

A new signature based on alternative polyadenylation for prognostic prediction and therapeutic responses in low-grade glioma

Yuzhi Wang^{1,*}, Yunfei Xu^{2,*}, Yi Zhang³

¹Department of Laboratory Medicine, People's Hospital of Deyang City, Deyang 618000, Sichuan, China

²Department of Laboratory Medicine, Chengdu Women's and Children's Central Hospital, Chengdu 610031, Sichuan, China

³Department of Blood Transfusion, People's Hospital of Deyang City, Deyang 618000, Sichuan, China

*Co-first authors

Correspondence to: Yi Zhang; email: 472189926@qq.com, <https://orcid.org/0000-0002-0010-1915>

Keywords: low-grade glioma, alternative polyadenylation, prognosis signature, therapeutic responses, TCGA

Received: October 27, 2021

Accepted: January 12, 2022

Published: January 18, 2022

Copyright: © 2022 Wang 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

Background: Evidence from research supports the significant role of alternative polyadenylation (APA) in the development of cancer. The aim of this study is to explore the prognostic and therapeutic value of APA events for patients with low-grade gliomas (LGG).

Methods: The gene expression and APA profiles of patients with low-grade gliomas were obtained from The Cancer Genome Atlas database. All patients were sorted randomly into training and test sets. The prognostic-associated events of alternative splicing were screened by univariate Cox regression. Subsequently, Least Absolute Shrinkage and Selection Operator and multivariate Cox analysis were performed to construct a prognostic signature. The patients were sorted into the high and low-risk groups based on their median risk score. Bioinformatics methods were used to identify genetic variation, pathway activation, immune heterogeneity, and drug response differences between the two groups.

Results: A prognostic signature was constructed shown to be capable of accurately predicting prognosis of patients with LGG. Notable variations were observed in the tumor mutation burden and copy number variations between the high-risk and low-risk patients. Besides, the high-risk group had enhanced immune cell abundance and immune checkpoint gene expression. In terms of drug response, we further found that the patients of high-risk group were more sensitive to immunotherapy, but chemotherapy was suggestively more appropriate for the low-risk group patients.

Conclusion: Our findings give new insights and methods related to prognosis prediction and treatment methods for LGG patients, and expand the understanding regarding the role of alternative splicing in LGG.

INTRODUCTION

Glioma is the most common primary intracranial tumor in the adult central nervous system with high recurrence and mortality rates [1]. The age-calibrated incidence of gliomas ranges from 4.67‰ to 5.73‰ [2]. As per the classification system of the World

Health Organization (WHO), gliomas can be sorted into grades I–IV, of which grades II and III are defined as low-grade gliomas (LGG) [3]. Although LGG is less malignant as compared to GBM, seemingly the recurrence of the tumor and malignant progression cannot be evaded completely even after standard treatment such as surgical resection, radiotherapy, and

chemotherapy is received by the patient [3]. Because of tumor heterogeneity, the prognosis of LGG patients varies widely, with survival periods ranging from 1 to 15 years [4]. As a result of the heterogeneity of LGG patients, there is a need for the development of effective biomarkers that can be used to stratify predictions based on patient prognostic risk, thereby facilitating the development of precise treatment regimens.

The alternative polyadenylation (APA) is a very common RNA processing mechanism that occurs during pre-mRNA maturation. Multiple mature mRNA subtypes can be generated from the same pre-mRNA by selecting different polyadenylation signaling sites (PAS) in the 3'-UTR [5]. Whole-genome deep sequencing indicates that at least 70% of multiple transcripts coded by human genes are derived as a result of APA [6]. The APA plays a significant function in controlling the stability of mRNA, its localization and translation, protein-coding, and the localization of protein [7]. According to recent studies, the change of APA is closely associated with the occurrence and development of different tumor types, for instance, low expression levels of PCF11 in neuroblastoma is related to extensive APA in the transcriptome, good prognosis of the patient, and spontaneous tumor regression, while PCF11 knockdown causes abnormal neural differentiation [8]. During glioblastoma, the down-regulation of CFIm25 results in 3'UTR shortening and the up-regulation of oncogenes such as CCND1 and Pak1 which increases tumorigenicity and enhances tumor size [9, 10]. Though the biological significance of APA in tumors is widely accepted, its prognostic value and biological function in LGG are still largely unknown.

In this research, a systematic analysis of APA events in LGG patients was performed using The Cancer Genome Atlas (TCGA) database and we identified numerous overall survival (OS) related APA events. The potential biological functions of these OS-related APA events were explored in detail. Subsequently, a signature based on multiple prognostic APA events was constructed to predict the prognosis of LGG patients. Furthermore, we analyzed the correlation between the signature and tumor immunity, tumor mutation burden (TMB), and copy number variations (CNVs). The value of this signature in predicting the response of LGG patients to various treatments was also evaluated. At the end of this study, a core regulator (CR)-APA network was constructed to disclose the underlying mechanism through which the events of APA affect LGG prognosis. The findings of this research may help in understanding mechanisms involved in the occurrence and development of LGG.

MATERIALS AND METHODS

Data collection and preprocessing

The expression profiles, mutations, CNVs, and clinical information of LGG patients were downloaded from the TCGA database. Moreover, data for APA events were obtained from the UCSC Xena database (<https://xena.ucsc.edu>). Patients were then randomly sorted into two sets namely a training set and a test set in the ratio of 1:1 by the software. The percentage of distal polyA site usage index (PDUI) value is an intuitive ratio from 0 to 1 that was used for the quantification of events related to APA. A set of strict filtering criteria (Percentage of samples with PDUI value $\geq 75\%$, mean PDUI value ≥ 0.05 and standard deviation PDUI value ≥ 0.01) were established for ensuring the reliability of the subsequent analysis of APA events. Afterward, for the completion of missing data regarding APA, we used the k-nearest neighbor algorithm.

Identification of prognostic APA events and functional enrichment analysis

For the determination of survival-related APA events, univariate Cox regression analysis was performed to evaluate the relation between APA events and the overall survival period of patients with LGG. The APA events with P -value < 0.05 were chosen as survival-related APA events. The web tool Metascape (<https://metascape.org/>) was utilized for analyzing the functional enrichment of parental genes for survival-related APA events. Terms having a calibrated P -value < 0.05 were taken as substantially enriched. Bar plots were used for visualizing key terms in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) function enrichment analysis (including biological process (BP), cellular component (CC), and molecular function (MF)).

Construction of prognosis signature for patients with LGG based on APA events

At first, the least absolute shrinkage and selection operator (LASSO) analysis was performed on the top 20 most prognostic-related APA events to further screen key APA events and avoid overfitting of subsequent models. Afterward, a stepwise multivariate Cox regression analysis was carried out for the development of a prognostic signature for APA events. The risk value of each patient was measured as per the following formula: $\beta_1 \times \text{Exp1} + \beta_2 \times \text{EXP2} + \beta_i \times \text{EXPi}$, where β was the coefficient value and exp was the PDUI value of APA time. The risk value was an indicator that measured the prognostic risk of each LGG patient. All patients were sorted into two groups: the high-risk and

low-risk groups according to their median risk values. The Kaplan-Meier survival analysis was carried out for the verification of survival differences among the high-risk and low-risk groups. The time-dependent receiver operating characteristic (ROC) curves were drawn for the evaluation of the predictive ability of APA signature.

Somatic mutation and copy number analysis

This study used the “maftools” package (version 2.10) for calculating and visualizing the mutation data. TMB was identified as the total number of somatic non-synonymous mutations in the coding regions. For each sample, mutation detection was done using the preprocessed and annotated MAF data files generated by the varscan platform for the calculation of the tumor mutation burden. The GISTI is a tool that is widely used for the identification of genes targeted for somatic copy number changes that trigger the growth of cancer. In this research, the GISTIC 2.0 software was used to identify regions with significant gene amplification or deletion in the CNV data of LGG [11]. The parameter thresholds were defined as amplifying or missing length >0.1 and *P*-value <0.05.

The evaluation of immune infiltration and immunotherapy response

Based on the expression profile information, the estimation of stromal and immune cells in malignant tumours using expression data (ESTIMATE) algorithm was used to generate matrix and immune scores to estimate the level of infiltrating matrix, immune cells, and tumor purity in LGG tissues [12]. In addition, the “GSVA” package (version 1.42) of Single-sample gene set enrichment analysis (ssGSEA) algorithm was used to quantify the enrichment scores of 16 immune cells and 13 immune functions for every sample of LGG. 14 genes have previously been reported as hub targets for immune checkpoint inhibitors [13]. Differences in the therapeutic efficacy of immune checkpoint inhibitors in malignancies are associated with differences in immune checkpoint gene expression. Therefore, the correlation between the expression levels of 14 immune checkpoint genes and our signature was analyzed in depth. Tumor immune dysfunction and exclusion (TIDE) is an algorithm that predicts response to immunotherapy based on mimicking tumor immune evasion mechanisms [14]. The algorithm was performed to preliminarily explore the possibility of each sample responding to immunotherapy.

Clinical drug response prediction

Based on the Genomics of Drugs Sensitivity in Cancer (GDSC) cell line dataset, the R package “pRRophetic”

(version 0.50) was used to predict the sensitivity of patients to signature-targeted chemotherapy agents. The R package can be evaluated by ridge regression for IC50 of included drugs, and the prediction accuracy can be observed by 10-fold cross-validation on the basis of the GDSC training set [15, 16].

Establishment of the nomogram

The clinicopathological parameters of LGG patients were added in univariate and multivariate Cox regression analyses for the verification of the independence of the risk scores based on the survival-related APA signature. Afterward, a nomogram signature was constructed using all independent prognostic factors to develop a scoring system to evaluate the OS of patients at 1-year, 3-year, and 5-year. To demonstrate the effectiveness of the system, the C index curve, calibration curves, and time ROC curves were used for the evaluation of the recognition performance of the system. The DCA curves were used to assess the clinical applicability of the scoring system.

Construction of correlation network between CR genes and APA events

A total of 22 CRs genes were obtained from previous studies [17] and CRs gene expression profiles were collected from the TCGA database. The Spearman correlation method was performed for the calculation of the correlation among the PDUI values and SF expression levels of survival-related APA events. *P*-value <0.001 and absolute values of correlation coefficient >0.5 were considered as thresholds. Lastly, to construct and draw the interaction network between APA and SF, the Cytoscape software was used in this study.

Statistical analysis

The R software (version 4.0.5) was used for all statistical analyses, and *P*-value <0.05 on both sides was considered statistically significant. The Kaplan-Meier survival analysis was done by log-rank test. The independent sample *T*-test and Wilcoxon test were performed to compare the two groups. The language was polished by Bullet Edits Services (<http://www.bulletedits.cn>).

RESULTS

Overview of APA events in patients with LGG

The general analysis flow chart of this study was illustrated in Supplementary Figure 1. 511 LGG patients in total were enrolled in the current research, and their clinical data were compiled in Table 1. According to the

Table 1. Clinical characteristics of the TCGA cohort.

Characteristics	Groups	Number (percentage)
Age	<60	442 (86%)
	>60	69 (14%)
Gender	Male	283 (55%)
	Female	228 (45%)
Grade	G2	246 (48%)
	G3	264 (52%)
Histology	Astrocytoma	193 (38%)
	Oligodendroglioma	187 (36%)
	Mixed glioma	131 (26%)

above screening criteria, 6,574 APA events were identified from 6,214 genes, indicating that one gene can produce one or more APA events.

Survival-related APA events and the analysis of functional enrichment

For further investigation of the impact of APA on prognosis, we randomized 511 LGG patients into either a training set ($n = 256$) or a test set ($n = 255$). After the analysis of all APA events in the training set by univariate Cox regression, a total of 1,213 prognostic-related APA events were screened out of 1,164 genes. These APA events were shown in the volcanic map (Figure 1A). Parental genes refer to genes involved in APA events, and their functions and pathways may reflect the potential role of APA events. To elucidate the potential biological functions of these parent genes related to the survival-related APA events in LGG, the

GO enrichment and KEGG pathway analysis were performed on these genes. The GO and KEGG analyses highlighted that these genes were enriched in large quantities in RNA processing, protein synthesis, cell metabolism, and other processes, revealing the correlation between APA and these basic biological processes (Figure 1B–E).

Construction of prognostic APA signature

The LASSO regression analysis was performed for further screening of significant APA events from the top 20 most important prognostic APA events (Figure 2), which were considered as candidate genes for the construction of a prediction signature using the multivariate Cox regression analysis. In the end, a prognostic signature was constructed that consisted of four APA events. The coefficients of the four events were showed in Supplementary Table 1. The risk value

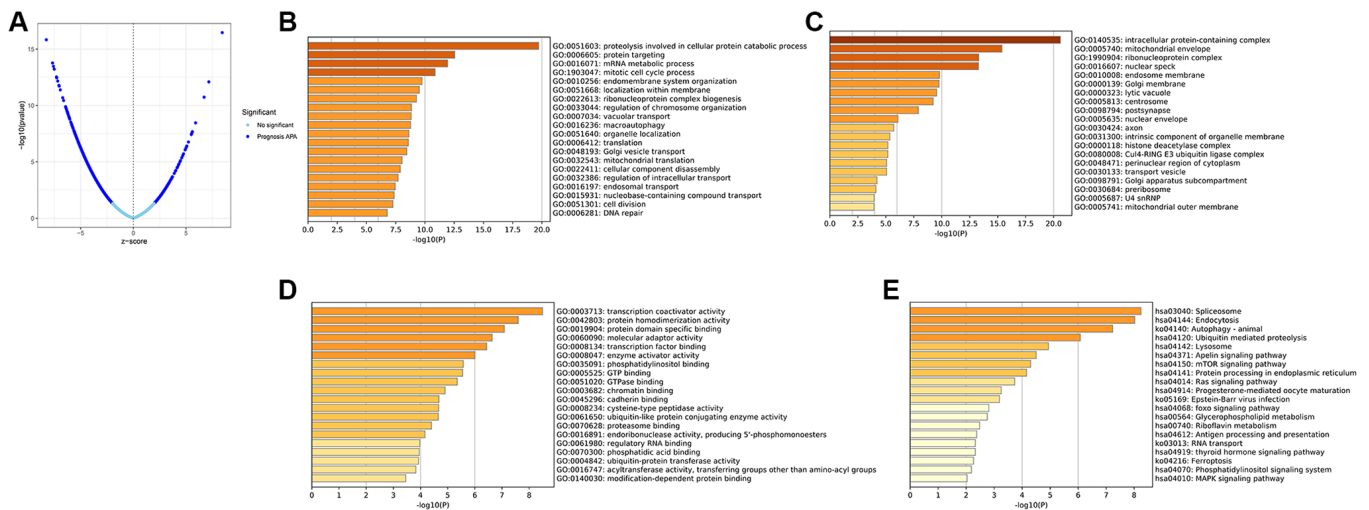


Figure 1. Identification of the prognosis-related APA events and enrichment analysis of the corresponding genes of prognosis-related APA events. (A) Volcano plot of prognosis-related APA events. The top 20 significant enrichment terms in BP (B), CC (C), and MF (D) in the GO analysis. (E) KEGG pathway analysis.

for individual LGG patients was measured by following the formula, and all patients were classified into two groups called the high-risk and low-risk groups according to the median risk score (Figure 3A–3C). According to the Kaplan-Meier analysis these prognostic models successfully stratified patients with

different results, it was observed that the patients that belonged to the high-risk group had a substantially shorter OS as compared to those in the low-risk group (Figure 4A). For the additional evaluation of the predictive power of the signature in 1-year, 3-year, and 5-year, the ROC analysis was conducted. The outcomes

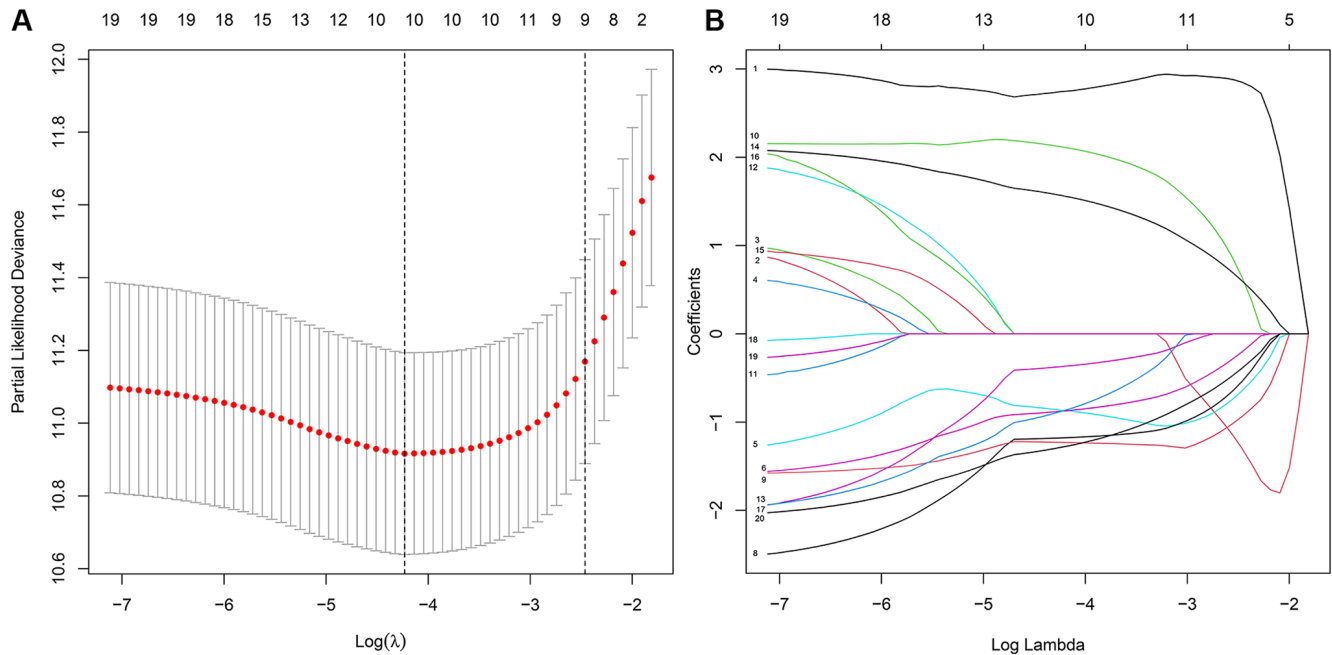


Figure 2. The key features were identified using LASSO regression. (A) Selection of the optimal parameter (lambda) via 5 times cross-validation. **(B)** LASSO coefficient profiles of the top 20 prognosis-related APA events.

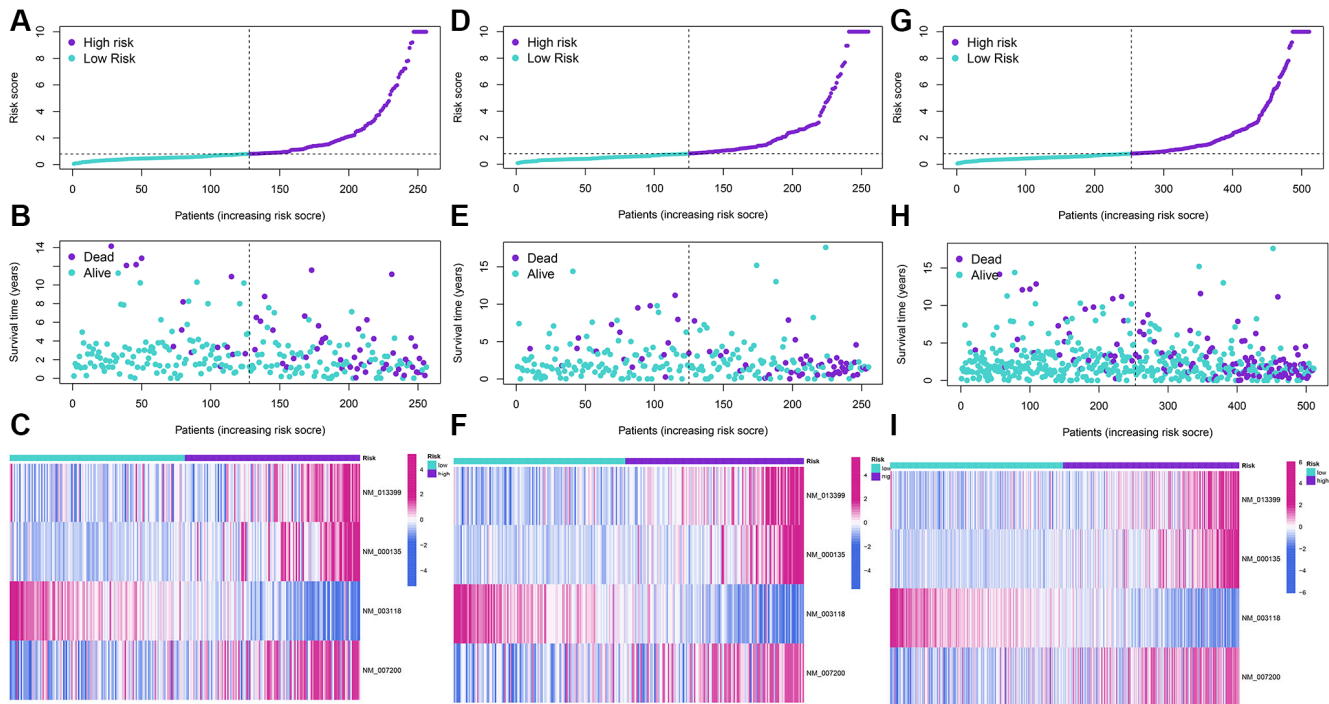


Figure 3. Development and validation of a four-APA-based prognostic signature. The risk score distribution, gene expression, and LGG patients' survival status in the training (A–C), test (D–F), and whole sets (G–I) based on the signature.

indicated that the signature had good prognostic performance in different years, with AUC values ranging from 0.817 to 0.956 (Figure 4D). Moreover, this study also verified the predictive power and accuracy of the signature in the training set and the whole set (Figure 3D–3I). The Kaplan-Meier survival analysis confirmed a substantial decline in the OS in high-risk groups of both datasets (Figure 4B, 4C). The AUC values of test set and whole set at 1-year, 3-year, and 5-year were 0.826, 0.830, 0.731 and 0.873, 0.839, 0.772, respectively (Figure 4E, 4F). According to the PCA and t-SNE analyses, patients in the two risk groups were distributed in two separate directions (Figure 4G, 4H). To confirm the general applicability of this signature, patients were subdivided by age, sex, grade, and pathological type. Considering the different stratified analysis results, the OS duration of LGG

patients in the low-risk group was significantly longer in comparison to that in the high-risk group (Figure 5). As indicated by these findings, the signature can accurately and effectively predict the prognosis of patients with LGG.

TMB and CNVs analysis

We analyzed and visualized somatic mutation data in LGG patients by distinguishing between the high-risk and low-risk groups. The top 10 drive genes with the highest variation frequency in the two risk groups were shown in Figure 6A, 6B. TMB of the two risk groups were calculated based on somatic mutation data, and the outcomes indicated that the TMB of the high-risk group was greatly enhanced as compared to that of the low-risk group (Figure 6C). We then identified whether

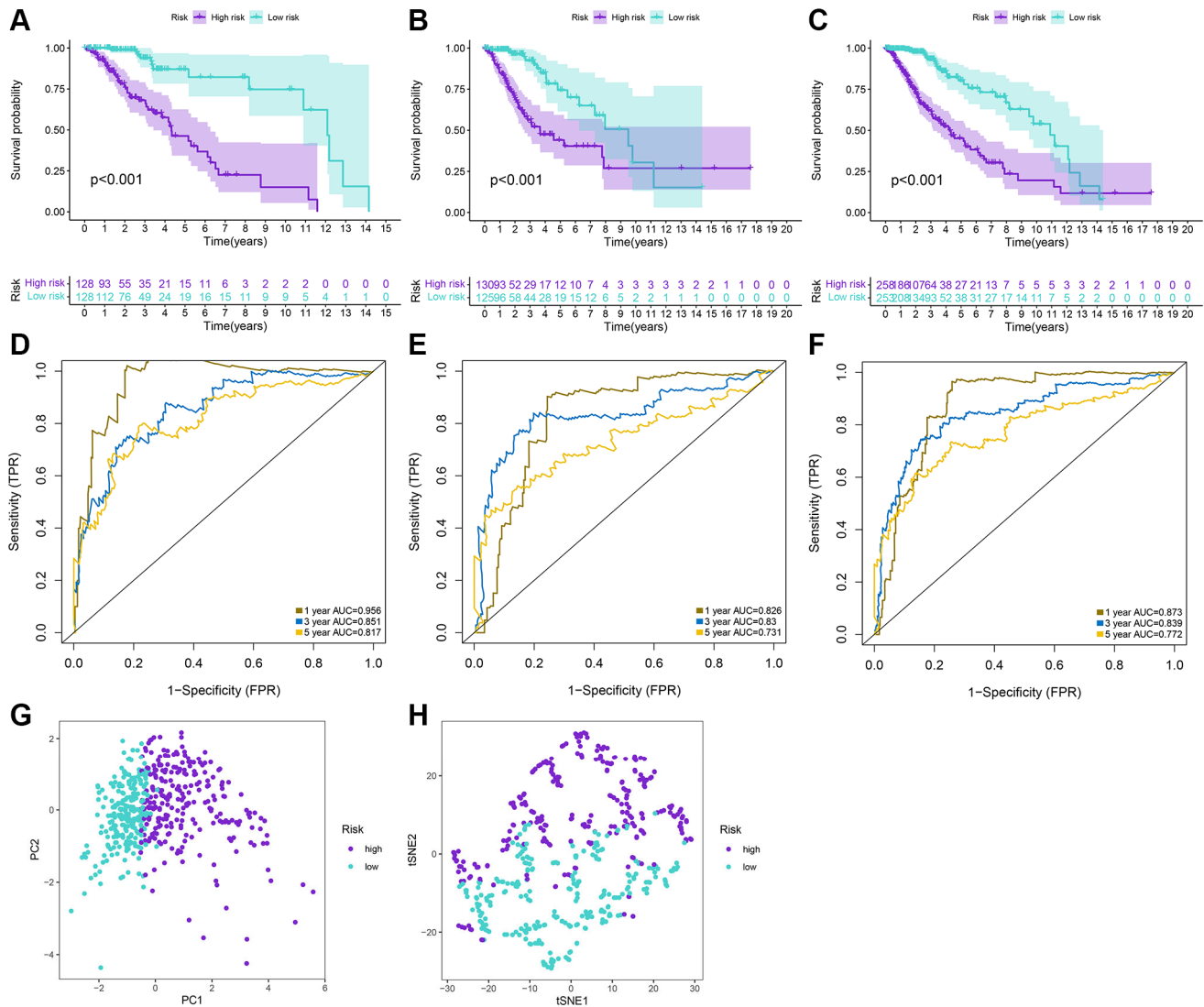


Figure 4. Prediction performances of the signature for LGG patients. (A–C) Kaplan-Meier analysis of high-risk and low-risk patients stratified by the median risk score in the training, test and whole sets. (D–F) Time-dependent ROC curves for 1-year, 3-year, and 5-year OS predictions by the signature in the training, test, and whole sets. (G, H) PCA and tSNE plots for LGG patients based on the risk groups.

TMB was an independent biomarker for LGG patients. LGG cohorts are split into high-TMB and low-TMB groups as per the median TMB value. According to the findings, the TMB cannot be used alone for predicting patient outcomes (Figure 6D). However, when the risk score and TMB were combined they could effectively predict LGG patient outcomes (Figure 6E). We also performed correlation analysis to explore the association of signature with tumor stem cells, and the results suggested that the risk score was negatively correlated with DNAss (Figure 6F). Additionally, we analyzed CNVs data using the GISTIC algorithm to find the gene regions with apparent amplification or deletion. The distribution of copy number changes in both the high and low-risk groups was demonstrated in Figure 6G. Overall, there were significant differences in the gene and frequency of copy number changes between the two groups. The frequency of gene changes, deletion, and gain was elevated in the high-risk group as compared to that in the low-risk group (Figure 6H, 6I).

Analysis of immune function and immunotherapy

The materials and methods mentioned in the algorithm were used to assess the immune status of each LGG patient, as shown in the heatmap (Figure 7A). Moreover, the Wilcoxon test was performed for the comparison of

the differences in individual cell markers among the high-risk and low-risk groups. The results indicated that stromal, immune, and ESTIMATE scores were substantially enhanced in the high-risk group as compared to those in the low-risk group, but tumor purity was lower in the low-risk group (Figure 7B–7E). Concerning the immune cells and functions, 11 types of immune cells and 13 immune function types were present in greater abundance in the high-risk group (Figure 7F, 7G). Additionally, 14 immune checkpoint genes between the two risk groups were compared. The outcomes indicated that 12 genes were highly expressed in the high-risk group (Figure 7H). We observed the TIDE scores to be closely correlated with immune checkpoint blocker response. As shown in Figure 7I, TIDE scores of the low-risk group patients were substantially enhanced as compared to those in the high-risk group, suggesting that LGG patients having high-risk scores showed more sensitivity to immune checkpoint blockade (ICB) treatment.

Chemotherapeutic drug sensitivity

Antitumor drug therapy is the basic treatment for patients with LGG. The response of high and low-risk LGG patients to anticancer drugs was evaluated in this study by using the 10-fold cross-validation prediction signature through the GDSC database. Out of the 137

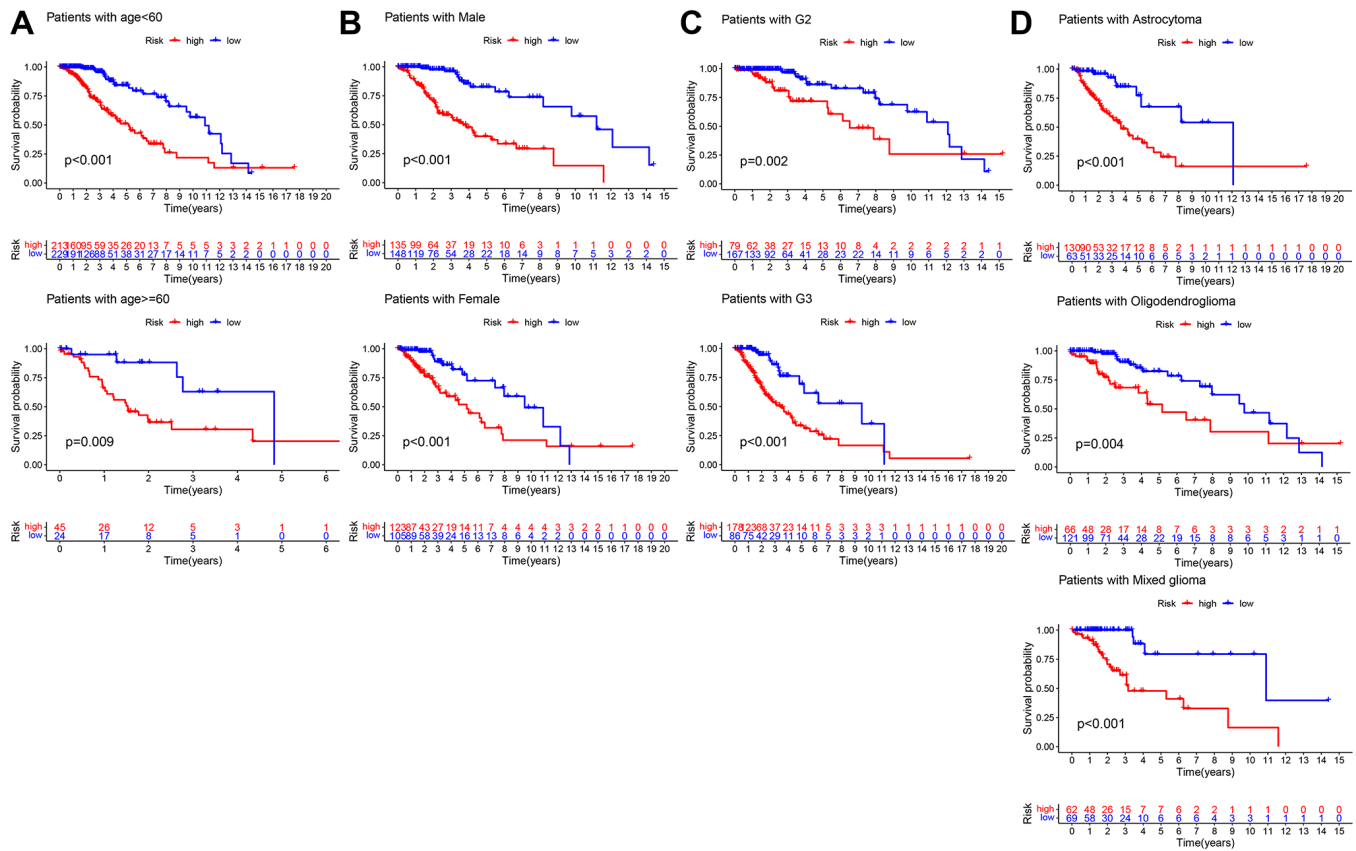


Figure 5. Kaplan–Meier survival curves of LGG patients in different clinical subgroups. (A) age. (B) Gender. (C) Grade. (D) Histology.

drugs, 29 drugs had IC50 values that differed among the two groups, suggesting that low-risk patients showed more sensitivity to these 29 drugs (Supplementary Figure 2). 16 drugs in particular, including Gefitinib, Lenalidomide, and Axitinib, had great potential for the treatment of LGG.

Construction of prognostic nomogram

Initially, the correlation between clinicopathological indicators and the signature was explored. The bar chart indicated prominent variations in gender, grade, and

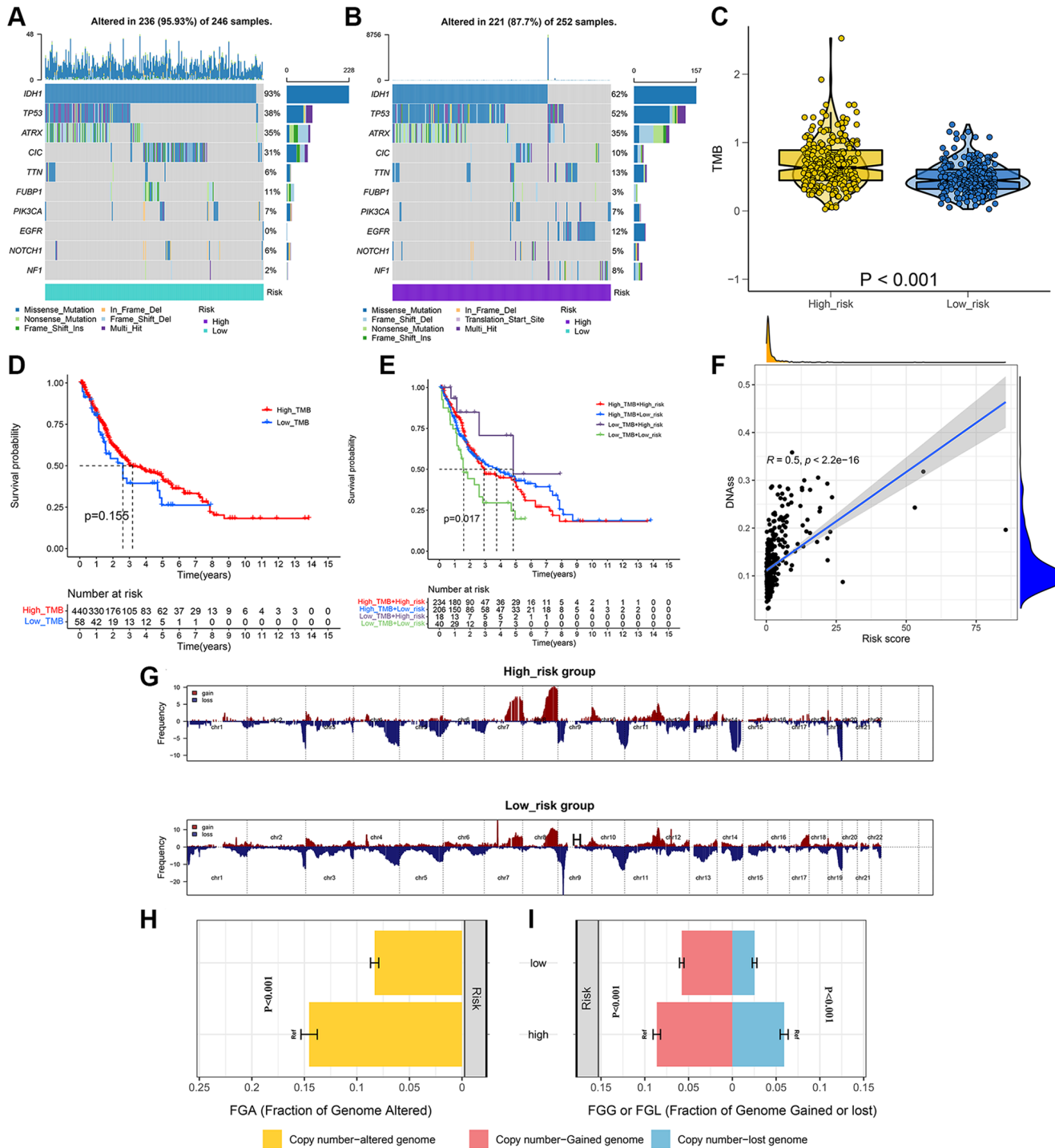


Figure 6. Integrated comparisons of somatic mutation and CNVs between high-risk and low-risk groups in the whole set. (A, B) Waterfall plots showing the mutation information of top 10 genes with the highest mutation frequency in two groups. **(C)** Distribution of TMB in two groups. **(D)** Survival curves for the OS of the high-TMB and low-TMB groups. **(E)** Survival curves for patients stratified by both TMB and signature. **(F)** Relationship between risk score and DNAss. **(G)** Gene fragments profiles with amplification (red) and deletion (blue) among the two groups. **(H, I)** Comparison of the fraction of the genome altered, lost, and gained between the two groups.

pathological type among the high and low-risk groups (Figure 8A). Subsequently, a nomogram was created on the basis of outcomes of univariate and multivariate Cox regression (Figure 8B, 8C), including pathological

type, sex, age, and risk score (Figure 9A). The OS was calculated for all patients and was predicted at 1-year, 3-year, and 5-year. It can be seen from the calibration curve that the results predicted by the nomogram

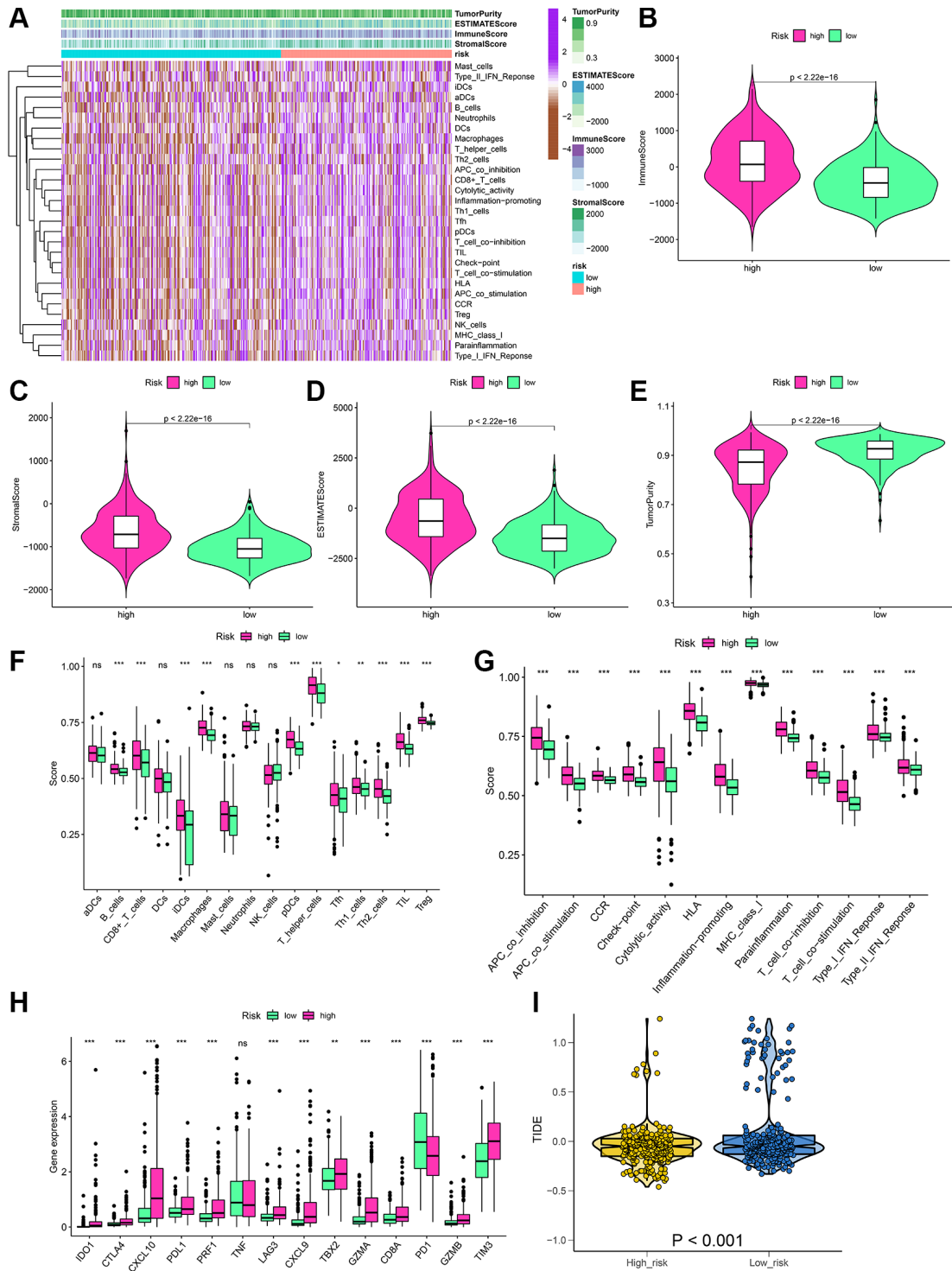


Figure 7. Estimation of the immune status and response to immunotherapy based on the signature in the high-risk and low-risk groups for the whole set. (A) Heatmap of the immune scores, stromal scores, tumor purity, ESTIMATE scores and immune-infiltrating cells in the two groups. **(B–E)** Violin plots for the immune scores, stromal scores, ESTIMATE scores, and tumor purity. **(F–H)** Boxplots of immune cells scores, immune-related functions scores and immune checkpoints expression. **(I)** Prediction of immunotherapy response according to TIDE score. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ns: no significance.

supported the actual outcomes (Figure 9B). The results of the time ROC curve and C index curve showed that AUC and C index at different times were the largest in the nomogram (Figure 9C, 9D). Furthermore, as compared with the risk score signature, the nomogram had improved the net benefit of clinical patients even more (Figure 9E–9G).

Potential regulatory network of APA

For a detailed analysis of the potential regulatory mechanism of survival-related APA events in LGG cohorts, an interaction network of APA events and key regulators were designed. The Spearman test was used to observe the correlation between PDUI values and CRs gene expression levels of OS-related APA events. Significant relationship pairs with the correlation coefficient >0.5 and P -value <0.001 were selected to create the correlation network. As demonstrated in Figure 10, the expression levels of 17 CRs (shown as the blue triangles) were greatly correlated with 180 survival-related APA events, including the 164 APA events with good prognosis (shown as the purple triangles) and 16 APA events with poor prognosis (shown as the yellow triangles). Interestingly, it was found that the ratios of SFs increased (shown as the red

lines) and decreased (shown as the green lines) were 1:1 for both APA events with poor survival prognosis and the APA events with good survival prognosis.

DISCUSSION

Heterogeneity in patients is a key contributing factor in the poor prognosis of glioma, and multifaceted evaluation can help in improving the precision of prognosis prediction for such patients [18]. Therefore, the latest version of gliomas' classification by the WHO incorporates molecular features into the classification criteria, thereby improving the homogeneity of clinical outcomes in patients with the same subtype [19]. Nevertheless, as one of the histological subtypes of glioma, patients with LGG vary greatly in survival and lack effective prognostic markers. APA is a common post-transcriptional regulatory mechanism in eukaryotic organisms. The differential expression of APA can be caused by affecting the stability of transcripts, their output to the cytoplasm, and their translation efficiency [5]. Studies have found that extensive APA occurs in the pathophysiological process of numerous diseases, including cancer. In such diseases, APA events are emerging as potential biomarkers that can be used in clinical practice. The majority of the dysfunctional APA

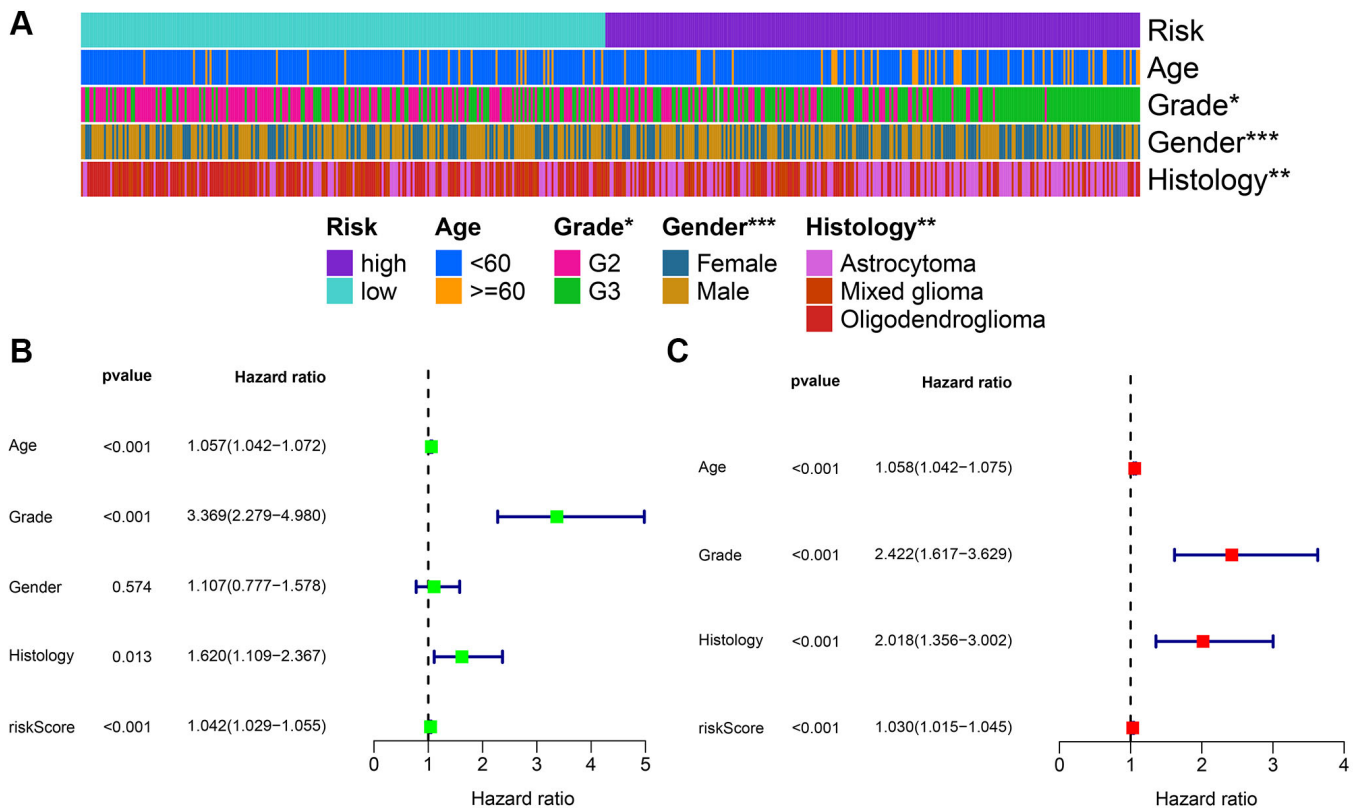


Figure 8. Combined analysis of signature and clinicopathological characteristics in the whole set. (A) Heatmap presents the distribution of clinical feature and corresponding risk score. (B) Univariate Cox regression analysis of clinical characteristics and signature. (C) Multivariate Cox regression analysis of clinical characteristics and signature. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

events result in the production of transcript isomers with variable lengths of 3'UTRs. These are often associated with various clinical traits. These APA events do not depend on the commonly used molecular data, such as gene expression and somatic mutations [20] and they are associated with disease prognosis, recurrence, tumor subtypes, and staging of multiple cancers [21–24]. However, there are few systematic studies on the function of APA events in LGG.

In the current research, systematic identification and analysis of survival-related APA events was performed in LGG patients from the TCGA database and 1,213 survival-related APA events were discovered in 1,164 genes. By performing the GO and KEGG analysis of

parental genes of these prognostic APA events, it was observed that these events were largely involved in RNA processing, protein synthesis, cell metabolism, and other processes, such as proteolysis involved in cellular protein catabolic process, mRNA metabolic process, and endocytosis. Meanwhile, we constructed a correlation network of CRs-APA to reveal the potential regulatory mechanism of prognostic APA events. For identifying the prognostic importance of APA events, a prognostic prediction signature was constructed for LGG based on the screening of the top 20 survival-related APA events. Patients were sorted into two groups called the low-risk and high-risk groups on the basis of the median risk score. The Kaplan-Meier method indicated that the OS of the training set, test set,

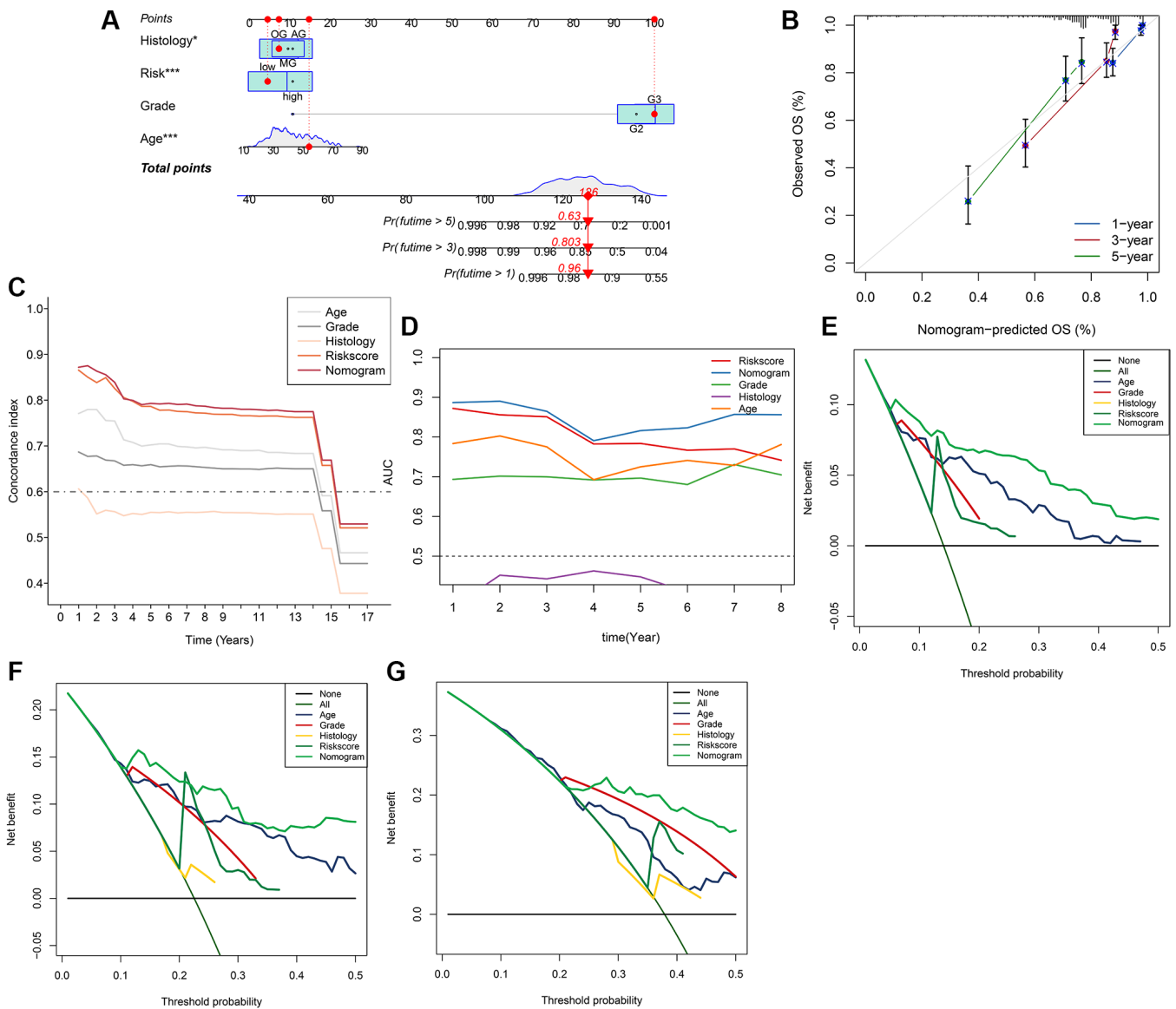


Figure 9. Nomogram model for the prediction of LGG prognosis for 1-year, 3-year, 5-year OS in the whole set. (A) Nomogram of applied to predict survival. **(B)** Calibration curve. **(C)** Graph showing concordance index changes over time. **(D)** AUC values of time-dependent ROC curves changes over time. **(E–G)** Decision curve analysis for 1-year, 3-year, 5-year OS.

discovered that SPARC is highly expressed in pancreatic cancer (PC) tissues, and overexpression of SPARC in PC cells can induce epithelial mesenchymal transition (EMT) and stimulate the migration and invasion of cancer cells [33]. In endometrial cancer (EC), low SPARC expression is related to aggressive EC phenotype and poor prognosis. In addition, down-regulation of SPARC promotes the EMT process and enhances EC cell proliferation and invasion [34]. The AKAP13 is an anchor protein found in the Rho signaling pathway. AKAP13 is expressed increasingly in hepatocellular carcinoma, but it is not expressed in healthy adult liver. Its overexpression in hepatocellular carcinoma cell lines promotes cell proliferation and leads to increased levels of downstream ERK and cyclin D1 [35]. AKAP13 plays a major role in PKA-induced phosphorylation of ER, which is a significant cause of tamoxifen resistance in breast cancer cells and cancer patients [36].

Recently, many studies have stressed the significance of immunotherapy in glioma and it has become a research hotspot [37]. Due to the heterogeneity of gliomas, the therapeutic effect of gliomas is not entirely satisfactory [38]. TMB is a potential biomarker for predicting immunotherapy outcomes for different types of cancers [39, 40]. At the same time, CNVs may provide a better predictor of immunotherapy response than traditional biomarkers. These two molecular characteristics provide new perspectives and good practical methods for immunotherapy [41]. Consequently, we explored variation in genetic mutations between the high-risk and low-risk groups. The high-risk group had substantially increased percentages of TMB and CNV as compared to the low-risk group. The immune infiltration level greatly affects the prognosis of tumor patients. The tumor immune microenvironment contains stroma and immune cells and is associated with immunotherapeutic responses. Additionally, targeting tumor immune checkpoints can serve as a new technique for killing tumor cells, and the expression of immune checkpoints is related to immunotherapeutic responses [42, 43]. Considering the importance of immune status in immunotherapy, the relationship between the signature and immune cell infiltration and immune checkpoint was identified in this study. The ESTIMATE and ssGSEA algorithms suggested that the high-risk group had a higher proportion of immune and stromal cells, stronger immune function, and lower tumor purity. Immune checkpoint expression analysis yielded similar results for immune cell infiltration, again with higher levels in the high-risk group. Due to the lack of open data on LGG patients receiving immunotherapy and APA testing, the TIDE algorithm was used for preliminary

identification of the response of this cohort to immunotherapy. The outcomes showed that the high-risk group had a lower TIDE score and might have increased sensitivity to immunotherapy. Chemotherapy and targeted gene therapy have been shown to have many advantages in prolonging survival in LGG patients [44, 45]. Hence, it is important to predict the treatment response of LGG patients to chemotherapy agents and molecularly targeted antitumor agents. The data in this research indicated that patients in the low-risk groups were more sensitive to 16 drugs, including Gefitinib, Lenalidomide, and Axitinib, as compared to those in the high-risk groups. These findings showed that the APA signature had a close correlation with immunotherapy and chemotherapy and it can be used to select individualized treatment strategies for LGG patients.

There are some deficiencies in this study that can be highlighted. For instance, this research was based on one study cohort only. Only internal validation was performed in the absence of another dataset for external validation. Moreover, this signature has not been clinically validated with a large sample, so its clinical practicality cannot be evaluated directly. Lastly, due to the lack of *in vitro* or *in vivo* experiments, the specific molecular mechanisms of these biomarkers remain unclear. In future studies, we will conduct detailed studies for the validation of our current results.

In conclusion, this study comprehensively analyzed the specific role of APA events in LGG and identified survival-related APA events for the first time. Furthermore, we created prediction signatures based on these events that can accurately stratify the risk and predict the prognosis of LGG patients. In addition, APA events in the signature are expected to be targets for LGG treatment in the future.

AUTHOR CONTRIBUTIONS

YZW and YZ conceived and designed the study. YZW and YFX performed the collection and assembly of data. YZW and YZ analyzed the data. YZW and YFX wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We're grateful for the TCGA project's contributions.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This study was supported by the incubated thesis capital of the People's Hospital of Deyang City (Grant number FHS202004).

REFERENCES

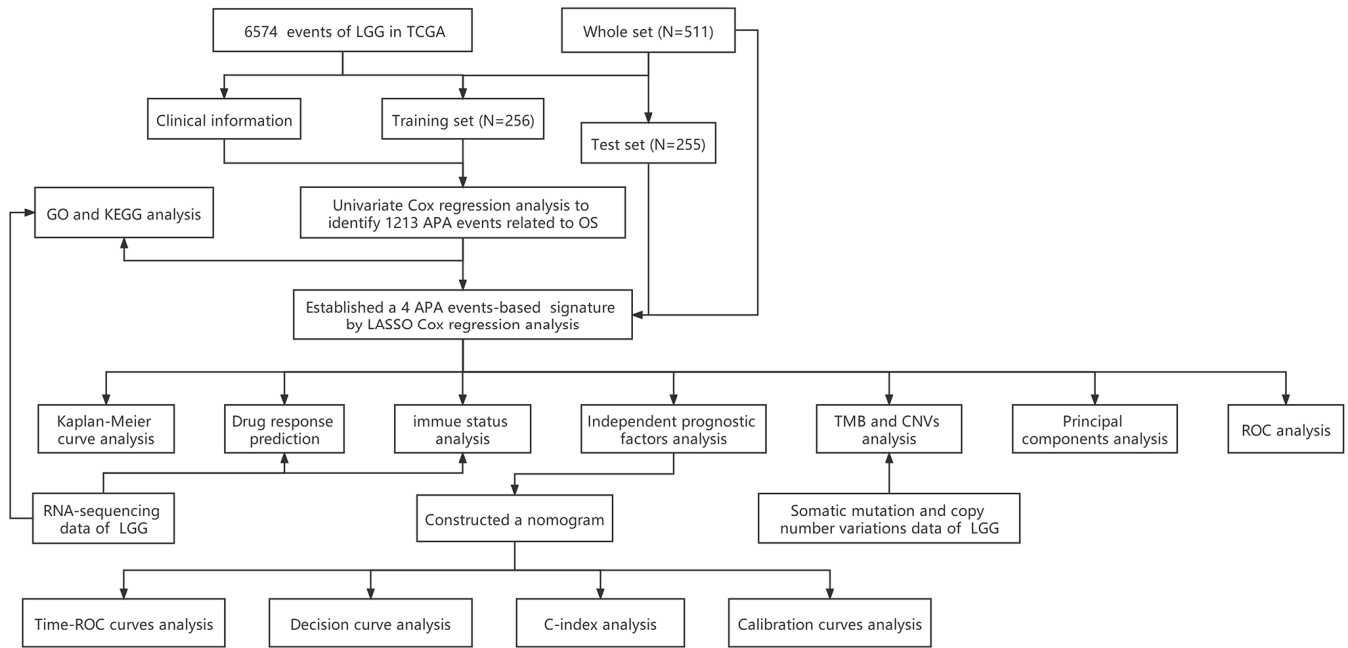
1. Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M. Epidemiology and molecular pathology of glioma. *Nat Clin Pract Neurol*. 2006; 2:494–503.
<https://doi.org/10.1038/ncpneuro0289>
PMID:[16932614](https://pubmed.ncbi.nlm.nih.gov/16932614/)
2. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, Pekmezci M, Schwartzbaum JA, Turner MC, Walsh KM, Wrensch MR, Barnholtz-Sloan JS. The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol*. 2014; 16:896–913.
<https://doi.org/10.1093/neuonc/nou087>
PMID:[24842956](https://pubmed.ncbi.nlm.nih.gov/24842956/)
3. Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, Rheinbay E, Miller CR, Vitucci M, Morozova O, Robertson AG, Nushmehr H, Laird PW, et al, and Cancer Genome Atlas Research Network. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med*. 2015; 372:2481–98.
<https://doi.org/10.1056/NEJMoa1402121>
PMID:[26061751](https://pubmed.ncbi.nlm.nih.gov/26061751/)
4. McNamara MG, Lwin Z, Jiang H, Chung C, Millar BA, Sahgal A, Laperriere N, Mason WP. Conditional probability of survival and post-progression survival in patients with glioblastoma in the temozolomide treatment era. *J Neurooncol*. 2014; 117:153–60.
<https://doi.org/10.1007/s11060-014-1368-7>
PMID:[24469855](https://pubmed.ncbi.nlm.nih.gov/24469855/)
5. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*. 2017; 18:18–30.
<https://doi.org/10.1038/nrm.2016.116>
PMID:[27677860](https://pubmed.ncbi.nlm.nih.gov/27677860/)
6. Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG. Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. *Proc Natl Acad Sci U S A*. 2011; 108:12533–8.
<https://doi.org/10.1073/pnas.1019732108>
PMID:[21746925](https://pubmed.ncbi.nlm.nih.gov/21746925/)
7. Gruber AJ, Zavolan M. Alternative cleavage and polyadenylation in health and disease. *Nat Rev Genet*. 2019; 20:599–614.
<https://doi.org/10.1038/s41576-019-0145-z>
PMID:[31267064](https://pubmed.ncbi.nlm.nih.gov/31267064/)
8. Ogorodnikov A, Levin M, Tattikota S, Tokalov S, Hoque M, Scherzinger D, Marini F, Poetsch A, Binder H, Macher-Göppinger S, Probst HC, Tian B, Schaefer M, et al. Transcriptome 3'end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat Commun*. 2018; 9:5331.
<https://doi.org/10.1038/s41467-018-07580-5>
PMID:[30552333](https://pubmed.ncbi.nlm.nih.gov/30552333/)
9. Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, Wagner EJ. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*. 2014; 510:412–6.
<https://doi.org/10.1038/nature13261>
PMID:[24814343](https://pubmed.ncbi.nlm.nih.gov/24814343/)
10. Chu Y, Elrod N, Wang C, Li L, Chen T, Routh A, Xia Z, Li W, Wagner EJ, Ji P. Nudt21 regulates the alternative polyadenylation of Pak1 and is predictive in the prognosis of glioblastoma patients. *Oncogene*. 2019; 38:4154–68.
<https://doi.org/10.1038/s41388-019-0714-9>
PMID:[30705404](https://pubmed.ncbi.nlm.nih.gov/30705404/)
11. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011; 12:R41.
<https://doi.org/10.1186/gb-2011-12-4-r41>
PMID:[21527027](https://pubmed.ncbi.nlm.nih.gov/21527027/)
12. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia 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/)
13. Zhang X, Shi M, Chen T, Zhang B. Characterization of the Immune Cell Infiltration Landscape in Head and Neck Squamous Cell Carcinoma to Aid Immunotherapy. *Mol Ther Nucleic Acids*. 2020; 22:298–309.
<https://doi.org/10.1016/j.omtn.2020.08.030>
PMID:[33230435](https://pubmed.ncbi.nlm.nih.gov/33230435/)
14. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. 2018; 24:1550–8.
<https://doi.org/10.1038/s41591-018-0136-1>
PMID:[30127393](https://pubmed.ncbi.nlm.nih.gov/30127393/)
15. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression

- levels and in vitro drug sensitivity in cell lines. *Genome Biol.* 2014; 15:R47.
<https://doi.org/10.1186/gb-2014-15-3-r47>
 PMID:24580837
16. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One.* 2014; 9:e107468.
<https://doi.org/10.1371/journal.pone.0107468>
 PMID:25229481
 17. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell.* 2009; 33:365–76.
<https://doi.org/10.1016/j.molcel.2008.12.028>
 PMID:19217410
 18. Cheng J, Meng J, Zhu L, Peng Y. Exosomal noncoding RNAs in Glioma: biological functions and potential clinical applications. *Mol Cancer.* 2020; 19:66.
<https://doi.org/10.1186/s12943-020-01189-3>
 PMID:32213181
 19. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and molecular epidemiology of adult diffuse glioma. *Nat Rev Neurol.* 2019; 15:405–17.
<https://doi.org/10.1038/s41582-019-0220-2>
 PMID:31227792
 20. Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun.* 2014; 5:5274.
<https://doi.org/10.1038/ncomms6274>
 PMID:25409906
 21. Lembo A, Di Cunto F, Provero P. Shortening of 3'UTRs correlates with poor prognosis in breast and lung cancer. *PLoS One.* 2012; 7:e31129.
<https://doi.org/10.1371/journal.pone.0031129>
 PMID:22347440
 22. Wang L, Hu X, Wang P, Shao ZM. The 3'UTR signature defines a highly metastatic subgroup of triple-negative breast cancer. *Oncotarget.* 2016; 7:59834–44.
<https://doi.org/10.18632/oncotarget.10975>
 PMID:27494850
 23. Wang L, Hu X, Wang P, Shao ZM. Integrative 3' Untranslated Region-Based Model to Identify Patients with Low Risk of Axillary Lymph Node Metastasis in Operable Triple-Negative Breast Cancer. *Oncologist.* 2019; 24:22–30.
<https://doi.org/10.1634/theoncologist.2017-0609>
 PMID:30082491
 24. Wang L, Lang GT, Xue MZ, Yang L, Chen L, Yao L, Li XG, Wang P, Hu X, Shao ZM. Dissecting the heterogeneity of the alternative polyadenylation profiles in triple-negative breast cancers. *Theranostics.* 2020; 10:10531–47.
<https://doi.org/10.7150/thno.40944>
 PMID:32929364
 25. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* 2008; 26:1364–70.
<https://doi.org/10.1200/JCO.2007.12.9791>
 PMID:18323559
 26. Brown L, Ongusaha PP, Kim HG, Nuti S, Mandinova A, Lee JW, Khosravi-Far R, Aaronson SA, Lee SW. CDIP, a novel pro-apoptotic gene, regulates TNFalpha-mediated apoptosis in a p53-dependent manner. *EMBO J.* 2007; 26:3410–22.
<https://doi.org/10.1038/sj.emboj.7601779>
 PMID:17599062
 27. Zhou X, Feng Y, Liu S, Li C, Teng Y, Li X, Lu J. IL-33 Promotes the Growth of Non-Small Cell Lung Cancer Cells Through Regulating miR-128-3p/CDIP1 Signalling Pathway. *Cancer Manag Res.* 2021; 13:2379–88.
<https://doi.org/10.2147/CMAR.S276297>
 PMID:33737835
 28. Joenje H. Fanconi anaemia complementation groups in Germany and The Netherlands. European Fanconi Anaemia Research group. *Hum Genet.* 1996; 97:280–2.
<https://doi.org/10.1007/BF02185753>
 PMID:8786063
 29. Savino M, Ianzano L, Strippoli P, Ramenghi U, Arslanian A, Bagnara GP, Joenje H, Zelante L, Savoia A. Mutations of the Fanconi anemia group A gene (FAA) in Italian patients. *Am J Hum Genet.* 1997; 61:1246–53.
<https://doi.org/10.1086/301632>
 PMID:9399890
 30. Abbasi S, Rasouli M. A rare FANCA gene variation as a breast cancer susceptibility allele in an Iranian population. *Mol Med Rep.* 2017; 15:3983–8.
<https://doi.org/10.3892/mmr.2017.6489>
 PMID:28440412
 31. Marshall CH, Fu W, Wang H, Baras AS, Lotan TL, Antonarakis ES. Prevalence of DNA repair gene mutations in localized prostate cancer according to clinical and pathologic features: association of Gleason score and tumor stage. *Prostate Cancer Prostatic Dis.* 2019; 22:59–65.
<https://doi.org/10.1038/s41391-018-0086-1>
 PMID:30171229
 32. Camacho D, Jesus JP, Palma AM, Martins SA, Afonso A, Peixoto ML, Pelham CJ, Moreno E, Gogna R. SPARC-p53: The double agents of cancer. *Adv Cancer Res.* 2020; 148:171–99.

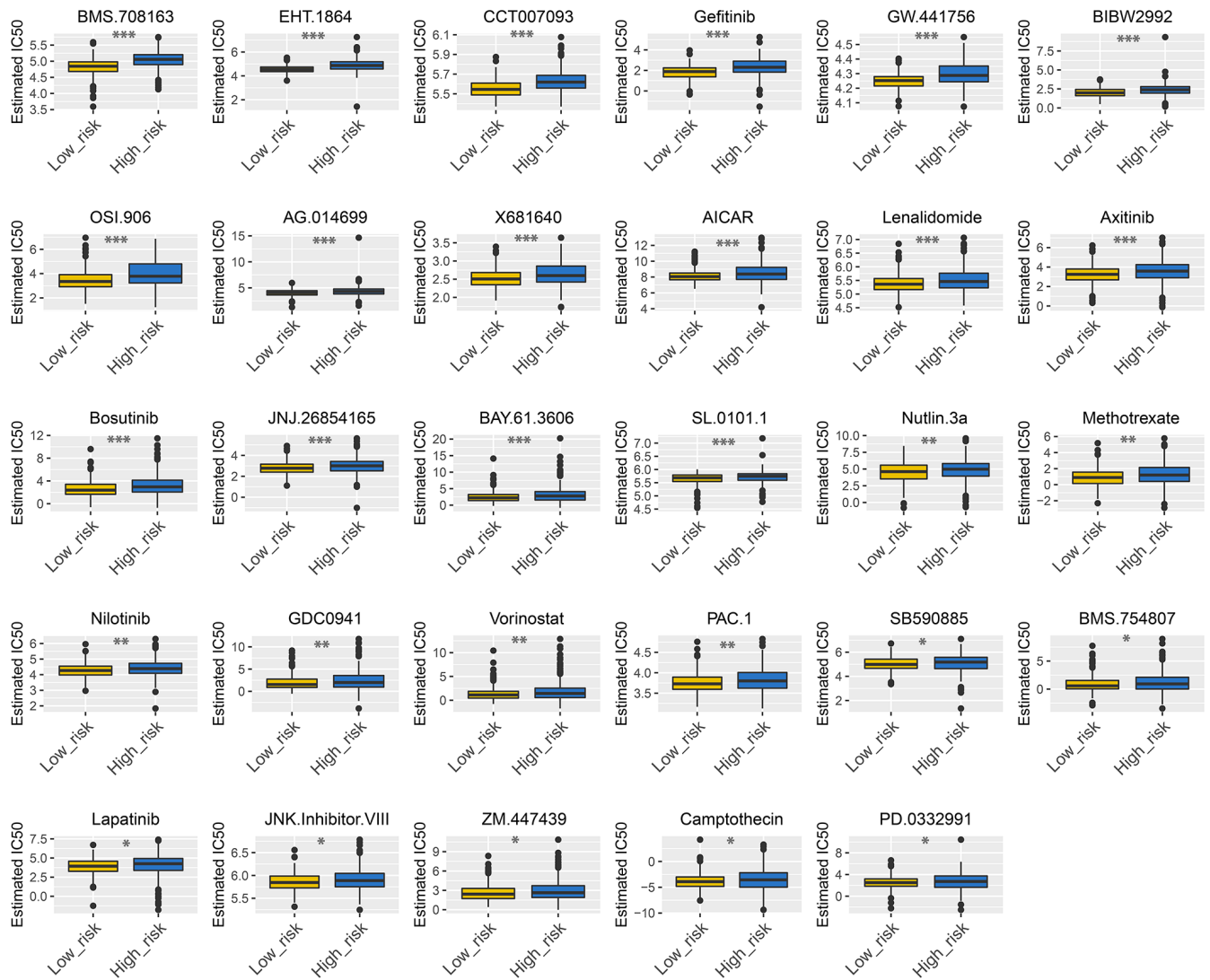
- <https://doi.org/10.1016/bs.acr.2020.05.004>
PMID:[32723563](https://pubmed.ncbi.nlm.nih.gov/32723563/)
33. López-Moncada F, Torres MJ, Castellón EA, Contreras HR. Secreted protein acidic and rich in cysteine (SPARC) induces epithelial-mesenchymal transition, enhancing migration and invasion, and is associated with high Gleason score in prostate cancer. *Asian J Androl.* 2019; 21:557–64.
https://doi.org/10.4103/aja.aja_23_19
PMID:[31031331](https://pubmed.ncbi.nlm.nih.gov/31031331/)
34. Wang L, Wang W, Xu Y, Wang Q. Low Levels of SPARC are Associated with Tumor Progression and Poor Prognosis in Human Endometrial Carcinoma. *Onco Targets Ther.* 2020; 13:11549–69.
<https://doi.org/10.2147/OTT.S277795>
PMID:[33204109](https://pubmed.ncbi.nlm.nih.gov/33204109/)
35. Sterpetti P, Marucci L, Candelaresi C, Toksoz D, Alpini G, Ugili L, Baroni GS, Macarri G, Benedetti A. Cell proliferation and drug resistance in hepatocellular carcinoma are modulated by Rho GTPase signals. *Am J Physiol Gastrointest Liver Physiol.* 2006; 290:G624–32.
<https://doi.org/10.1152/ajpgi.00128.2005>
PMID:[16322093](https://pubmed.ncbi.nlm.nih.gov/16322093/)
36. Bentin Toaldo C, Alexi X, Beelen K, Kok M, Hauptmann M, Jansen M, Berns E, Neefjes J, Linn S, Michalides R, Zwart W. Protein Kinase A-induced tamoxifen resistance is mediated by anchoring protein AKAP13. *BMC Cancer.* 2015; 15:588.
<https://doi.org/10.1186/s12885-015-1591-4>
PMID:[26272591](https://pubmed.ncbi.nlm.nih.gov/26272591/)
37. Strepkos D, Markouli M, Klonou A, Piperi C, Papavassiliou AG. Insights in the immunobiology of glioblastoma. *J Mol Med (Berl).* 2020; 98:1–10.
<https://doi.org/10.1007/s00109-019-01835-4>
PMID:[31650201](https://pubmed.ncbi.nlm.nih.gov/31650201/)
38. Sampson JH, Gunn MD, Fecci PE, Ashley DM. Brain immunology and immunotherapy in brain tumours. *Nat Rev Cancer.* 2020; 20:12–25.
<https://doi.org/10.1038/s41568-019-0224-7>
PMID:[31806885](https://pubmed.ncbi.nlm.nih.gov/31806885/)
39. Wang R, Yang Y, Ye WW, Xiang J, Chen S, Zou WB, Wang XJ, Chen T, Cao WM. Case Report: Significant Response to Immune Checkpoint Inhibitor Camrelizumab in a Heavily Pretreated Advanced ER+/HER2- Breast Cancer Patient With High Tumor Mutational Burden. *Front Oncol.* 2021; 10:588080.
<https://doi.org/10.3389/fonc.2020.588080>
PMID:[33634015](https://pubmed.ncbi.nlm.nih.gov/33634015/)
40. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, Kurzrock R. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther.* 2017; 16:2598–608.
<https://doi.org/10.1158/1535-7163.MCT-17-0386>
PMID:[28835386](https://pubmed.ncbi.nlm.nih.gov/28835386/)
41. Lu Z, Chen H, Li S, Gong J, Li J, Zou J, Wu L, Yu J, Han W, Sun H, Jiao X, Zhang X, Peng Z, et al. Tumor copy-number alterations predict response to immune-checkpoint-blockade in gastrointestinal cancer. *J Immunother Cancer.* 2020; 8:e000374.
<https://doi.org/10.1136/jitc-2019-000374>
PMID:[32792358](https://pubmed.ncbi.nlm.nih.gov/32792358/)
42. Liu X, Guo CY, Tou FF, Wen XM, Kuang YK, Zhu Q, Hu H. Association of PD-L1 expression status with the efficacy of PD-1/PD-L1 inhibitors and overall survival in solid tumours: A systematic review and meta-analysis. *Int J Cancer.* 2020; 147:116–27.
<https://doi.org/10.1002/ijc.32744>
PMID:[31633798](https://pubmed.ncbi.nlm.nih.gov/31633798/)
43. Ribas A, Hu-Lieskovan S. What does PD-L1 positive or negative mean? *J Exp Med.* 2016; 213:2835–40.
<https://doi.org/10.1084/jem.20161462>
PMID:[27903604](https://pubmed.ncbi.nlm.nih.gov/27903604/)
44. Gnekow AK, Walker DA, Kandels D, Picton S, Perilongo G, Grill J, Stokland T, Sandstrom PE, Warmuth-Metz M, Pietsch T, Giangaspero F, Schmidt R, Faldum A, et al, and of the Low Grade Glioma Consortium and the participating centers. A European randomised controlled trial of the addition of etoposide to standard vincristine and carboplatin induction as part of an 18-month treatment programme for childhood (≤ 16 years) low grade glioma - A final report. *Eur J Cancer.* 2017; 81:206–25.
<https://doi.org/10.1016/j.ejca.2017.04.019>
PMID:[28649001](https://pubmed.ncbi.nlm.nih.gov/28649001/)
45. de Blank P, Fouladi M, Huse JT. Molecular markers and targeted therapy in pediatric low-grade glioma. *J Neurooncol.* 2020; 150:5–15.
<https://doi.org/10.1007/s11060-020-03529-1>
PMID:[32399739](https://pubmed.ncbi.nlm.nih.gov/32399739/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. Flow chart of the present study.



Supplementary Figure 2. Boxplots of estimated IC50 values of potential compounds between high-risk and low-risk groups.
 * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Supplementary Table

Supplementary Table 1. The coefficients of the four events.

Id	Coef	HR	HR.95L	HR.95H	p value
`NM_013399 C16orf5 chr16 -'`	5.846687	346.0859	4.230633	28311.47	0.009273
`NM_000135 FANCA chr16 -'`	7.129974	1248.845	11.95803	130423.9	0.002645
`NM_003118 SPARC chr5 -'`	-4.96019	0.007012	0.000385	0.127679	0.000808
`NM_007200 AKAP13 chr15 +'`	1.243714	3.468473	0.671534	17.91467	0.13764