF regulatory network.

**Conclusion:** The AS events and clinical factors that can predict the prognosis of LGG patients were screened, and a prognostic prediction model was established. The results of this study can play an important role in clinical work to better evaluate the prognosis of patients and impact treatment options.

**KEYWORDS**

alternative splicing, lower-grade glioma, nomogram, prediction model

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

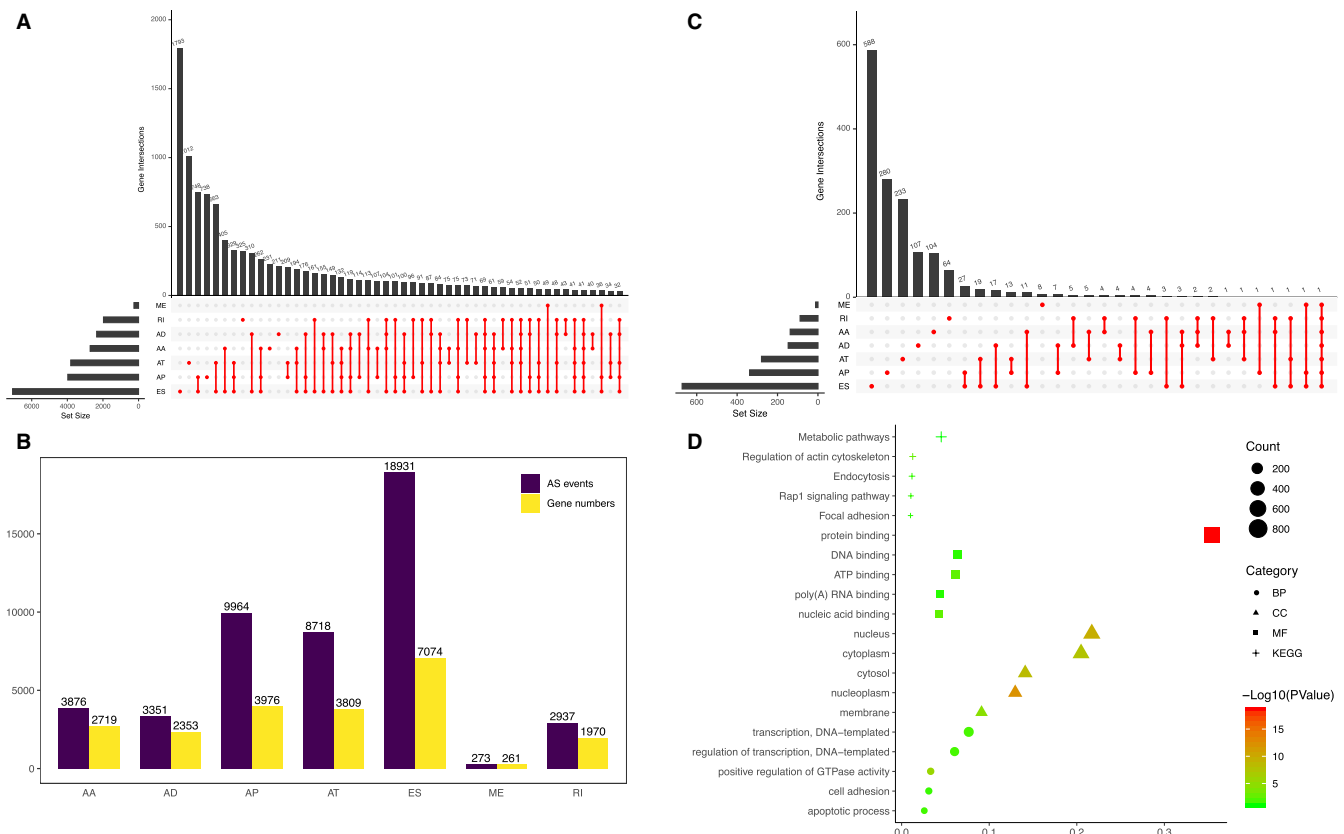
## 1 | INTRODUCTION

Glioma, the most common primary intracranial malignant tumor, can be classified into four grades according to the World Health Organization (WHO), with diffuse lower-grade glioma (LGG) including grades II and III.<sup>1</sup> The clinical characteristics of LGG patients vary, and the overall survival (OS) and progression-free survival (PFS) are significantly different, so neither clinical workers nor patients can predict the prognosis.<sup>2</sup> Numerous studies have been conducted on the prognostic biomarkers of LGG patients. These studies have included not only the well-known IDH1 mutation status, but also some factors discovered in recent years by bioinformatics analysis, such as eukaryotic initiation factor and some N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) RNA methylation regulators<sup>3,4</sup>; however, according to the 2016 WHO brain tumor classification, IDH1 mutation status and 1p/19q are the only universally recognized significant prognostic biomarkers of LGG.<sup>5</sup> Therefore, biomarkers with predictive value for the prognosis of LGG patients are still necessary to better guide the follow-up review and adjuvant therapy of these patients.

Splicing introns from precursor mRNA, preserving only exons and forming mature mRNA, are necessary steps in

intracellular translation. This process plays a crucial role in gene expression.<sup>6</sup> AS events entrust the diversity of the cellular proteome and influence the biological characteristics of cells in this way.<sup>7</sup> SFs, which are proteins involved in the regulation of AS, also play an important role in the process of splicing by influencing AS.<sup>8</sup> Recent studies have shown that AS, due to its influence on gene mutations and expression levels, may cause tumor cell proliferation by affecting oncogenes and inhibiting tumor cell apoptosis.<sup>9,10</sup> In addition, AS plays a role in drug resistance and tumor metastasis.<sup>11,12</sup>

Considering the broad implications of AS, an increasing number of studies have explored AS events as a prognostic biomarker for cancer patients. In recent years, with the establishment of a tumor database, the bioinformatics analysis between AS events and the outcome of cancer patients has also become a research hotspot.<sup>13</sup> Although this research continues to emerge, an effective and practical prediction model for LGG patient prognosis based on AS events has not yet been established. Therefore, in this study, bioinformatics analysis technology was used to comprehensively and deeply analyze the clinical information and Splice-Seq information of LGG patients in The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer>).



**FIGURE 1** Landscape of alternative splicing (AS) events in lower-grade glioma (LGG) patients. (A) The distribution of the seven AS events shown in the UpSet plot. (B) Seven types of AS events and the number of their source genes. (C) An UpSet plot to show the prognostic AS events in LGG patients. (D) Functional enrichment analysis of source genes corresponding to survival-related AS events

Variables	Training set ( <i>n</i> = 340)	Validation set ( <i>n</i> = 170)	<i>p</i> value
Age (years)	42.9 ± 13.4	42.4 ± 13.1	0.511
Sex			0.166
Female	158	68	
Male	182	102	
New event			0.414
None or NA	151	82	
Yes	189	88	
KPS			0.593
<80	25	15	
≥80	172	91	
NA	143	64	
Histology			0.737
Astrocytoma	127	67	
Oligoastrocytoma	85	45	
Oligodendroglioma	128	58	
WHO grade			0.754
Grade II	163	84	
Grade III	177	86	
Pharmaceutical therapy			0.779
TMZ	155	82	
PCV	22	8	
TMZ + PCV	3	2	
Others	11	3	
NA	149	75	
Radiation therapy			0.843
No	113	53	
Yes	184	93	
NA	43	24	
IDH mutation status			0.764
Wild type	78	37	
Mutant	262	133	
TP53 mutation status			0.060
Wild type	166	98	
Mutant	174	72	
PTEN mutation status			0.887
Wild type	323	161	
Mutant	17	9	
EGFR mutation status			0.091
Wild type	320	153	
Mutant	20	17	
ATRX mutation status			0.699
Wild type	208	107	
Mutant	132	63	

**TABLE 1** Clinicopathological characteristics of the TCGA LGG patients in the training set (*n* = 340) and internal validation set (*n* = 170)

Note: "New event" included progression and recurrence. "Others" in pharmaceutical therapy included CT + TMT, CT + TMT + Immunotherapy, and CT + Immunotherapy.

Abbreviations: CT, chemotherapy; KPS, Karnofsky performance score; LGG, lower-grade glioma; NA, not available; PCV, procarbazine, lomustine (CCNU), and vincristine; TMT, targeted molecular therapy; TMZ, temozolomide.

gov/) database, and a prognostic prediction model was established to better serve both clinical work and the patients themselves.

## 2 | METHODS

### 2.1 | Data inclusion and processing

This study downloaded the RNA sequencing results and clinical data of 529 patients with LGG in the TCGA (<https://portal.gdc.cancer.gov/>) and in the Splice-Seq section of the same database. Information on the AS events of these patients was also downloaded, including the percent-spliced-in (PSI) score. This score ranges from 0 to 1, is often used for evaluating AS events, and describes the existence of AS junctions in the exons of the clinical samples we downloaded.<sup>14</sup> This study used this score to screen all AS events and chose seven of them, including alternate acceptor site (AA), exon skipping (ES), alternate donor site (AD), alternative promoter (AP), mutually exclusive exons (ME), alternate terminator (AT), and retained intron (RI; Figure 1A).

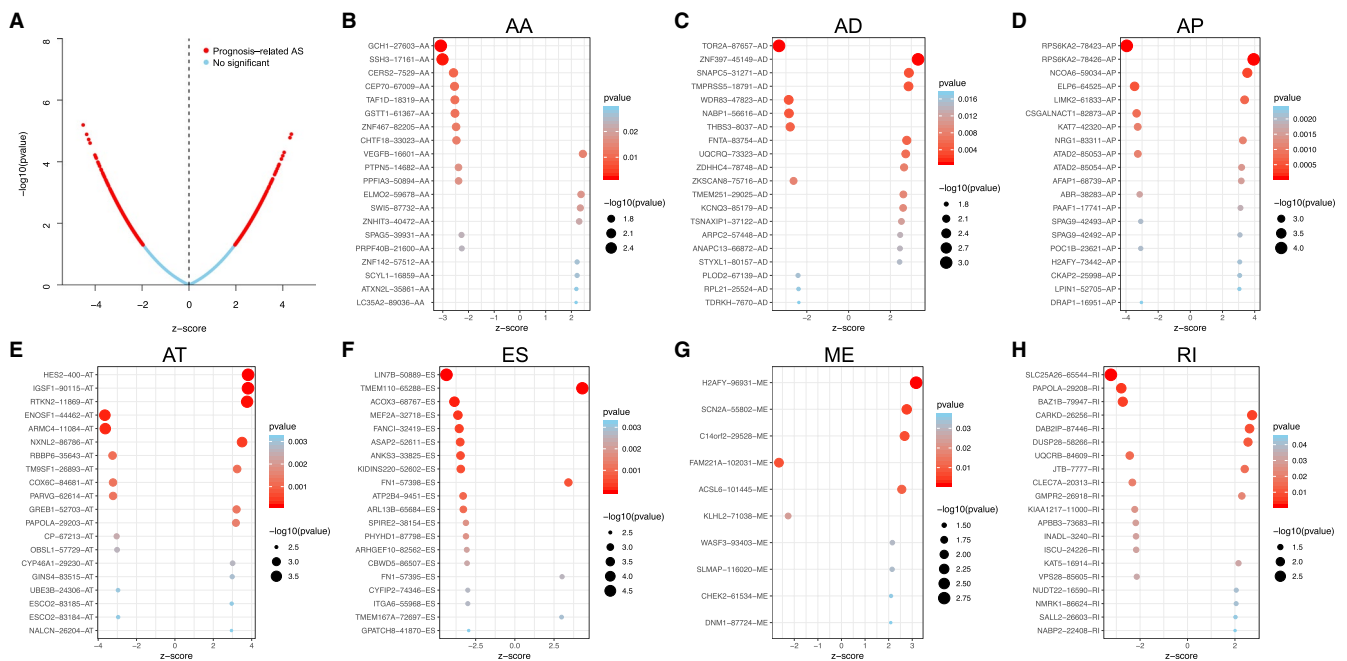
The inclusion criteria of the clinical samples were as follows: (a) age  $\geq 18$  years old; (b) patients were diagnosed with LGG; (c) RNA-Seq data could be obtained; (d) splice-Seq data could be obtained; (e) complete clinical information was available. After screening all 529 patients according to the above inclusion criteria, 510 LGG patients were finally enrolled.

### 2.2 | Constructing a prognostic prediction model based on AS events

First, we divided the data of the above 510 patients into a training set ( $n = 340$ ) and a validation set ( $n = 170$ ) according to a 2:1 ratio randomly. Their Clinicopathological characteristics are shown in Table 1. Then, to clarify the correlation between the screened AS events and the survival of LGG patients, we used univariate Cox regression analysis to analyze the PSI score of AS events and the OS of patients using R (Ver 3.5.1). Then, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID) to complete the functional annotation and pathway analysis for the genes that corresponded to the above AS events. The tools used in the above analysis were Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses, and  $p < 0.05$  was regarded as statistically significant (Table S1).

After the univariate Cox regression analysis, we further screened prognostic AS events using least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox regression, and a prognostic scoring formula was established for these AS events as follows: risk score = PSI value of AS event<sub>1</sub>  $\times$   $\beta_1$  + PSI value of AS event<sub>2</sub>  $\times$   $\beta_2$  + PSI value of AS event<sub>3</sub>  $\times$   $\beta_3$  ... + PSI value of AS event<sub>n</sub>  $\times$   $\beta_n$ . In this formula,  $\beta$  represents the regression coefficient in the multivariate Cox regression analysis.

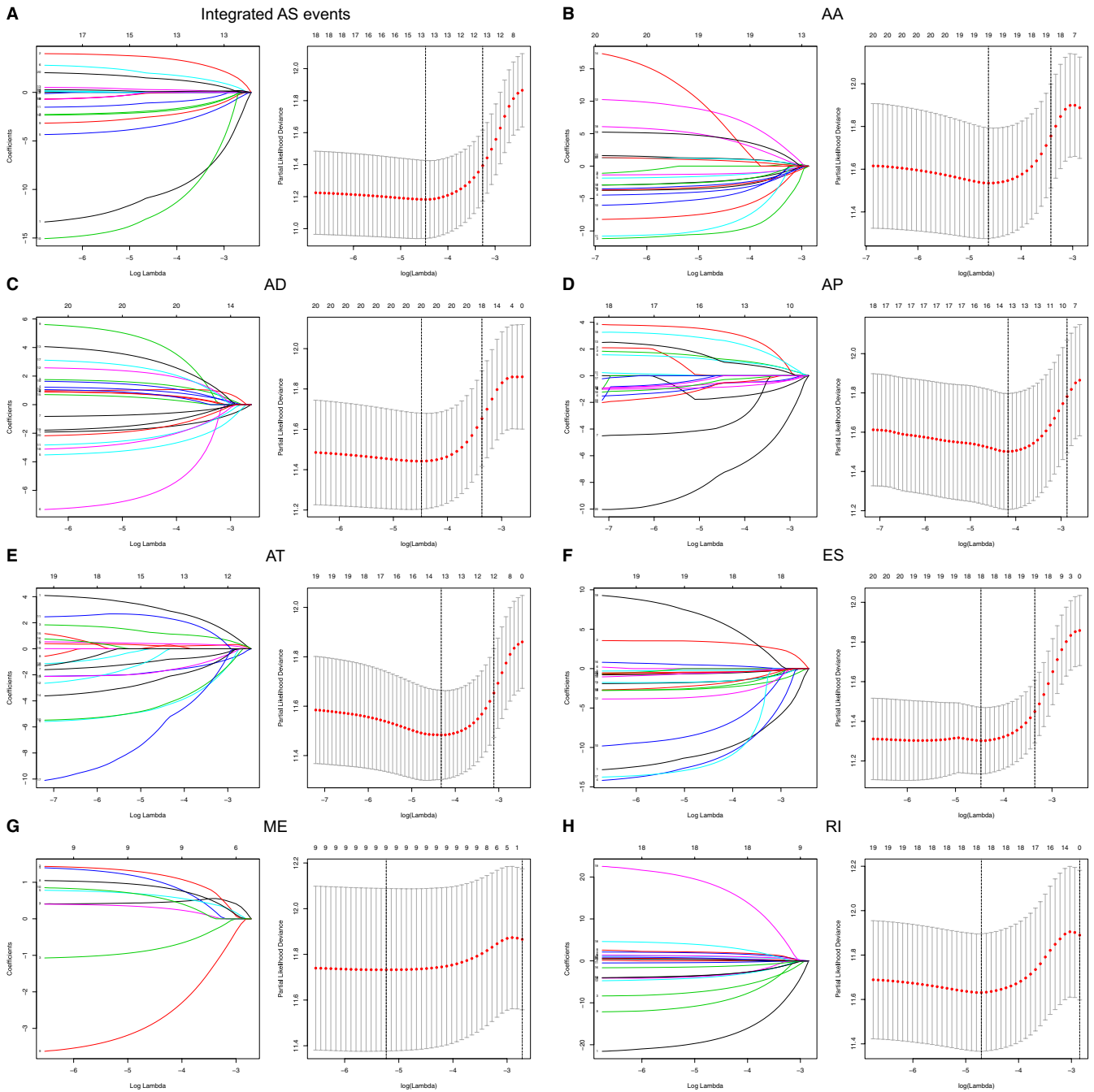
To verify the clinical value of the prediction model, we first scored 340 patients in the training set and divided them into high- and low-risk groups according to the scores, with the median of the score as the boundary. Then,



**FIGURE 2** Prognostic alternative splicing (AS) events in lower-grade glioma (LGG) patients. (A) A volcano plot shows the overview of prognostic AS events. (B–H) Bubble plots show the top 20 prognostic AS events ( $z\text{-score} > 0$ , HR  $> 1$ ;  $z\text{-score} < 0$ , HR  $< 1$ )

Kaplan–Meier (K–M) survival curve analysis was used to analyze the survival of these two groups. The difference in survival between the high- and low-risk groups was analyzed by a two-sided log-rank test. Then, we used Harrell's concordance index (C-index) and time-dependent receiver operating characteristic (ROC) curve analysis to verify the accuracy and effectiveness of the AS signature prognostic

prediction. In addition, to solve the problem of whether AS events are an independent influencing factor, we included all clinical information and carried out a multivariate Cox proportional hazards regression analysis for all factors. Finally, to make the conclusions more rigorous, we performed a further validation using a similar method in the validation set ( $n = 170$ ). Considering that in different LGG



**FIGURE 3** Alternative splicing (AS) events related to the prognosis of lower-grade glioma (LGG) patients were screened and verified with LASSO regression. The left panel is a coefficient profile plot. A vertical line is shown at the value that was selected by 10-fold cross-validation. The optimal log produced 10 characteristics of nonzero coefficients. The right panel depicts the process of selecting optimal parameters by LASSO regression using 10-fold cross-validation. We plotted the partial likelihood deviance curve with log, and a dotted vertical line is shown at the optimal values. This process was accomplished using the 1 standard error of the minimum and maximum criteria

patients different biomarkers may have a greater impact on the prognosis, we also conducted a subgroup analysis of the patients.

To ensure that the results of this study would be meaningful for clinical work, we further included the clinical information of patients in the TCGA database, performed univariate Cox analysis, and finally, screened out the factors that have prognostic significance for patients.

### 2.3 | Establishing a nomogram of the prognostic model according to AS events and clinical information

The clinical information of the LGG patients, including age, sex, new event (including progression and recurrence), Karnofsky performance status (KPS) score, surgery, pharmaceutical therapy, radiotherapy, IDH1 mutation status, TP53 mutation status,

**TABLE 2** The prognostic risk score models based on the PSI values of AS event types of the LGG training set ( $n = 340$ )

AS event types	Formula (risk score model)	C-index (95% CI)	<i>p</i> value
Integrated AS signature	$\text{LIN7B-50889-ES} \times (-12.74) + \text{TMEM110-65288-ES} \times 4.29 + \text{RPS6KA2-78423-AP} \times (-2.35) + \text{ACOX3-68767-ES} \times (-4.42) + \text{HES2-400-AT} \times 2.33 + \text{ENOSF1-44462-AT} \times (-3.06) + \text{MEF2A-32718-ES} \times (-15.58) + \text{ANKS3-33825-ES} \times (-2.41) + \text{LIMK2-61833-AP} \times 1.85$	0.791 (0.752–0.830)	$1.54 \times 10^{-17}$
AA signature	$\text{GCH1-27603-AA} \times (-4.03) + \text{SSH3-17161-AA} \times (-3.70) + \text{CERS2-7529-AA} \times (-16.89) + \text{CEP70-67009-AA} \times (-4.25) + \text{TAF1D-18319-AA} \times (-2.15) + \text{ZNF467-82205-AA} \times (-2.86) + \text{CHTF18-33023-AA} \times (-8.09) + \text{PTPN5-14682-AA} \times (-5.85) + \text{ELMO2-59678-AA} \times 10.57 + \text{ZNHIT3-40472-AA} \times 18.52 + \text{SPAG5-39931-AA} \times (-3.54) + \text{PRPF40B-21600-AA} \times (-3.84) + \text{ZNF142-57512-AA} \times 1.65 + \text{SCYL1-16859-AA} \times 6.96 + \text{ATXN2L-35861-AA} \times 5.83$	0.739 (0.699–0.778)	$4.88 \times 10^{-16}$
AD signature	$\text{TOR2A-87657-AD} \times (-2.36) + \text{SNAPC5-31271-AD} \times 2.06 + \text{TMPRSS5-18791-AD} \times 1.40 + \text{WDR83-47823-AD} \times (-3.64) + \text{NABP1-56616-AD} \times (-7.22) + \text{THBS3-8037-AD} \times (-1.14) + \text{ZDHHC4-78748-AD} \times 1.75 + \text{ZKSCAN8-75716-AD} \times (-2.53) + \text{TMEM251-29025-AD} \times 3.20 + \text{KCNQ3-85179-AD} \times 4.58 + \text{TSNAXIP1-37122-AD} \times 1.06 + \text{STYXL1-80157-AD} \times 3.67 + \text{PLOD2-67139-AD} \times (-3.40) + \text{RPL21-25524-AD} \times (-2.02) + \text{TDRKH-7670-AD} \times (-1.94)$	0.739 (0.685–0.794)	$7.92 \times 10^{-18}$
AP signature	$\text{RPS6KA2-78423-AP} \times (-2.16) + \text{NCOA6-59034-AP} \times 1.71 + \text{LIMK2-61833-AP} \times 1.61 + \text{CSGALNACT1-82873-AP} \times (-1.07) + \text{NRG1-83311-AP} \times 4.44 + \text{CKAP2-25998-AP} \times 3.89 + \text{DRAP1-16951-AP} \times (-9.90)$	0.727 (0.675–0.778)	$2.32 \times 10^{-17}$
AT signature	$\text{HES2-400-AT} \times 4.00 + \text{RTKN2-11869-AT} \times 1.62 + \text{ENOSF1-44462-AT} \times (-1.99) + \text{ARMC4-11084-AT} \times (-5.67) + \text{PARVG-62614-AT} \times (-5.35) + \text{CP-67213-AT} \times (-2.58) + \text{OBSL1-57729-AT} \times (-3.05) + \text{UBE3B-24306-AT} \times (-8.60)$	0.697 (0.642–0.753)	$4.35 \times 10^{-12}$
ES signature	$\text{LIN7B-50889-ES} \times (-14.51) + \text{TMEM110-65288-ES} \times 3.66 + \text{ACOX3-68767-ES} \times (-2.95) + \text{MEF2A-32718-ES} \times (-13.74) + \text{ANKS3-33825-ES} \times (-1.93) + \text{KIDINS220-52602-ES} \times (-0.87) + \text{ATP2B4-9451-ES} \times (-10.29) + \text{ARL13B-65684-ES} \times (-2.03) + \text{SPIRE2-38154-ES} \times (-5.10) + \text{ARHGEF10-82562-ES} \times (-2.94) + \text{CBWD5-86507-ES} \times (-3.01) + \text{CYFIP2-74346-ES} \times (-14.51) + \text{TMEM167A-72697-ES} \times 10.70$	0.735 (0.686–0.785)	$1.81 \times 10^{-20}$
ME signature	$\text{ACSL6-101445-ME} \times 1.46 + \text{KLHL2-71038-ME} \times (-1.24) + \text{SLMAP-116020-ME} \times 1.41 + \text{C14orf2-29528-ME} \times 0.93 + \text{SCN2A-55802-ME} \times 1.20 + \text{FAM221A-102031-ME} \times (-3.97) + \text{WASF3-93403-ME} \times 1.02$	0.681 (0.642–0.720)	$2.32 \times 10^{-8}$
RI signature	$\text{SLC25A26-65544-RI} \times (-20.42) + \text{BAZ1B-79947-RI} \times (-8.98) + \text{CARKD-26256-RI} \times 2.59 + \text{DUSP28-58266-RI} \times 1.94 + \text{UQCRB-84609-RI} \times (-4.45) + \text{CLEC7A-20313-RI} \times (-13.52) + \text{APBB3-73683-RI} \times (-5.29) + \text{INADL-3240-RI} \times (-4.11) + \text{ISCU-24226-RI} \times (-4.40) + \text{VPS28-85605-RI} \times (-1.98) + \text{NMRK1-86624-RI} \times 4.32 + \text{SALL2-26603-RI} \times 22.37$	0.727 (0.679–0.775)	$2.01 \times 10^{-20}$

Abbreviations: AA, alternate acceptor site; AD, alternate donor site; AP, alternate promoter; AS, alternative splicing; AT, alternate terminator; CI, confidence interval; ES, exon skip; LGG, lower-grade glioma; ME, mutually exclusive exons; PSI, percent-spliced-in; RI, retained intron.



the patients' AS signature, and integrated AS signature, which means considering AS events as a whole for the analysis, were included in the multivariate Cox regression analysis. After this analysis, we established the nomogram scoring system by using all the independent prognostic influencing factors, aiming to form a scoring system capable of evaluating the patients' 1-, 3-, and 5-year OS. To demonstrate the effectiveness of the system, its discrimination performance was evaluated using the C-index, area under the ROC curve (AUC) and calibration plots.<sup>15,16</sup> The clinical applicability of this scoring system was evaluated using decision curve analysis.<sup>17</sup> Finally, to verify the effectiveness of the prognostic prediction model, we used the validation set ( $n = 170$ ) to perform analyses including C-index, AUC, and calibration plots. All the above analyses were conducted using R (version 3.5.1), and a  $p < 0.05$  was considered statistically significant.

## 2.4 | Correlation analysis between AS events and SFs and the establishment of a regulatory network among them

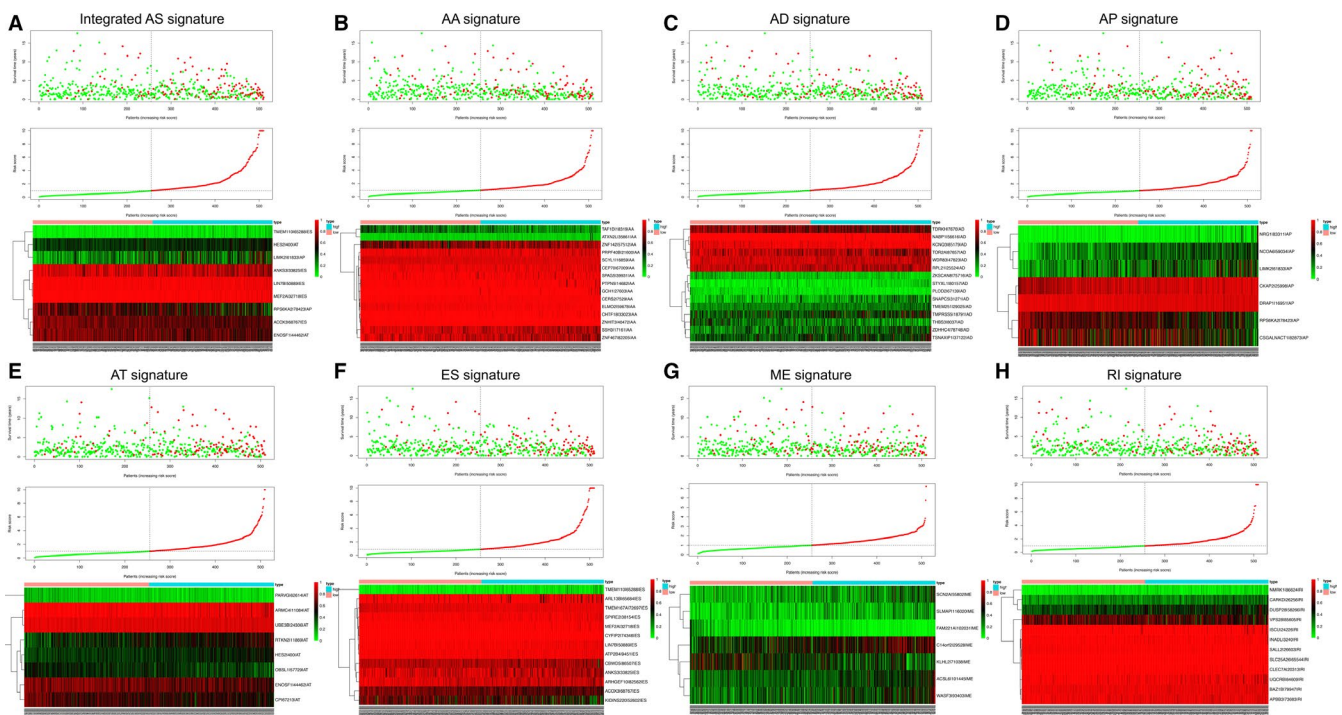
In the tumor microenvironment, SFs can regulate AS events, therefore, the establishment of a prognostic prediction model based on AS events needs to explore the relationship between SFs and AS events. First, Pearson's correlation test

was used to analyze the PSI value of AS events and the expression level of SFs that could regulate these AS events.  $P < 0.001$  and correlation coefficients  $>0.6$  or  $<-0.6$  were considered to be statistically significant. Then, based on the analysis results, we established a regulatory network between AS events and SFs and visualized the regulatory network using Cytoscape.

## 3 | RESULTS

### 3.1 | Distribution of AS events in LGG patients in the TCGA

By analyzing the Splice-Seq data of 510 LGG patients in the TCGA database, we obtained a total of 48,050 AS events and 22,162 corresponding genes. Among them were 3876 AA events and 2719 corresponding genes; 3351 AD events and 2353 corresponding genes; 9964 AP events and 3976 corresponding genes; 8718 AT events and 3809 corresponding genes; 18,931 ES events and 7074 corresponding genes; 273 ME events and 261 corresponding genes; and 2937 RI events and 1907 corresponding genes (Figure 1B). Among all AS events, ES was the most common, accounting for 39.40%. The second most common was AP and the least common was ME.



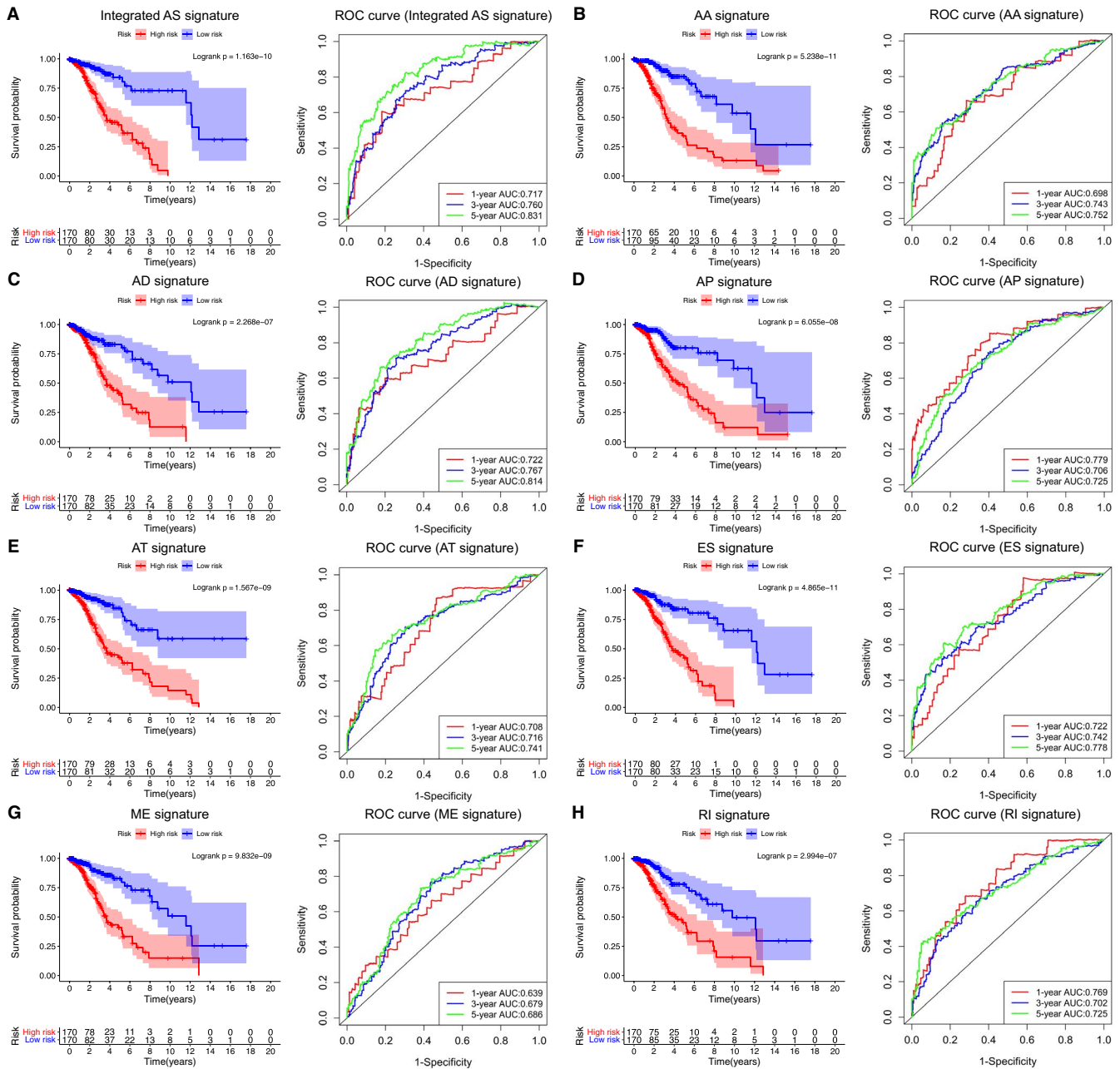
**FIGURE 4** We analyzed the risk scores of seven types of alternative splicing (AS) signatures and integrated AS signatures in lower-grade glioma (LGG) patients. The upper panel shows the distribution of the survival time of patients with different risk scores. The middle panel figure shows the variation trend of patient survival time with risk scores. The bottom panel shows the heat map of survival-related AS events. The red and green in the figure represent high- and low-expression levels, respectively. The patients were divided into high- and low-risk groups according to the relationship between the risk score and survival time

### 3.2 | Screening of prognostic AS events in LGG patients and functional annotation of corresponding genes

After completing a univariate Cox regression analysis of all AS events, we screened 2,185 prognostic AS events and 1,520 corresponding source genes (Figure 1C). In the corresponding source genes, we conducted a GO analysis, which included biological processes (BPs), cellular components

(CCs), and molecular functions (MFs). According to the BP results, these genes were mainly enriched in transcription and the regulation of transcription. The results of the CC analysis showed that these genes were mainly enriched in the nucleus and cytoplasm. The MF results indicated that these genes were enriched in protein, DNA and ATP binding (Figure 1D).

According to the hazard ratio (HR) or  $z$ -score, AS events that are predictive of patient prognosis were divided into two categories: favorable prognostic factors (HR < 1 or



**FIGURE 5** Survival analysis of the patients based on the alternative splicing (AS) signature and validation of the prognostic performance of all AS signatures and integrated AS signature in the training set ( $n = 340$ ). The left panel shows the K–M survival curve of the patients in the high- and low-risk groups. The right panel shows the ROC curve demonstrating the prognostic performance of each AS signature and integrated AS signature



**TABLE 3** Univariate and multivariate cox proportional hazards analysis of clinical parameters and AS event-based risk score model of the TCGA LGG patients in the training set ( $n = 340$ ) and internal validation set ( $n = 170$ )

Variables	Training set ( $n = 340$ )				Validation set ( $n = 170$ )			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	$p$	HR (95% CI)	$p$	HR (95% CI)	$p$	HR (95% CI)	$p$
Age	1.03 (1.02–1.04)	$5.67 \times 10^{-6}$	1.02 (1.01–1.04)	$2.44 \times 10^{-3}$	1.03 (1.02–1.05)	$3.13 \times 10^{-5}$	1.03 (1.01–1.05)	$4.18 \times 10^{-3}$
Sex (Female/Male)	0.90 (0.63–1.27)	0.54	—	—	0.88 (0.58–1.34)	0.53	—	—
New event (None or NA/Yes)	2.11 (1.35–3.30)	$1.08 \times 10^{-3}$	1.65 (1.03–2.63)	$3.57 \times 10^{-2}$	2.07 (1.20–3.57)	$8.92 \times 10^{-3}$	1.70 (1.66–1.74)	$1.03 \times 10^{-2}$
KPS (<80/ $\geq 80$ /NA)	0.61 (0.45–0.82)	$1.25 \times 10^{-3}$	0.84 (0.80–0.88)	$3.05 \times 10^{-5}$	0.62 (0.47–0.88)	$8.27 \times 10^{-3}$	0.84 (0.79–0.88)	$3.86 \times 10^{-2}$
Histology (Astrocytoma/ Oligoastrocytoma/ Oligodendroglioma)	0.98 (0.80–1.20)	0.83	—	—	1.10 (0.87–1.40)	0.43	—	—
WHO grade (Grade II/III)	2.15 (1.48–3.11)	$5.96 \times 10^{-5}$	1.46 (1.07–1.85)	$7.53 \times 10^{-3}$	2.24 (1.43–3.50)	$3.96 \times 10^{-4}$	1.74 (1.05–2.93)	$3.18 \times 10^{-3}$
Pharmaceutical therapy (TMZ/PCV/TMZ + PCV/ Others/NA)	0.84 (0.80–0.88)	$1.23 \times 10^{-6}$	0.71 (0.67–0.75)	$8.90 \times 10^{-3}$	0.82 (0.76–0.87)	$1.47 \times 10^{-2}$	1.09 (1.05–1.13)	$2.43 \times 10^{-2}$
Radiation therapy (No/Yes/NA)	1.13 (0.84–1.52)	0.42	—	—	1.01 (0.72–1.44)	0.94	—	—
Surgery (Biopsy only/Tumor resection)	0.82 (0.59–1.44)	0.77	—	—	0.93 (0.54–1.32)	0.81	—	—
IDH1 mutation status (Wildtype/Mutant)	0.40 (0.28–0.57)	$3.85 \times 10^{-7}$	0.64 (0.44–0.94)	$2.17 \times 10^{-2}$	0.37 (0.24–0.56)	$3.90 \times 10^{-6}$	0.70 (0.66–0.74)	$1.58 \times 10^{-2}$
TP53 mutation status (Wildtype/Mutant)	2.01 (1.39–2.92)	$2.13 \times 10^{-4}$	1.46 (1.07–1.86)	$5.95 \times 10^{-3}$	2.03 (1.30–3.17)	$1.85 \times 10^{-3}$	1.85 (1.12–3.06)	$1.63 \times 10^{-2}$
PTEN mutation status (Wildtype/Mutant)	1.27 (0.59–2.74)	0.53	—	—	1.95 (0.90–4.22)	0.09	—	—
EGFR mutation status (Wildtype/Mutant)	0.88 (0.46–1.68)	0.69	—	—	1.27 (0.61–2.63)	0.53	—	—
ATRX mutation status (Wildtype/Mutant)	0.85 (0.59–1.23)	0.39	—	—	0.75 (0.48–1.18)	0.21	—	—

(Continues)

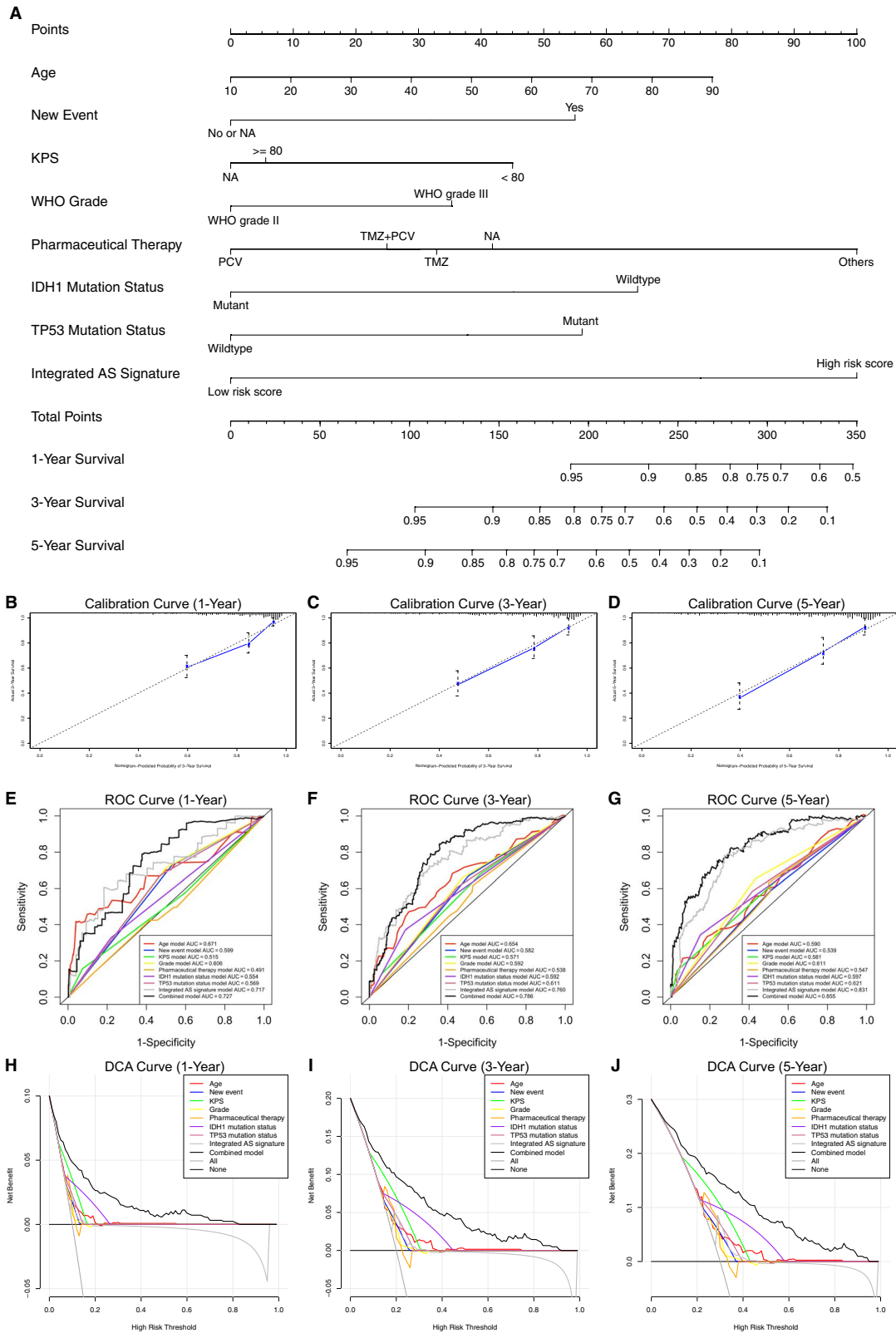
TABLE 3 (Continued)

Variables	Training set ( <i>n</i> = 340)			Validation set ( <i>n</i> = 170)				
	Univariate analysis			Univariate analysis				
	HR (95% CI)	<i>p</i>		HR (95% CI)	<i>p</i>			
Integrated AS signature (Low-risk score/High-risk score)	4.81 (3.06–7.56)	<b>8.76 × 10<sup>-12</sup></b>	1.99 (1.32–3.02)	<b>3.63 × 10<sup>-9</sup></b>	5.23 (2.99–9.13)	<b>5.81 × 10<sup>-9</sup></b>	2.91 (1.72–4.94)	<b>7.03 × 10<sup>-5</sup></b>
AA signature (Low-risk score/High-risk score)	3.71 (2.51–5.48)	<b>4.48 × 10<sup>-11</sup></b>	1.30 (0.74–2.26)	0.36	4.40 (2.72–7.11)	<b>1.55 × 10<sup>-9</sup></b>	1.30 (0.63–2.66)	0.47
AD signature (Low-risk score/High-risk score)	3.45 (2.32–5.13)	<b>9.20 × 10<sup>-10</sup></b>	1.16 (0.77–1.55)	0.096	3.29 (2.05–5.28)	<b>8.67 × 10<sup>-7</sup></b>	1.37 (0.76–2.46)	0.30
AP signature (Low-risk score/High-risk score)	3.66 (2.44–5.50)	<b>4.21 × 10<sup>-10</sup></b>	1.06 (0.67–1.45)	0.26	3.55 (2.18–5.79)	<b>3.56 × 10<sup>-7</sup></b>	1.49 (0.82–2.70)	0.19
AT signature (Low-risk score/High-risk score)	4.28 (2.74–6.69)	<b>1.64 × 10<sup>-10</sup></b>	1.50 (0.89–2.51)	0.13	4.24 (2.55–7.05)	<b>2.80 × 10<sup>-8</sup></b>	1.37 (0.75–2.49)	0.30
ES signature (Low-risk score/High-risk score)	4.05 (2.64–6.20)	<b>1.45 × 10<sup>-10</sup></b>	1.05 (0.65–1.44)	0.32	5.28 (3.06–9.12)	<b>2.27 × 10<sup>-9</sup></b>	2.26 (0.88–4.54)	0.21
ME signature (Low-risk score/High-risk score)	2.68 (1.84–3.92)	<b>3.34 × 10<sup>-7</sup></b>	1.09 (0.71–1.66)	0.69	3.68 (2.29–5.92)	<b>7.08 × 10<sup>-8</sup></b>	1.43 (0.81–2.51)	0.22
RI signature (Low-risk score/High-risk score)	3.14 (2.12–4.66)	<b>1.26 × 10<sup>-8</sup></b>	1.29 (0.81–2.06)	0.28	3.14 (1.98–4.98)	<b>1.08 × 10<sup>-6</sup></b>	1.10 (0.63–1.91)	0.74

Note: "New event" included progression and recurrence. "Others" in pharmaceutical therapy included CT + TMT, CT + TMT + Immunotherapy, and CT + Immunotherapy. All statistical tests were two-sided.

Bold numbers indicate that the P value is significant, which means that the corresponding factors are associated with the outcome of patients significantly.

Abbreviations: AA, alternate acceptor site; AD, alternate donor site; AP, alternate promoter; AS, alternative splicing; AT, alternate terminator; CI, confidence interval; CT, chemotherapy; ES, exon skip; HR, hazard ratio; KPS, Karnofsky performance score; LGG, lower-grade glioma; ME, mutually exclusive exons; NA, not available; PCV, procarbazine, lomustine (CCNU), and vincristine; RI, retained intron; TMT, targeted molecular therapy; TMZ, temozolomide.



**FIGURE 6** The nomogram of the prognostic prediction model was established based on the alternative splicing (AS) signature and the validation of the model on the predictive efficacy of the survival time in lower-grade glioma (LGG) patients. (A) The prognostic prediction model established for LGG patients that can predict the 1-, 3-, and 5-year survival. (B–D) A calibration plot to verify the prognostic performance of the model with the predicted value on the x-axis and the actual value on the y-axis. (E–G) A ROC curve to verify the prognostic performance of the model through comparison with other single factors. (H–J) A DCA curve to verify the prognostic performance of the model by comparison with other single factors

$z$ -score  $< 0$ ) and unfavorable prognostic factors (HR  $> 1$  or  $z$ -score  $> 0$ ). Figure 2A–H shows the top 20 AS events of all 7 AS events, with most being favorable events (1247 vs. 938).

### 3.3 | Establishment and validation of the prognostic risk score model based on AS events

We used the above seven AS events and integrated the AS events to further carry out LASSO regression and multivariate Cox regression analyses based on the results of the univariate Cox regression analysis (Figure 3A–H). As shown in Table 2, we established prognostic risk score formulas for the seven AS events and integrated the AS events (Table S2), and these formulas were used to score each LGG patient in our study. Then, with the median-risk score as the boundary, 340 patients in the training set were divided into the high-risk group (score higher than the median-risk score) and the low-risk group (score lower than the median-risk score; Figure 4A–H). For the two groups of patients, a K-M survival curve analysis was conducted, which showed that the OS of the eight groups of patients in the high-risk group was significantly shorter than the low-risk group ( $p < 0.05$ , Figure 5A–H). Among them, the integrated AS signature had the highest degree of differentiation for OS, with a C-index of 0.791 (95% CI, 0.752–0.830;  $p = 1.54 \times 10^{-17}$ ), which indicates that this score plays the most significant role in all AS-based risk scores (Table 2). Although the impact of the integrated AS signature was the most significant among all AS signatures, the remaining seven signatures also showed a good predictive effect on patients' 1-, 3-, and 5-year OS rates after being analyzed by a time-dependent ROC analysis. The above results demonstrate the importance of AS events for LGG patients.

Finally, to verify that AS events have a significant impact on patient survival, we used a similar method for verification in the validation set ( $n = 170$ ), and further determined the prognostic value of the AS signatures (Figure S1). Considering that some important biomarkers such as IDH may have a greater impact on the survival of LGG patients, we divided the patients into different subgroups based on biomarkers and analyzed the survival in the different subgroups. The results showed that the integrated AS signature still has the most significant predictive effect on the survival of patients in different subgroups (Figure S2).

### 3.4 | The AS signature and clinical factors are independent prognostic factors of LGG patients

After demonstrating the predictive effect of the above eight AS signatures on prognosis, we further included various clinical

information of the patients and analyzed them together with the AS signature (Table 3). Univariate Cox analysis showed that patient age ( $p < 0.001$ ), new event ( $p = 0.001$ ), KPS score ( $p = 0.001$ ), WHO grade ( $p < 0.001$ ), pharmaceutical treatment ( $p < 0.001$ ), IDH1 mutation status ( $p < 0.001$ ), TP53 mutation status ( $p < 0.001$ ), and all AS signatures ( $p < 0.001$ ) were significantly correlated with patient OS.

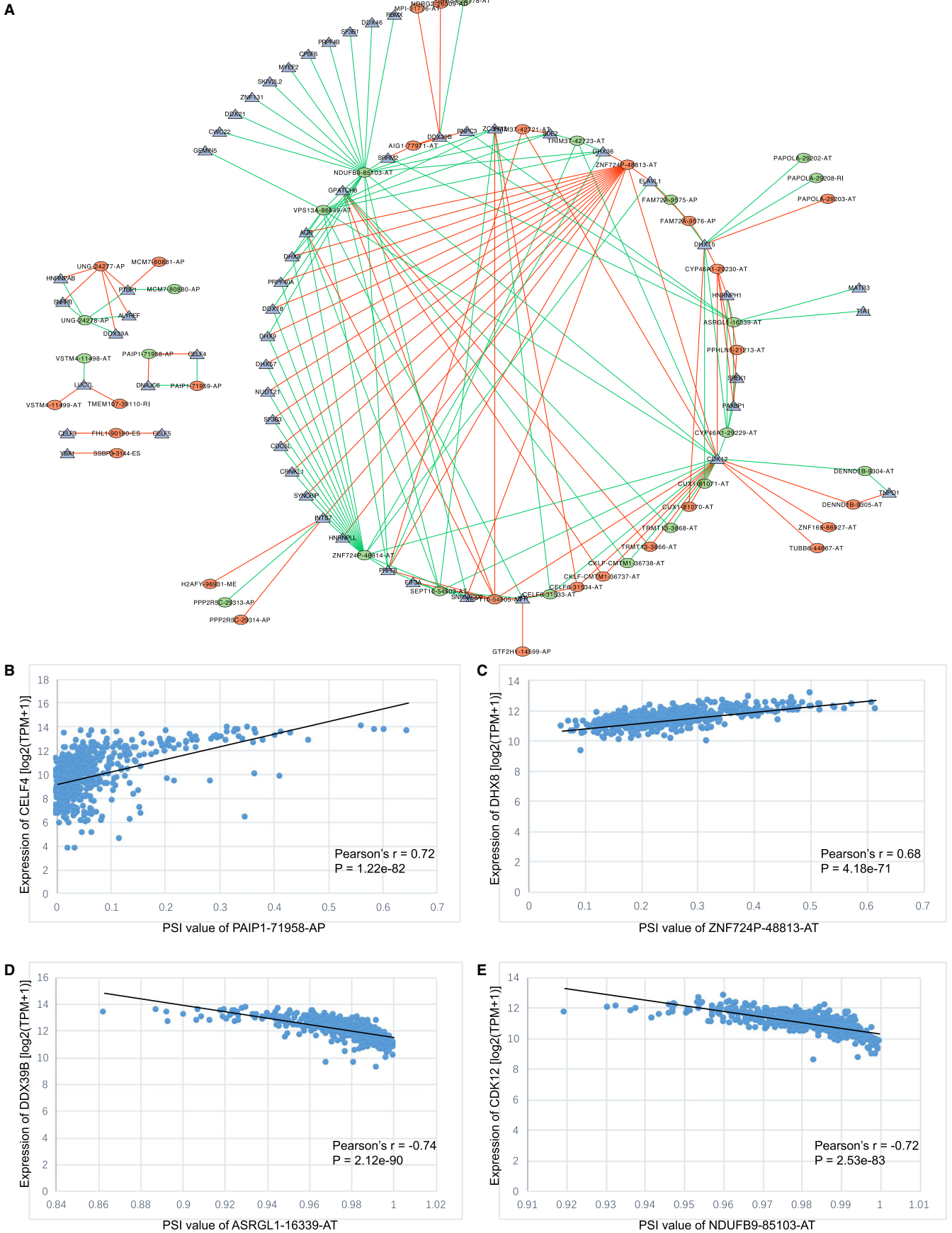
Based on the univariate Cox analysis, a multivariate Cox analysis was used to further screen independent factors that could affect the prognosis of patients including age ( $p = 0.002$ ), new event ( $p = 0.036$ ), KPS score ( $p < 0.001$ ), WHO grade ( $p = 0.002$ ), pharmaceutical treatment ( $p = 0.009$ ), IDH1 mutation status ( $p = 0.022$ ), TP53 mutation status ( $p = 0.006$ ), and integrated AS signature ( $p < 0.001$ ).

To further verify the above conclusions, we also analyzed the data of 170 LGG patients in the validation set and screened out the same prognostic prediction factors (Table 3). Therefore, we used the eight factors obtained above to carry out the next series of analyses, and established a nomogram model, which is more practical in clinical work.

### 3.5 | Establishment of a nomogram of the prognostic model that included the AS signature

To better apply the results of this study to clinical work, we established a nomogram prognostic score system based on the above results to predict the 1-, 3-, and 5-year survival of patients. This scoring system includes age, new event, KPS score, WHO grade, pharmaceutical therapy, IDH1 mutation status, TP53 mutation status, and integrated AS signature (Figure 6A). Then, to verify the reliability of this model, we conducted a calibration plot (Figure 6B–D) and an AUC of the ROC curve analysis (Figure 6E–G) again. The results all verified the practical value of the model. As shown in Figure 6E–G, the prediction effect of a single factor on the prognosis of patients is far less significant than that of a multifactor.

Then, we analyzed the above data again using decline curve analysis (DCA) to verify the ability of the nomogram to predict the 1-, 3-, and 5-year survival of patients (Figure 6H–J). The results showed that the nomogram as a whole was significantly better than individual factors in predicting prognosis. To verify the universality of the nomogram prognostic prediction model, we used the data of 170 LGG patients in the validation set to verify again, and further affirmed the predictive value of the model (Figure S3). Considering that no effective survival prediction model currently exists for LGG patients, the establishment of this model is undoubtedly of great significance to both clinical workers and patients.





**FIGURE 7** Analysis of the correlation between alternative splicing (AS) events and splicing factors (SFs). (A) The regulatory network between AS events and SFs. The green and red dots in the figure represent the favorable and unfavorable prognostic AS events, respectively; the blue triangle represents the SFs, and the green and red lines represent the negative and positive correlation between AS events and SFs, respectively. (B–C) The two AS-SF pairs with the strongest positive correlation are CELF4 and PAIP1-71958-AP (Pearson correlation coefficient = 0.72) and DHX8 and ZNF724P-48813-AT (Pearson correlation coefficient = 0.68). (D–E) The two AS-SF pairs with the strongest negative correlation are DDX39B and ASRGL1-16339-AT (Pearson correlation coefficient = -0.74) and CDK12 and NDUFB9-85103-A (Pearson correlation coefficient = -0.72)

### 3.6 | Regulatory network analysis between AS events and SFs

By conducting survival and correlation analyses on the AS sequencing data and RNA sequencing expression data, we screened 55 SFs and 48 AS events that were correlated with patient survival (Pearson correlation coefficient  $>0.6$  or  $<-0.6$ ,  $p < 0.001$ ; Table S3). The above SFs and AS events constitute a total of 173 SF-AS pairs, of which 73 pairs are positively correlated and 100 pairs are negatively correlated; we constructed these paired SFs and AS events into a regulatory network (Figure 7A) and found that the following four SF-AS pairs were most significantly correlated with each other: CELF4 and PAIP1-71958-AP, DHX8 and ZNF724P-48813-AT, DDX39B and ASRGL1-16339-AT, and CDK12 and NDUFB9-85103-AT (Figure 7B–E).

## 4 | DISCUSSION

As one of the factors that plays an important role in the development of tumors, AS events have gradually become a popular research topic for oncologists. Currently, AS events have been studied in lung cancer, prostate cancer, and acute myeloid leukemia. Studies have shown that in lung cancer, after aberrant splicing, the expression of BCL2L1, MDM2, MDM4, NUMB, and MET may play an important role in tumorigenesis, which may be due to the effects on cell apoptosis, cell proliferation, and intercellular cohesion-related pathways.<sup>18-20</sup> Coomer et al also mentioned the role of SFs in tumorigenesis when they summarized the impact of AS events in lung cancer.<sup>21</sup> SFs, including QKI, RBM4, RBM5, RBM6, RBM10, and SRSF1, were involved in the development of lung cancer, as SFs play a regulatory role in splicing. Paschalis et al. summarized the effect of AS events in prostate cancer, stating that they can influence tumor development and metastasis by acting on androgen receptors, as well as the effect of hormone therapy and radiotherapy.<sup>22</sup> Anande et al reported that RNA splicing alterations may be closely related to the prognosis of AML patients, and therefore, serve as a biomarker of this disease.<sup>23</sup> In glioma, the study of AS events is still in the initial stage, but the relationship between AS events and the prognosis of glioma patients has been analyzed.<sup>13</sup> Chai et al. proposed a new classification of LGG based on AS events.<sup>24</sup>

The above studies demonstrate that AS events play a crucial role in tumors. Therefore, translating the results of these studies into clinical work has become the purpose of all researchers. One of the highlights of our study is the construction of a prognostic model with a nomogram that uses both clinical factors and AS events, making this model more clinically practical. In recent years, the nomogram of the prognostic model has been developed in a variable sub-field of glioma. Gittleman et al. screened the several factors related to patient survival through the analysis of LGG patients in the TCGA database and The Ohio Brain Tumor Study (OBTS) database.<sup>25</sup> These factors included tumor grade, age of diagnosis, KPS score, and IDH1 mutation status. The authors further analyzed the above factors and constructed a prognostic prediction model using a nomogram. Similarly, Zhao et al. also constructed a prognostic prediction model based on clinical data downloaded from the Surveillance, Epidemiology, and End Result (SEER) database.<sup>26</sup> New genes related to prognosis have been found based on database analysis, and a prognostic prediction model has been constructed.<sup>27,28</sup> With the development of radiology in recent years, research on establishing nomogram models in glioma based on radiomics data has also gradually increased, and this includes the preoperative prediction of tumor grade and the prediction of molecular biomarkers.<sup>29,30</sup> Based on clinical factors and AS events, an epigenetic factor closely related to the genesis, progression, metastasis, and drug resistance of tumors, this study established a nomogram of the prognostic prediction model for LGG patients, demonstrating the value of basic science research and the importance of translational medicine.

The RNA-Seq data of 510 LGG patients in the TCGA database were analyzed. Univariate Cox analysis was used to screen 2185 AS events related to prognosis, and a prognostic risk score model based on AS events was further established based on multivariate Cox regression and LASSO regression. Although all seven AS Signatures can be related to the prognosis of LGG patients separately, we found that when the seven AS events are considered as a whole, the integrated AS signature could be an independent prognostic factor. Therefore, we used the integrated AS signature and clinical information of the patients to establish a nomogram prognosis prediction model. To further clarify the role of AS Events, the regulatory network between AS events and SFs was analyzed. The prognosis of LGG patients is known

to be significantly different, and few remarkable biomarkers are available except for IDH and 1p/19q. This study focused on the AS events that have not been well explored in LGG. A large amount of information from 510 samples was enrolled in the TCGA database for analysis and further verification. The integrated AS signature was significantly correlated with prognosis and could be used as an important biomarker. Based on this, a clinically applicable prediction model was established, which is not only beneficial for the diagnosis and treatment of LGG patients, but also a direction for further exploration.

This study has some limitations. First, as mentioned above, the effect of AS events on RNA is regulated by SFs.<sup>8</sup> Although this study analyzed the regulatory network between AS events and SFs and found four AS-SF pairs with the strongest correlation, we did not perform an additional in-depth analysis of their relationship, which may warrant future research. The second disadvantage is that this study included only well-known clinical and pathological indicators in LGG patients without exploring other new potential biomarkers due to the consideration of clinical practicability. To make the model more accurate, future studies should be conducted to explore new prognostic factors. Additionally, surgical factors and imaging data could also be considered for analysis. The above two deficiencies can be further studied in the future.

In conclusion, the value of AS events, as a factor closely related to tumorigenesis and development, is fully reflected in glioma. In this study, AS events and clinical data of LGG patients in the TCGA database were analyzed, and prognostic factors were screened. Then, a nomogram of the prognostic prediction model was established based on the factors that can predict the 1-, 3-, and 5-year survival of LGG patients. The establishment of this model will help clinical workers better evaluate the prognosis of patients and will play a role in guiding follow-up and treatment processes.

## CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

## AUTHOR CONTRIBUTIONS

All authors designed and conducted this article. All authors read and approved the final manuscript. Notably, Yaning Wang and Zihao Wang equally share the first authorship; Yu Wang and Wenbin Ma equally share the corresponding authorship.

## DATA AVAILABILITY STATEMENT

The data are available upon request of the corresponding authors.

## ORCID

Yaning Wang  <https://orcid.org/0000-0002-0999-0499>

Yu Wang  <https://orcid.org/0000-0002-7175-4649>

## REFERENCES

1. Brat DJ, Verhaak RG, Aldape KD, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med*. 2015;372(26):2481-2498.
2. van den Bent MJ, Wefel JS, Schiff D, et al. Response assessment in neuro-oncology (a report of the RANO group): assessment of outcome in trials of diffuse low-grade gliomas. *Lancet Oncol*. 2011;12(6):583-593.
3. Chai R-C, Wu F, Wang Q-X, et al. m(6)A RNA methylation regulators contribute to malignant progression and have clinical prognostic impact in gliomas. *Aging (Albany NY)*. 2019;11(4):1204-1225.
4. Chai R-C, Wang N, Chang Y-Z, et al. Systematically profiling the expression of eIF3 subunits in glioma reveals the expression of eIF3i has prognostic value in IDH-mutant lower grade glioma. *Cancer Cell Int*. 2019;19:155.
5. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131(6):803-820.
6. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008;456(7221):470-476.
7. Lee SC, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. *Nat Med*. 2016;22(9):976-986.
8. Kedzierska H, Piekietko-Witkowska A. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. *Cancer Lett*. 2017;396:53-65.
9. Yanagisawa M, Huveltdt D, Kreinest P, et al. A p120 catenin isoform switch affects Rho activity, induces tumor cell invasion, and predicts metastatic disease. *J Biol Chem*. 2008;283(26):18344-18354.
10. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol*. 2007;14(3):185-193.
11. Marzese DM, Manughian-Peter AO, Orozco JIJ, Hoon DSB. Alternative splicing and cancer metastasis: prognostic and therapeutic applications. *Clin Exp Metastasis*. 2018;35(5-6):393-402.
12. Siegfried Z, Karni R. The role of alternative splicing in cancer drug resistance. *Curr Opin Genet Dev*. 2018;48:16-21.
13. Li Y, Ren Z, Peng Y, et al. Classification of glioma based on prognostic alternative splicing. *BMC Med Genomics*. 2019;12(1):165.
14. Ryan MC, Cleland J, Kim R, Wong WC, Weinstein JN. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics*. 2012;28(18):2385-2387.
15. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.
16. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA*. 2017;318(14):1377-1384.
17. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53.
18. Yin XM, Oltvai ZN, Korsmeyer SJ. BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax. *Nature*. 1994;369(6478):321-323.
19. Weng M-W, Lai J-C, Hsu C-P, et al. Alternative splicing of MDM2 mRNA in lung carcinomas and lung cell lines. *Environ Mol Mutagen*. 2005;46(1):1-11.

20. Kong-Beltran M, Seshagiri S, Zha J, et al. Somatic mutations lead to an oncogenic deletion of met in lung cancer. *Cancer Res.* 2006;66(1):283-289.
21. Coomer AO, Black F, Greystoke A, Munkley J, Elliott DJ. Alternative splicing in lung cancer. *Biochim Biophys Acta Gene Regul Mech.* 2019;1862(11-12):194388.
22. Paschalis A, Sharp A, Welti JC, et al. Alternative splicing in prostate cancer. *Nat Rev Clin Oncol.* 2018;15(11):663-675.
23. Anande G, Deshpande NP, Mareschal S, et al. RNA splicing alterations induce a cellular stress response associated with poor prognosis in Acute Myeloid Leukemia. *Clin Cancer Res.* 2020.
24. Chai R-C, Li Y-M, Zhang K-N, et al. RNA processing genes characterize RNA splicing and further stratify lower-grade glioma. *JCI Insight.* 2019;5(17):e130591
25. Gittleman H, Sloan AE, Barnholtz-Sloan JS. An independently validated survival nomogram for lower grade glioma. *Neuro Oncol.* 2020;22(5):665-674.
26. Zhao YY, Chen SH, Hao Z, Zhu HX, Xing ZL, Li MH. A nomogram for predicting individual prognosis of patients with low-grade glioma. *World Neurosurg.* 2019;130:e605-e612.
27. Wang L, Yan Z, He X, Zhang C, Yu H, Lu Q. A 5-gene prognostic nomogram predicting survival probability of glioblastoma patients. *Brain Behav.* 2019;9(4):e01258.
28. Wang Y, Liu X, Guan G, Xiao Z, Zhao W, Zhuang M. Identification of a five-pseudogene signature for predicting survival and its ceRNA network in glioma. *Front Oncol.* 2019;9:1059.
29. Lao J, Chen Y, Li Z-C, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep.* 2017;7(1):10353.
30. Wang Q, Li Q, Mi R, et al. Radiomics nomogram building from multiparametric MRI to predict grade in patients with glioma: a cohort study. *J Magn Reson Imaging.* 2019;49(3):825-833.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Wang Y, Wang Z, Zhao B, Chen W, Wang Y, Ma W. Development of a nomogram for prognostic prediction of lower-grade glioma based on alternative splicing signatures. *Cancer Med.* 2020;9:9266–9281. <https://doi.org/10.1002/cam4.3530>