Heliyon 9 (2023) e18075

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

A novel, effective machine learning-based RNA editing profile for predicting the prognosis of lower-grade gliomas

Boshen Wang [a,b,1], Peijie Tian [c,1], Qianyu Sun [b], Hengdong Zhang [a], Lei Han [a], Baoli Zhu [a,b,*]

[a] Jiangsu Provincial Center for Disease Prevention and Control, Nanjing 210000, Jiangsu, China
[b] Key Laboratory of Environmental Medicine Engineering of Ministry of Education, School of Public Health, Southeast University, Nanjing 210009, Jiangsu, China
[c] Department of Pathology, Weifang Medical University, China

ARTICLE INFO

Keywords:
Low-grade glioma
Machine learning
Prognosis
TCGA

ABSTRACT

Patients with low-grade glioma (LGG) may survive for long time periods, but their tumors often progress to higher-grade lesions. Currently, no cure for LGG is available. A-to-I RNA editing accounts for nearly 90% of all RNA editing events in humans and plays a role in tumorigenesis in various cancers. However, little is known regarding its prognostic role in LGG. On the basis of The Cancer Genome Atlas (TCGA) data, we used LASSO and univariate Cox regression to construct an RNA editing site signature. The results derived from the TCGA dataset were further validated with Gene Expression Omnibus (GEO) and Chinese Glioma Genome Atlas (CGGA) datasets. Five machine learning algorithms (Decision Trees C5.0, XGboost, GBDT, Lightgbm, and Catboost) were used to confirm the prognosis associated with the RNA editing site signature. Finally, we explored immune function, immunotherapy, and potential therapeutic agents in the high- and low-risk groups by using multiple biological prediction websites. A total of 22,739 RNA editing sites were identified, and a signature model consisting of four RNA editing sites (PRKCSH|chr19:11561032, DSEL|chr18:65174489, UGGT1|chr2:128952084, and SOD2|chr6:160101723) was established. Cox regression analysis indicated that the RNA editing signature was an independent prognostic factor, according to the ROC curve (AUC = 0.823), and the nomogram model had good predictive power (C-index = 0.824). In addition, the predictive ability of the RNA editing signature was confirmed with the machine learning model. The sensitivity of PCI-34051 and Elephantin was significantly higher in the high-risk group than the low-risk group, thus potentially providing a marker to predict the effects of lung cancer drug treatment. RNA editing may serve as a novel survival prediction tool, thus offering hope for developing editing-based therapeutic strategies to combat LGG progression. In addition, this tool may help optimize survival risk assessment and individualized care for patients with low-grade gliomas.

* Corresponding author. Institute of Occupational Disease Prevention, Jiangsu Provincial Center for Disease Prevention and Control. No. 172 Jiangsu Road, Nanjing, Jiangsu, China.
E-mail address: zhublcdc@sina.com (B. Zhu).
[1] Boshen Wang and Peijie Tian contributed equally to this article.

https://doi.org/10.1016/j.heliyon.2023.e18075
Received 14 January 2023; Received in revised form 2 July 2023; Accepted 5 July 2023
Available online 7 July 2023

1. Introduction

Despite decades of global research, low-grade glioma (LGG) treatment remains controversial [1]. LGG is a progressive and invasive central nervous system disorder [2,3]. Although LGG develops relatively slowly, its diffuse and infiltrative nature makes cure difficult [4,5], and malignant progression to high-grade glioma inevitably occurs. The median survival of patients with LGG varies from 5.6 to 13.3 years, on the basis of various parameters, such as the extent of resection and molecular characteristics [6]. With the adoption of evidence-based and personalized medicine, individualized biomarkers play crucial roles in treatment [7]; hence, we sought to identify valuable predictive biomarkers for LGG, to improve its clinical diagnosis and treatment.

RNA editing refers to alterations in the chemical structure of RNA molecules occurring after DNA transcription and synthesis by the RNA polymerase enzyme [8]. These post-transcriptional alterations may include the replacement of nucleotides, M6A methylation, nucleotide insertion, and several other modifications that alter the RNA sequences encoded by the genomic sequences [9]. RNA editing has been shown to play roles in cancer progression, including tumor growth, metastasis, and drug resistance. Aberrant RNA editing events have been identified in various cancer types, including breast, lung, liver, and prostate cancers [10]. For example, increased adenosine-to-inosine (A-to-I) editing of transcripts encoding ion channels and receptors has been observed in lung cancer, and has been found to alter calcium signaling and increase cell proliferation [11]. In breast cancer, estrogen receptor RNA editing has been shown to alter the response to endocrine therapy, thereby leading to resistance [12]. Thus, RNA editing is a critical biological process that contributes to the diversity and complexity of the transcriptome. Dysregulation of RNA editing has been implicated in cancer progression and targeting RNA editing is a promising therapeutic strategy for cancer treatment [13].

Maas et al. [14] first identified an association between RNA editing and glioma, specifically the pathogenesis of glioblastoma. They have demonstrated decreased editing at the Q/R site of GluR-B, with no corresponding change in ADAR2 expression and no difference in the alternative splicing of ADAR2 mRNA in tumor and normal tissues. Unknown biological mechanisms may influence the catalytic activity of ADAR2, including posttranslational modification or subcellular localization. ADAR2 catalytic activity has further been correlated with a decrease in editing levels in juvenile astrocytomas [15]. In addition, alternative splicing at exon 2 of the ADAR1 pre-mRNA produces a 110-kDa protein rather than the full-length 150-kDa protein in high-grade malignancies. When ADAR1 ASVs are overexpressed, they form heterodimers with ADAR2, thus perturbing the balance among ADAR1, ADAR2, and ADAR3, and compete with ADAR2 for editing activity at the Q/R site of GluR-B. Together, these findings indicate that alternative splicing in ADAR1 regulates RNA editing mediated by ADAR2. RNA editing is more tumor-specific than gene expression, and is independent of inter-individual heterogeneity in extracted RNA quantity and reference gene selection [16,17]. RNA editing is preferable to RNA expression as a biomarker in terms of test stability and reliability.

A-to-I RNA editing is the most ubiquitous type of RNA editing in humans, accounting for almost 90% of all RNA editing events [18]. Therefore, using The Cancer Genome Atlas (TCGA), Chinese Glioma Genome Atlas (CGGA), and Gene Expression Omnibus (GEO) data, we developed a novel predictive model that elucidates the prognostic and regulatory potential of RNA editing in LGG. In addition, we analyzed immune function, immunotherapy, and potential therapeutic agents in high and low LGG risk groups through a bioinformatics approach.

2. Results

2.1. Overview of RNA editing sites in LGG

Most RNA editing sites were found to be situated in non-coding regions, such as 3′ untranslated regions (40.26%), intron regions (32.13%), and intergenic regions (14.22%), according to annotations of RNA editing site distribution (Fig. 1A). In addition, 95.76% of all A-to-I editing sites were found in Alu repetitive regions (Fig. 1B).

2.2. Identification of LGG survival-associated RNA editing sites

The Manhattan plot for assessing correlations between 22,739 RNA sites and LGG overall survival (OS) in TCGA cohort data indicated 21 RNA sites in the Cox-PH model with $p < 1 \times 10^{-6}$ (Fig. 1C and D). To develop a predictive model for RNA editing sites in LGG, we performed selection operator (least absolute shrinkage and selection operator, LASSO) Cox regression and least absolute shrinkage (Fig. 2A and B). Consequently, four prognostic RNA editing sites ($p < 0.01$) were filtered to build the prognostic model. Four tagging RNA editing sites (PRKCSH|chr19:11561032, DSEL|chr18:65174489, UGGT1|chr2:128952084, and SOD2|chr6:160101723) were then used to calculate risk scores for each sample, and patients were divided into a high-risk group and a low-risk group (Fig. 2C and D). The high-risk group had a higher mortality rate than the low-risk group (Fig. 2E and F). We next performed univariate and multivariate analyses using the Cox proportional hazard model in the cohort. Higher risk scores were correlated with poorer survival in both the univariate and multivariate Cox regression models (Fig. 2G and H).

In the training and testing cohorts, Kaplan-Meier survival curves revealed a significant difference in OS between the high- and low-risk groups ($P < 0.001$, Fig. 3A and B), thus indicating that the newly discovered signature accurately predicted survival. In agreement with these findings, for four tagging RNA editing sites, the high-risk group displayed poorer OS than the low-risk group ($P < 0.001$, Fig. 3C–F). The Kaplan-Meier curves revealed that the patients in the high-risk group had greater chances of tumor progression and metastasis (Fig. 3G).

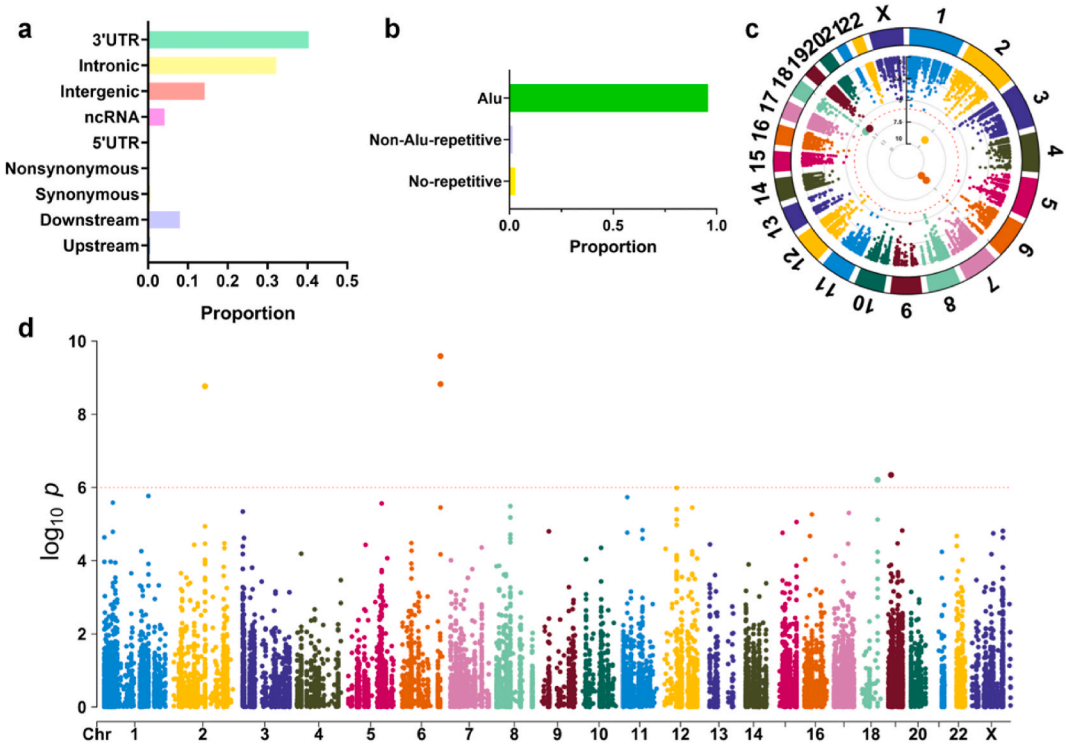


Fig. 1. LGG RNA editing profiles. (A–B) Distribution of RNA editing sites in various gene areas, Alu regions, and overlap with known sites, on the basis of TCGA data. (C–D) Scatter plot showing p-values on a $\log_{10}$ scale derived from the univariate Cox-PH model for correlations between all RNA editing sites and OS in TCGA-LGG patients. The y-axis represents p-values on a $\log_{10}$ scale, whereas the x-axis represents the chromosomal locations of these RNA editing sites.

2.3. *Prognostic evaluation of the RNA editing site signature in LGG, according to TCGA, GEO, and CGGA data*

Correlation analysis was performed between gene expression and RNA editing levels. SOD2 and UGGT1 exhibited positive expression-editing correlations in LGG (Fig. 4A and B). The expression of SOD2 and UGGT1 increased proportionally with increasing risk score (Fig. 4C and D). To determine whether SOD2 and UGGT1 expression might be associated with malignancy prognosis, we obtained tumor data from GEO and CGGA. Poor prognosis of LGG was associated with greater expression levels of SOD2 and UGGT1 in the CGGA cohort (Fig. 4E–F, $P < 0.01$). In addition, the expression levels of SOD2 and UGGT1 in LGG tissues were significantly higher than those in the normal control group in the GEO database (Fig. 4G–K, $P < 0.01$). These results demonstrated that SOD2 and UGGT1 levels were elevated in LGG tissues, thus suggesting that SOD2 and UGGT1 may play crucial regulatory roles in the progression of LGG.

2.4. *Clinical correlation analyses based on TCGA data*

Analysis of 532 TCGA samples revealed a significant association between the risk score and the histological type, tumor location, sex, and age (Fig. 5A–D). Furthermore, the hybrid nomogram integrating clinical features and RNA editing site signatures was accurate and stable (C-index = 0.824), and therefore could be used to evaluate OS in patients with LGG at 1, 3, and 5 years (Fig. 5E and F). According to the ROC curve analysis, the AUC values of the risk score and nomogram in the LGG-TCGA cohort were 0.823 and 0.899, respectively (Fig. 6A). To further assess the predictive accuracy of OS, we performed DCA analysis of TCGA cohort data, in which age and nomogram score had the highest net benefit (Fig. 6B).

2.5. *Evaluation of machine learning predictive models*

XGboost outperformed other machine learning methods in terms of the correct prediction rate and F1 score, with a 99.7% correct prediction rate and F1 score for the training group (Fig. 6C). Compared with the other four models, the Catboost method was more stable, with accuracy rates of 89.3% and 81.4%, and F1 scores of 88.4% and 80.2%, for the training and testing groups, respectively (Fig. 6C–F). To predict survival time, we tested multiple models, among which Lightgbm had the best performance. To evaluate the relative importance of features (risk score, clinical grade, sex, and age), we used the XGboost algorithm to calculate the proportion of importance for each feature on a 10-fold cross-validated training dataset (Fig. 6G). The MSE, RMSE, and MAE values of the XGboost

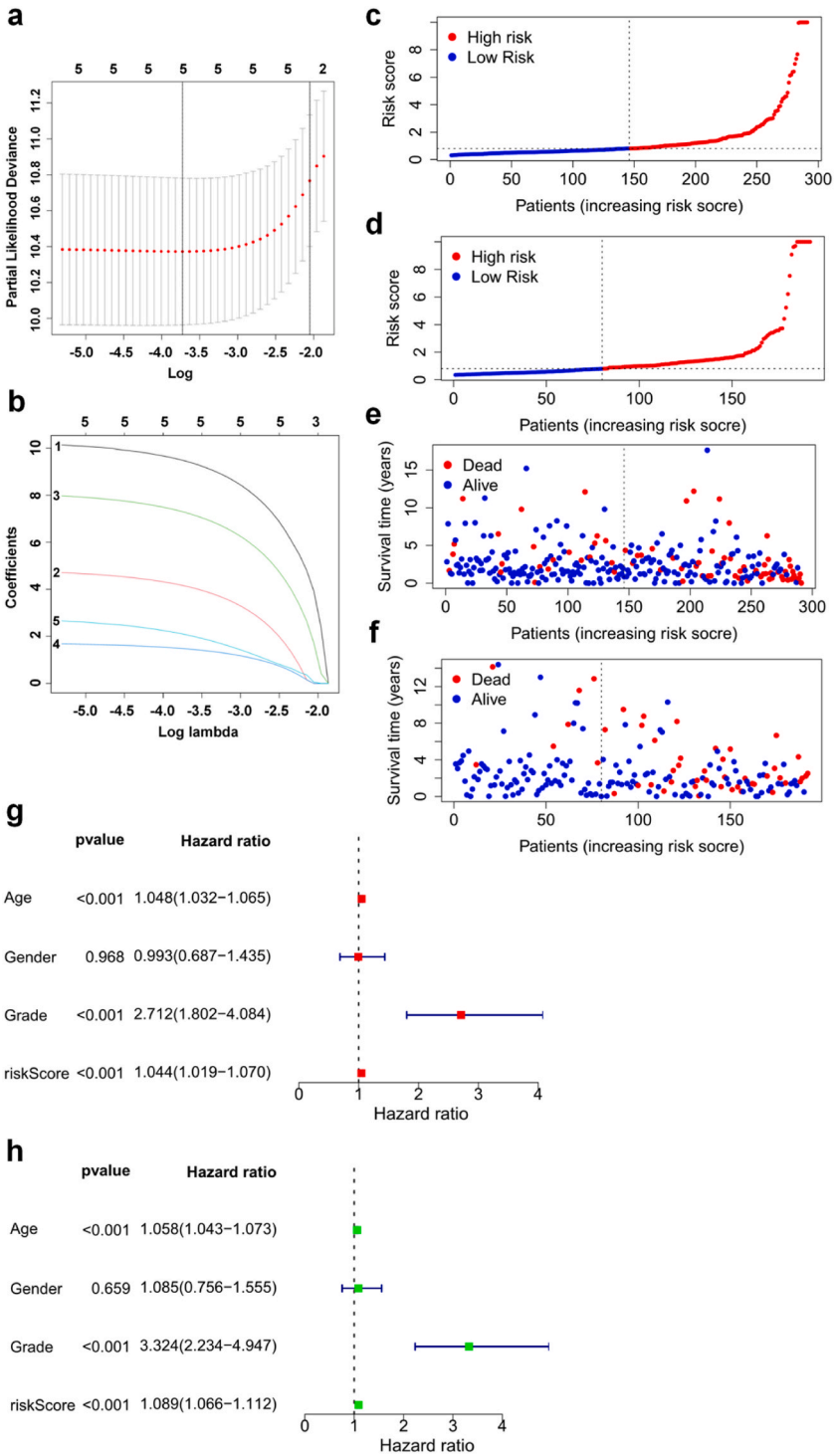


Fig. 2. Independent prognostic value of the RNA editing site signature. (A–B) LASSO regression was performed with the optimal value of λ. (C–D) Risk curves based on each patient's risk score, with blue indicating low risk and red indicating high risk. Each patient's survival status is shown via a scatterplot. (E–F) Survival and death are shown by red and blue dots, respectively. Cox regression for the RNA editing site signature, both univariate (G) and multivariate (H). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

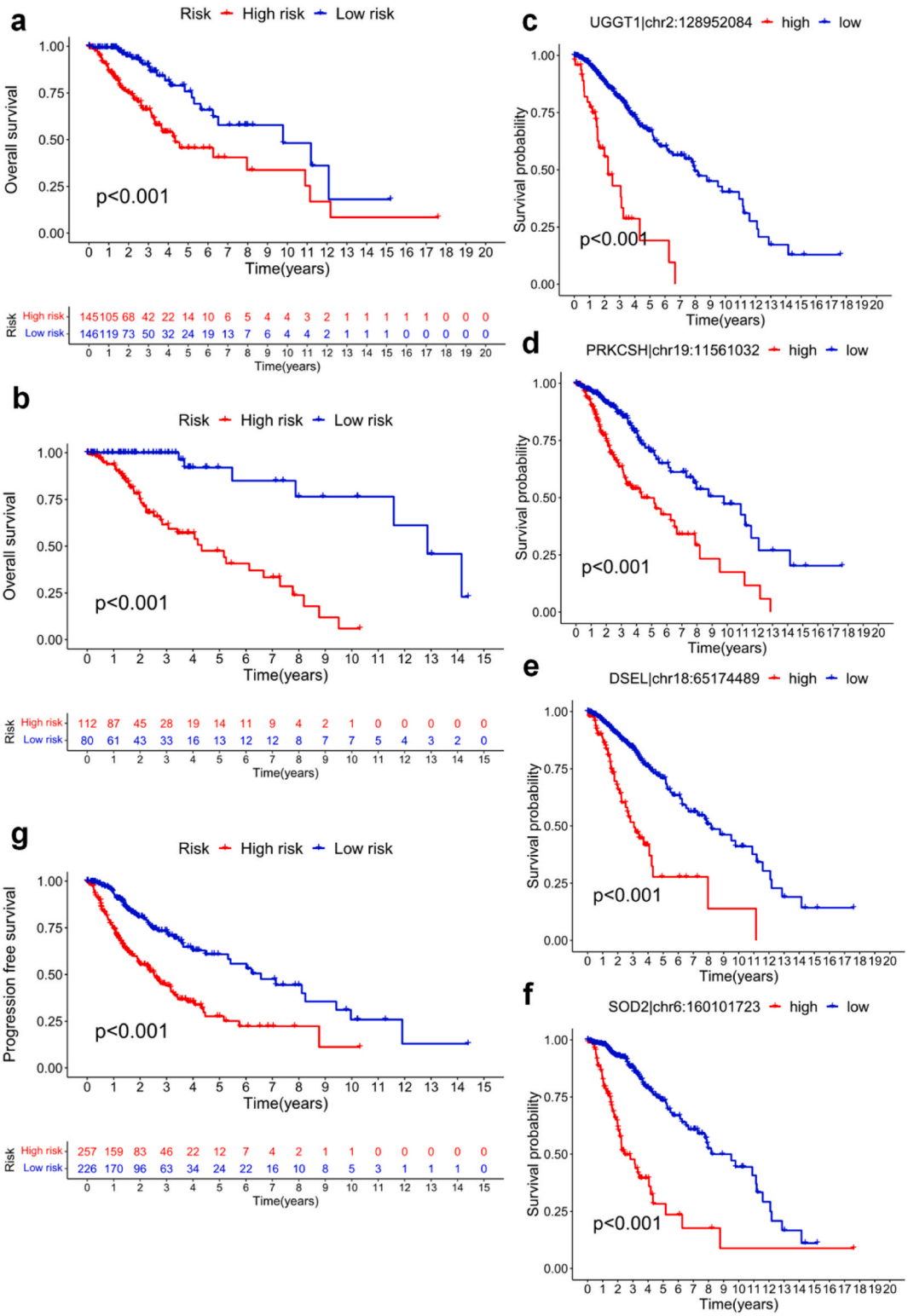
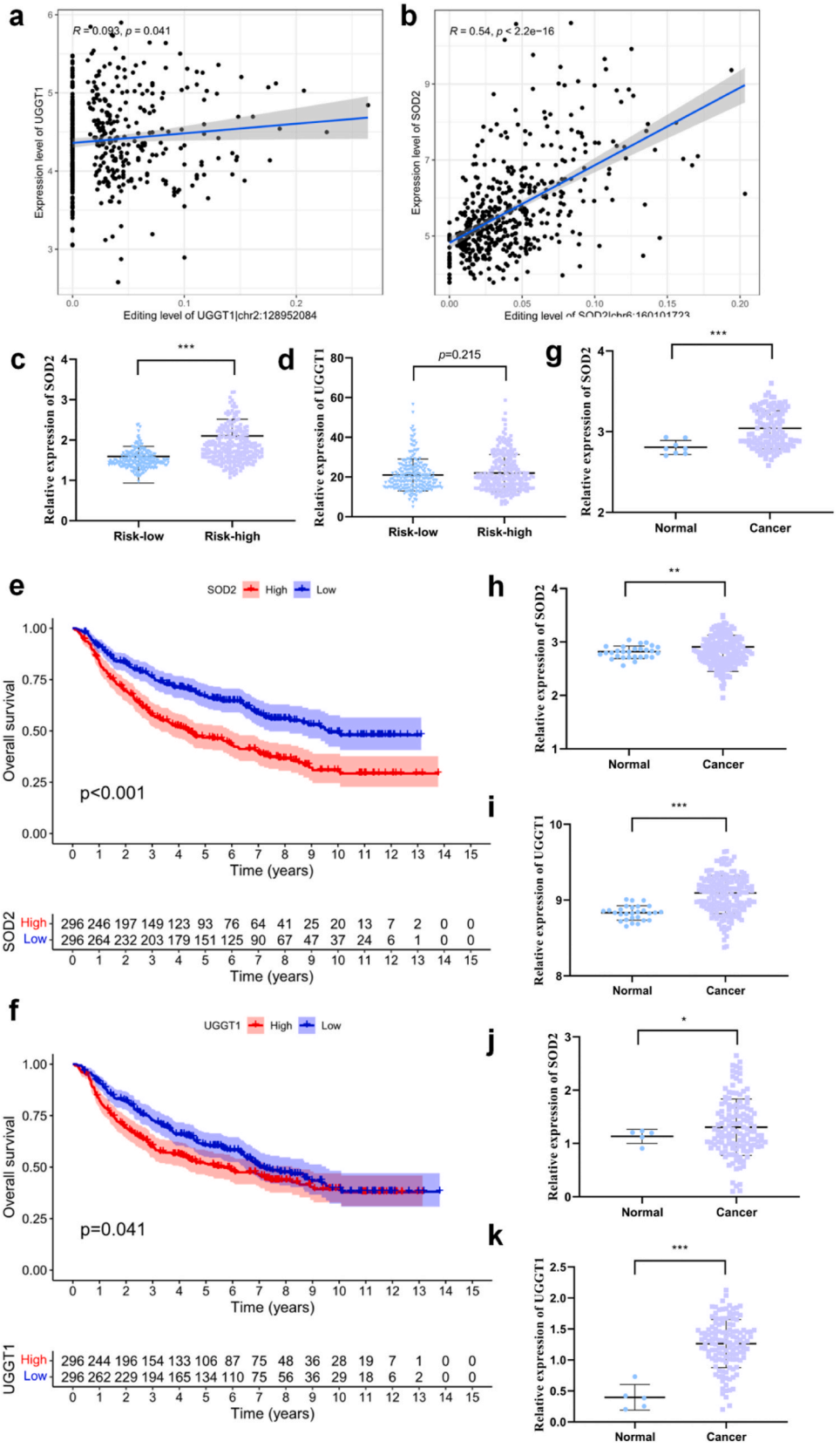


Fig. 3. Kaplan-Meier survival curves in the training (A) and testing cohorts (B) from TCGA data, stratified by risk score. (C–F) Survival curves of the four tagging RNA editing sites. (G) Kaplan-Meier analysis of progression-free survival.



(caption on next page)

Fig. 4. (A–B) Significant correlation between RNA editing level and mRNA expression in tumor tissues. Rank-based Pearson correlations were used and plotted. (C–D) Box plot visualization of SOD2 and UGGT1 expression in the low- and high-risk groups. (E–F) Survival curves for SOD2 and UGGT1, on the basis of CGGA data. Gene expression of SOD2 and UGGT1 in tumors from 3 GEO datasets (G: GSE16011, H–I: GSE68848, J–K: GSE109857) is shown.

predictive models were the smallest, thus indicating that the model using the XGboost algorithm had high predictive power (Supplement Table S1, Fig. 6H).

2.6. Assessment of enrichment of RNA editing site signatures

To determine the roles of RNA editing site signatures in LGG, we examined these genes by using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) databases. The KEGG study demonstrated that RNA editing-associated genes were involved in functions including calcium signaling pathway, regulation of actin cytoskeleton, and Parkinson's disease (Supplement Figure 1A). RNA editing-associated genes were involved in functions including signaling receptor activator activity, receptor-ligand activity, and collagen-containing extracellular matrix, according to the GO database (Supplement Fig. S1B).

2.7. Immune-associated characteristics and drug susceptibility analysis

Tumor-infiltrating lymphocytes are significant predictors of the OS rate and sentinel lymph node status. Therefore, we investigated whether the risk score might correlate with the degree of immune cell infiltration in various tumor types, by using the TIMER database. In the LGG, the risk score was significantly associated with B cells; CD8+T cells; and macrophage, TIL, and neutrophil infiltration levels (Fig. 7A). On the basis of TCGA-LGG data, a correlation analysis of immune-associated functions revealed that functions including checkpoint, APC-CO-inhibition, CCR, HLA, type II INF response, and T cell co-inhibition significantly differed between groups (Fig. 7B). The relationship between risk groups and immunotherapy was also investigated, and a significant difference in TME scores was observed between low- and high-risk individuals (Fig. 8A). Furthermore, the Tumour Immune Dysfunction and Exclusion (TIDE) score in the low-risk group was much higher than that in the high-risk group (Fig. 8B). Overall, patients in the high-risk group had better responses to immunotherapy for LGG, thus indicating that risk scores have value in treatment selection. We performed drug susceptibility analysis for patients with LGG with high and low risk levels to prospectively identify chemotherapeutic or targeted medicines. Patients in the high-risk group exhibited relatively lower half-maximal inhibitory concentration (IC50) values for PCI-34051 and Elephantin, whereas patients in the low-risk group had relatively lower IC50 values for KU-55933 (Fig. 8C–E). These findings support the significance of risk scores in the selection of potential beneficiaries of specific therapeutic chemotherapy and targeted therapeutics among patients with LGG.

3. Discussion

LGGs are slow-growing tumors with an ineluctable propensity for malignancy [19]. The biological characteristics of LGG are notably distinct from those of GBM, and it is regarded as a therapeutic window before the malignant change. Given the prevalence of LGGs in adolescents and children, preserving quality of life and cognitive ability is equally important as extending progression-free survival [20]. Consequently, establishing an accurate diagnosis in early tumor stages might aid in improving patients' clinical prognosis, clarifying the processes underlying the evolution of LGG, and discovering molecular pathways for specialized therapies that may minimize damage.

Recently, RNA editing has emerged as an intriguing field of study [21]. A-to-I RNA editing sites are crucial to the pathophysiology of human malignancies, because they confer apoptosis resistance and selective advantages in the proliferation of tumor cells [22]. In addition to influencing alternative splicing [23,24], the expression of cancer-associated genes [25,26], and the secondary structure of lncRNA [22], they affect a variety of other aspects of cancer. In this research, we systematically explored RNA editing events in LGG and discovered that more than 90% of detected A-to-I editing sites resided in Alu regions. Importantly, we developed a new model for predicting the prognosis of patients with LGG according to RNA editing. The model included four RNA editing sites, two of which were positively correlated with gene expression level, and validated its prognostic value by using the CGGA and GEO databases. Extensive literature has explored the potential roles of SOD2 and UGGT1 in cancer [27–29], but no studies to date have described a mechanism associated with these two genes and LGG. Our bioinformatics analysis sheds new light on the pathogenesis and treatment of LGG.

Furthermore, because of the heterogeneity of tumors and individual patient differences [30], we integrated multiple indicators to build several models for prognosis prediction. First, we constructed a linear visual prognostic prediction model, a nomogram, to assess patient risk in clinical settings by combining risk scores. Machine learning, particularly supervised learning, has the unique ability to anticipate outcomes according to patterns learned from historical data [31]. Regardless of the machine learning method used, machine learning methods appear to improve the accuracy of predictions by an average of 15–25% with respect to alternative or conventional approaches [32]. Combining clinical data with machine learning can give potential meaning to massive clinical datasets [33]. Compared with conventional machine learning algorithms, XGboost exhibits exceptional predictive accuracy and value. Notably, this method surpasses traditional models, such as the Cox model, by significantly enhancing the correct prediction rate from 82.3% to 99.7%. Moreover, machine learning regression models have the potential to reveal credible correlations between risk factors and survival; the highest weight of risk scores has notably reached 53.4%. Additionally, the XGboost model has valuable efficacy in

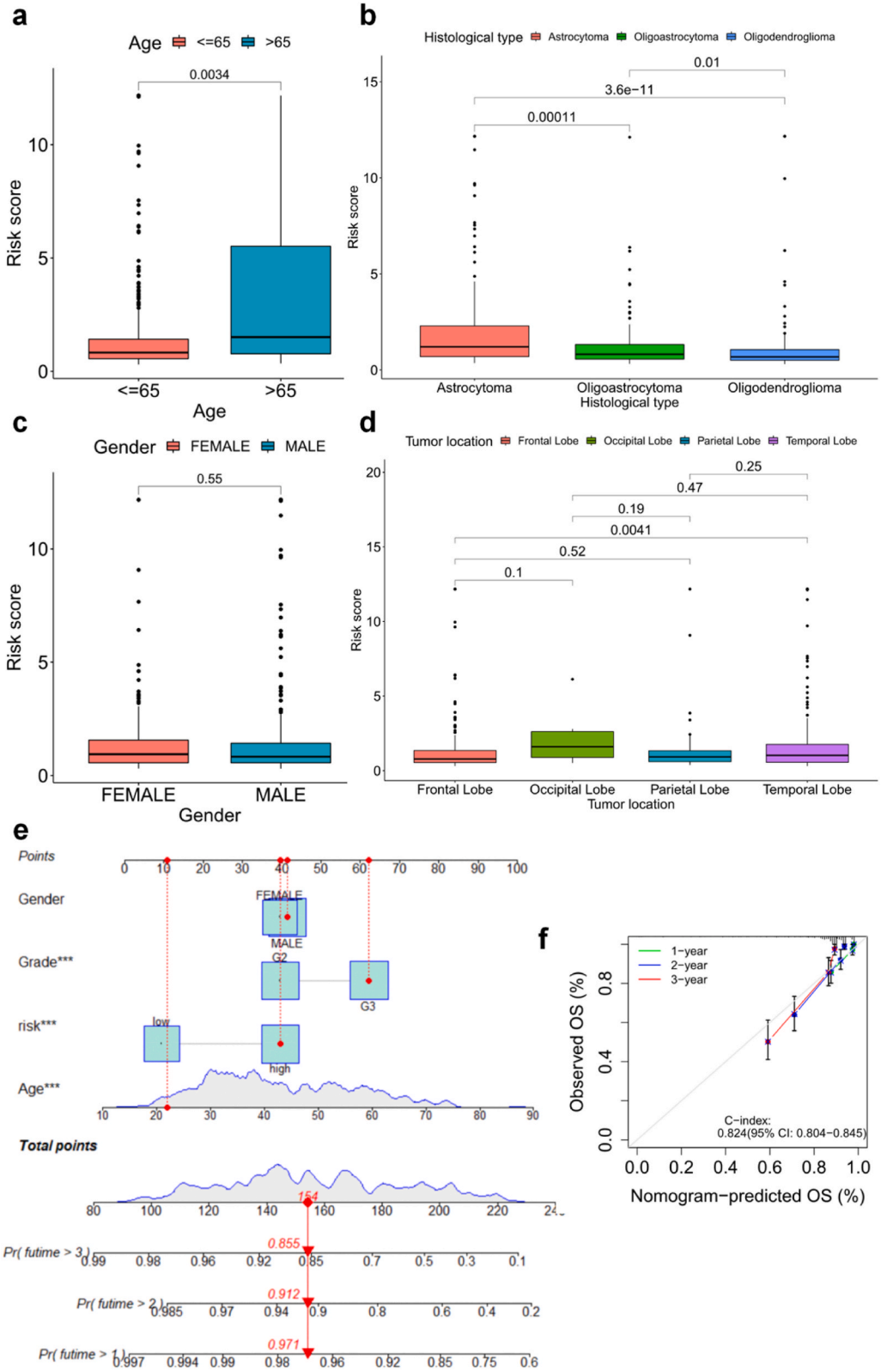


Fig. 5. Association between clinical information and risk score. (A–D) Box plots illustrating the association of age, sex, histological type, tumor location, and risk score. (E) Nomogram prediction model for 1-year, 3-year, and 5-year OS in patients with LGG. (F) C-index for the nomogram in the TCGA cohort.

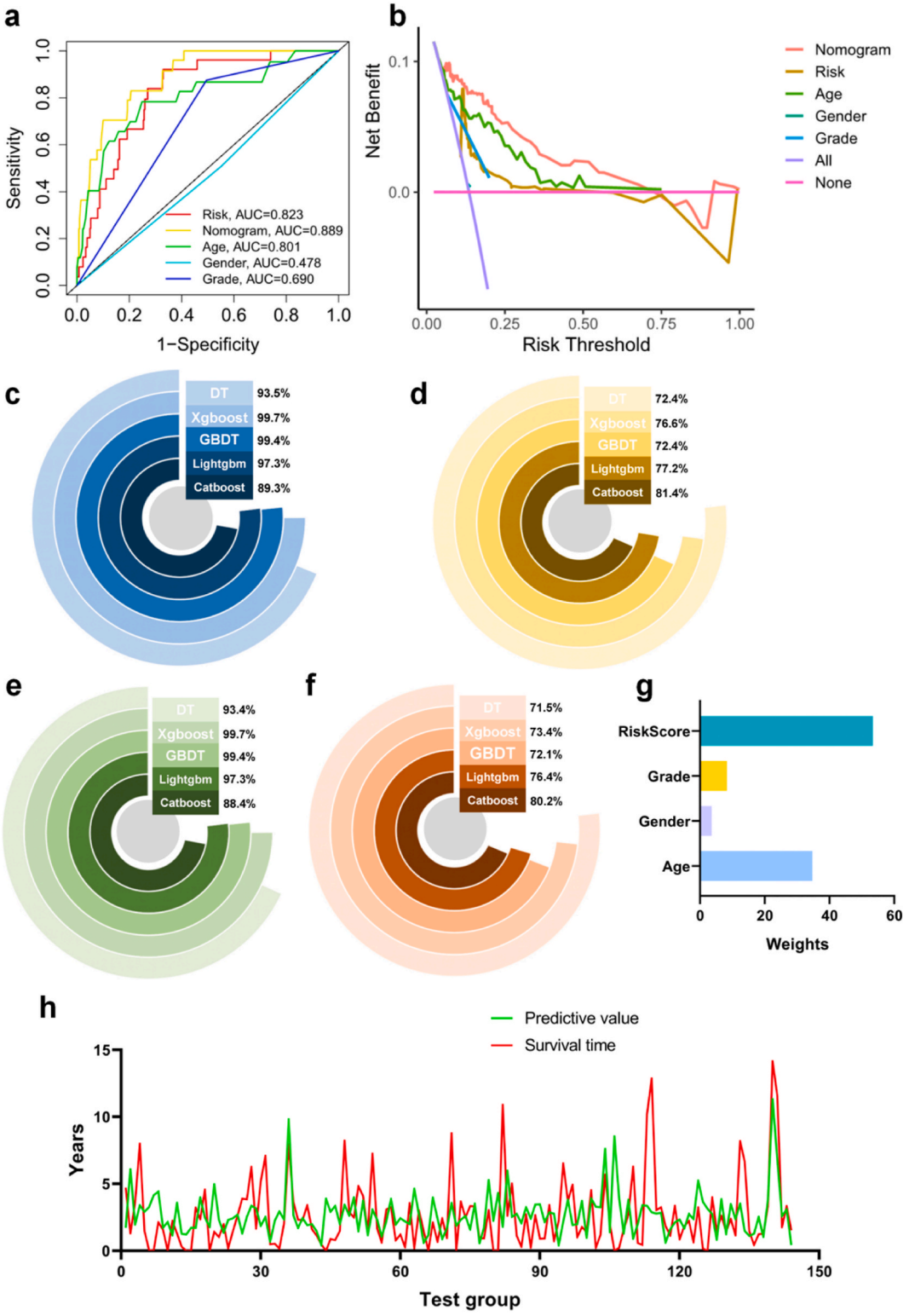
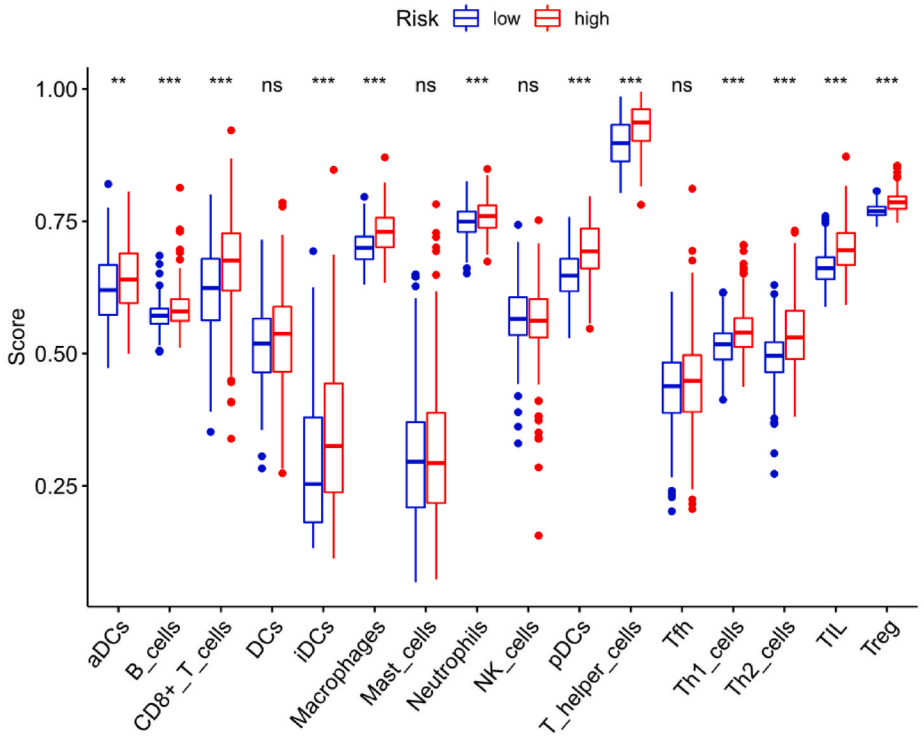
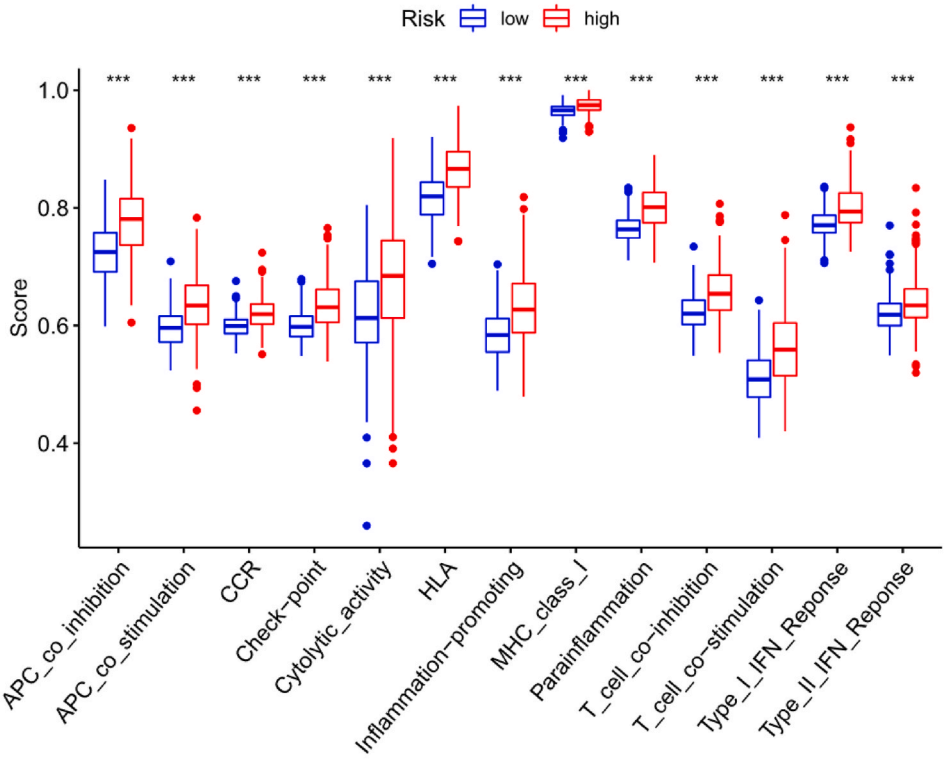


Fig. 6. (A) Risk score and other clinical prognostic factors in the TCGA-LGG cohort, compared with the AUC of ROC curves. (B) Application of DCA to assess the utility of nomograms, risk scores, and clinical indicators. Machine learning algorithms were used to estimate the accuracy and F1 of the diagnosis. Training data were used to train the machine learning model (C: accuracy, E: F1), whereas testing data were used to evaluate model performance (D: accuracy, F: F1). (G) Feature rankings of weight-based importance, obtained after XGBoost fitting. (H) Predicted survival time, according to the XGBoost Algorithm.



A.



B.

(caption on next page)

Fig. 7. Score comparisons performed in 16 immune cell types (A) and 13 types of immune-associated functions (B) between risk groups.

predicting survival duration.

Recent research has indicated that RNA editing increases the risk of many malignancies by inducing abnormal immune responses [8,34]. However, the connection between RNA editing and the immune microenvironment of LGG tumors remains poorly characterized. In this research, the risk score was positively correlated with several immune infiltration cells. Patients with a high risk score had more immune cell infiltration and poorer prognosis, in agreement with earlier research on other types of cancer. Few pharmacotherapeutic treatments are currently available for LGG. This model revealed four possible medications for patients with LGG, including KU-55933, PCI-34051, and Elephantin. KU-55933 is a selective inhibitor of the kinase activity of the protein expressed by Ataxia telangiectasia mutated (ATM), a critical tumor suppressor gene with an essential role in DNA repair [35]. Our investigation

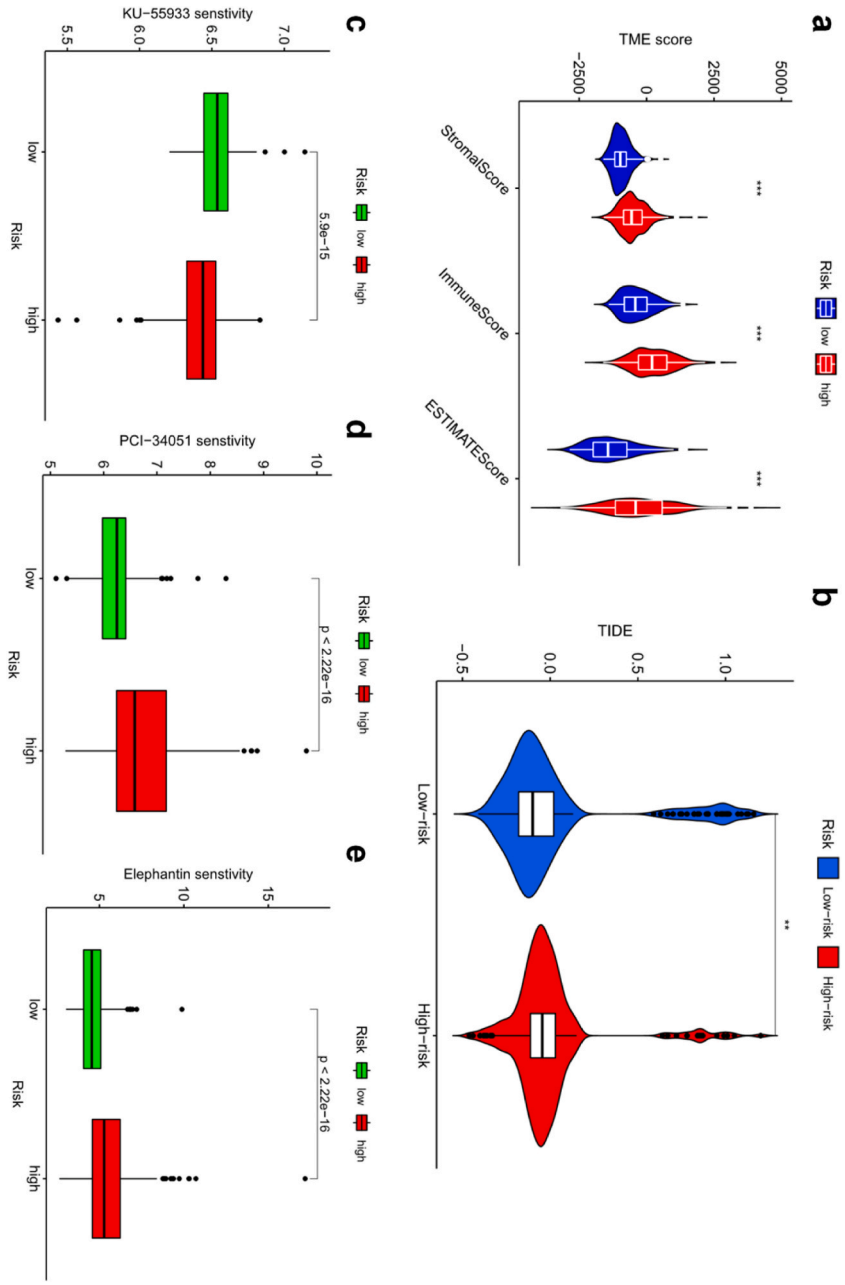


Fig. 8. Box plot of TIDE scores (A) and TME scores (B) for samples from the low- and high-risk groups. (C–E) Differences in IC50 values for PCI-34051, Elephantin, and KU-55933 between the low- and high-risk groups.

revealed that lower risk scores were associated with greater responsiveness of patients to KU-55933 inhibitors. PCI-34051 causes caspase-dependent apoptosis in T-cell lymphomas and leukemia cell lines [36]. Elephantopus elatus is the source of the germacrene sesquiterpenoid cancer inhibitor known as elephantin [37], which functions as an anticancer agent. As their risk scores increased, patients' sensitivity to Elephantin and PCI-34051 increased. In summary, the above findings suggested that our risk model may be used to identify and select patients who will respond to immunotherapy and targeted therapy.

This study used an innovative strategy integrating machine learning methods with bioinformatics to forecast the survival rates of patients with LGG by using RNA editing as a potential biomarker. The application of machine learning methods can reveal hidden patterns and regularities in vast biological data sets, thereby augmenting efficiency and accuracy in the investigation of biological problems [38]. In comparison with traditional biological research methods, the integration of machine learning methods and bioinformatics not only expedites the analysis of large-scale data sets but also aids in the discovery of intricate biological patterns, such as interaction networks, signal transduction pathways, and gene regulation, thereby enhancing understanding and interpretation of life science problems [39]. Furthermore, the combination of machine learning methods and bioinformatics may provide more innovative solutions in domains such as precision medicine, personalized medicine, and drug development. Nonetheless, the present study has several limitations [40]. We were unable to determine the underlying RNA editing mechanisms involved in the regulation of OS in LGG and glioma malignancy, because of technological limitations. This is a preliminary study, and additional consideration of this critical topic should prove valuable.

4. Conclusion

In conclusion, our study comprehensively described the general pattern of RNA editing in LGG and constructed a model consisting of four RNA editing sites as predictors of survival in patients with LGG. More importantly, the above observations were validated with a machine-learning model, thus indicating the value of RNA editing in survival prediction. In addition, given its critical role in immunotherapy and the immune microenvironment, RNA editing may have the potential to mediate LGG progression and may serve as a therapeutic target for the development of novel targeted drugs.

5. Methods

5.1. Datasets

TCGA was queried for information on 532 LGG samples (TCGA-LGG), including mRNA sequencing data, clinical information, and survival statistics. From the synapse [41], the RNA editing profiles of TCGA-LGG patients with cancerous tissues were retrieved. Two CGGA datasets, including 1018 tumor samples, were obtained and used as validation sets. The GEO database entries GSE16011, GSE68848, and GSE109857 were subjected to differential gene expression analysis. GSE16011, GSE68848, and GSE109857 contained 178, 117, and 131 tumor samples, respectively, and 28, 8, and 5 normal samples, respectively. The caret program randomly separated TCGA samples into training and internal validation groups, and GEO and CGGA samples served as the external validation group. Detailed data on the selection criteria for the TCGA and GEO low-grade glioma cohort can be found in the Supplementary Methods.

5.2. Identification of LGG-associated RNA editing sites and calculation of the Alu editing index

We evaluated the significance of each RNA editing site's prognostic value by using Cox regression analysis. RNA editing sites substantially associated with OS were included in the subsequent study. The level of editing was determined by dividing the height of the peak corresponding to G by the sum of the heights of the peaks corresponding to G and A. Detailed data on the identification of RNA editing events can be found in the Supplementary Methods.

5.3. Development and validation of an RNA editing site prognostic signature

The predictive model was then trained by using LASSO Cox regression analysis to prevent overfitting. Four RNA editing sites were identified, their coefficients were calculated with multivariable Cox regression, and the minimum criteria established the parameter for the penalty.

5.4. Independent prognostic value of the risk model

TCGA-LGG patients were divided into high- and low-risk groups by using the median score as the cut-off. Additionally, the "Kaplan-Meier" and "survival" packages were used to evaluate survival in both low- and high-risk groups. The risk score model was then examined with univariate and multivariate Cox regression tests to determine whether it could independently predict the prognosis of patients with LGG, by taking into account sex, age, and clinical stage. Using the "survival ROC" R package, we built a multi-ROC curve to test the predictive accuracy of the risk score and clinical factors for survival. We created a nomogram to assess the survival probability of patients with LGG, according to their clinical features and risk ratings. The C-index was used to evaluate discrimination. Regarding the C-index, a value of 1 represents perfect accuracy in prediction, whereas a value of 0.5 implies random guessing. The Kaplan-Meier approach was used to calculate PFS and pathologic characteristics for both the low- and high-risk groups.

5.5. Machine learning workflow

The data from the LGG cohort were randomly divided into training and testing groups in a ratio of 7:3. Machine learning was performed with the "train" function of the "caret" R package. Five machine learning algorithms were applied to the dataset: Decision Trees C5.0, XGboost, GBDT, Lightgbm, and Catboost. After optimization of parameters according to the features of each algorithm, we determined the optimal solution matching each algorithm. We performed tenfold cross-validation 100 times with shuffling to identify predictors that were selected more than 900 times in a total of 1000 trials (>90%). F1 score = 2 × precision × recall/(precision + recall). The detailed data can be found in the Supplementary Methods.

5.6. Gene Ontology functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes pathway analysis

GO enrichment analysis of differentially expressed genes across low- and high-risk groups was used to identify the molecular activities, biological processes, and cellular components associated with the RNA editing signature. Systematic investigation of KEGG gene functions was performed to determine the signaling pathways associated with the RNA editing signature.

5.7. Immunotherapy and drug susceptibility analysis

The TIDE algorithm was used to determine the likelihood of individual patients responding to immunotherapy. In addition, single-sample gene set enrichment analysis was used to measure and evaluate immune function in both groups. The "pRRophetic" R package was used to evaluate drug susceptibility in the low- and high-risk groups; determine the IC50 values for regularly used chemotherapy or targeted therapeutic drugs; and investigate the utility of the risk score in clinical medication selection.

5.8. Statistical analysis

Bioconductor tools in R version 4.2.0 was used to analyze the data. The Wilcoxon test was used for non-normally distributed data, whereas unpaired Student's t-test was used for regularly distributed data. Kaplan-Meier survival analysis was applied to assess the survival of patients with LGG, according to the RNA editing signature. $P < 0.05$ was considered to indicate statistical significance in each analysis.

Data availability

In this study, data from publicly available sources was examined. This information may be accessed at the following URL: https://portal.gdc.cancer.gov/; https://www.ncbi.nlm.nih.gov/geo/.

Author contributions

Boshen Wang, Peijie Tian, Qianyu Sun, Hengdong Zhang and Lei Han conceived and designed the experiments. Boshen Wang, Baoli Zhu, and Peijie Tian analyzed and interpreted the data. Boshen Wang and Baoli Zhu contributed reagents, materials, analysis tools or data. Boshen Wang, Peijie Tian, Hengdong Zhang and Qianyu Sun wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Scientific Research Project of Jiangsu Health Committee (M2022083); Jiangsu Province's Outstanding Medical Academic Leader program (CXTDA2017029).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge contributions from the TCGA, CGGA and GEO network.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18075.

References

[1] K. Tanaka, T. Sasayama, K. Mizukawa, et al., Combined IDH1 mutation and MGMT methylation status on long-term survival of patients with cerebral low-grade glioma, Clin. Neurol. Neurosurg. 138 (2015 Nov) 37–44.
[2] K.J. Marcus, L. Goumnerova, A.L. Billett, et al., Stereotactic radiotherapy for localized low-grade gliomas in children: final results of a prospective trial, Int. J. Radiat. Oncol. Biol. Phys. 61 (2) (2005 Feb 1) 374–379.
[3] H. Duffau, L. Taillandier, New concepts in the management of diffuse low-grade glioma: proposal of a multistage and individualized therapeutic approach, Neuro Oncol. 17 (3) (2015 Mar) 332–342.
[4] P.H. Wessels, W.E. Weber, G. Raven, et al., Supratentorial grade II astrocytoma: biological features and clinical course, Lancet Neurol. 2 (7) (2003 Jul) 395–403.
[5] P.D. Delgado-Lopez, E.M. Corrales-Garcia, J. Martino, et al., Diffuse low-grade glioma: a review on the new molecular classification, natural history and current management strategies, Clin. Transl. Oncol. 19 (8) (2017 Aug) 931–944.
[6] G. Lombardi, V. Barresi, A. Castellano, et al., Clinical management of diffuse low-grade gliomas, Cancers 12 (10) (2020 Oct 16).
[7] G.J. Downing, Policy perspectives on the emerging pathways of personalized medicine, Dialogues Clin. Neurosci. 11 (4) (2009) 377–387.
[8] T. Christofi, A. Zaravinos, RNA editing in the forefront of epitranscriptomics and human health, J. Transl. Med. 17 (1) (2019 Sep 23) 319.
[9] K. Nishikura, Functions and regulation of RNA editing by ADAR deaminases, Annu. Rev. Biochem. 79 (2010) 321–349.
[10] J. Xu, L. Li, P. Shi, et al., The crucial roles of Bmi-1 in cancer: implications in pathogenesis, metastasis, drug resistance, and targeted therapies, Int. J. Mol. Sci. 23 (15) (2022 Jul 26) 8231.
[11] B.Z. Tan, H. Huang, R. Lam, et al., Dynamic regulation of RNA editing of ion channels and receptors in the mammalian nervous system, Mol. Brain 2 (2009 May 29) 13.
[12] D. Fumagalli, D. Gacquer, F. Rothé, et al., Principles governing A-to-I RNA editing in the breast cancer transcriptome, Cell Rep. 13 (2) (2015 Oct 13) 277–289.
[13] M. Nakano, T. Fukami, S. Gotoh, et al., A-to-I RNA editing up-regulates human dihydrofolate reductase in breast cancer, J. Biol. Chem. 292 (12) (2017 Mar 24) 4873–4884.
[14] S. Maas, S. Patt, M. Schrey, et al., Underediting of glutamate receptor GluR-B mRNA in malignant gliomas, Proc. Natl. Acad. Sci. U. S. A. 98 (25) (2001 Dec 4) 14687–14692.
[15] K. Nishikura, A-to-I editing of coding and non-coding RNAs by ADARs, Nat. Rev. Mol. Cell Biol. 17 (2) (2016 Feb) 83–96.
[16] S. Ben-Aroya, E.Y. Levanon, A-to-I RNA editing: an overlooked source of cancer mutations, Cancer Cell 33 (5) (2018 May 14) 789–790.
[17] S. Wu, Z. Fan, P. Kim, et al., The integrative studies on the functional A-to-I RNA editing events in human cancers, bioRxiv (2022).
[18] L. Han, L. Diao, S. Yu, et al., The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers, Cancer Cell 28 (4) (2015 Oct 12) 515–528.
[19] Y. Zhu, Z. Song, Z. Wang, et al., Protective prognostic biomarkers negatively correlated with macrophage M2 infiltration in low-grade glioma, JAMA Oncol. 2022 (2022), 3623591.
[20] C. Koutsarnakis, E. Neromyliotis, S. Komaitis, et al., Effects of brain radiotherapy on cognitive performance in adult low-grade glioma patients: a systematic review, Radiother. Oncol. 160 (2021 Jul) 202–211.
[21] H. Zhu, L. Huang, S. Liu, et al., REIA: a database for cancer A-to-I RNA editing with interactive analysis, Int. J. Biol. Sci. 18 (6) (2022) 2472–2483.
[22] B.E. Baysal, S. Sharma, S. Hashemikhabir, et al., RNA editing in pathogenesis of cancer, Cancer Res. 77 (14) (2017 Jul 15) 3733–3739.
[23] L. Goldberg, M. Abutbul-Amitai, G. Paret, et al., Alternative splicing of STAT3 is affected by RNA editing, DNA Cell Biol. 36 (5) (2017 May) 367–376.
[24] D. Wu, Y.Y. Zang, Y.Y. Shi, et al., Distant coupling between RNA editing and alternative splicing of the osmosensitive cation channel Tmem63b, J. Biol. Chem. 295 (52) (2020 Dec 25) 18199–18212.
[25] T.H. Chan, A. Qamra, K.T. Tan, et al., ADAR-mediated RNA editing predicts progression and prognosis of gastric cancer, Gastroenterology 151 (4) (2016 Oct) 637–650 e10.
[26] K. Gumireddy, A. Li, A.V. Kossenkov, et al., The mRNA-edited form of GABRA3 suppresses GABRA3-mediated Akt activation and breast cancer metastasis, Nat. Commun. 7 (2016 Feb 12), 10715.
[27] Y.S. Kim, P. Gupta Vallur, R. Phaeton, et al., Insights into the dichotomous regulation of SOD2 in cancer, Antioxidants 6 (4) (2017 Nov 3).
[28] N. Alateyah, I. Gupta, R.S. Rusyniak, et al., SOD2, a potential transcriptional target underpinning CD44-promoted breast cancer progression, Molecules 27 (3) (2022 Jan 26).
[29] G. Tax, A. Lia, A. Santino, et al., Modulation of ERQC and ERAD: a broad-spectrum spanner in the works of cancer cells? JAMA Oncol. 2019 (2019), 8384913.
[30] M. Guo, Y. Peng, A. Gao, et al., Epigenetic heterogeneity in cancer, Biomark. Res. 7 (1) (2019) 1–19.
[31] M. Stern, D. Hexner, J.W. Rocks, et al., Supervised learning in physical networks: from machine learning to learning machines, Phys. Rev. X 11 (2) (2021), 021045.
[32] J.A. Cruz, D.S. Wishart, Applications of machine learning in cancer prediction and prognosis, Cancer Inf. 2 (2007 Feb 11) 59–77.
[33] A.N. Richter, T.M. Khoshgoftaar, A review of statistical and machine learning methods for modeling cancer risk using structured clinical data, Artif. Intell. Med. 90 (2018 Aug) 1–14.
[34] M. Zhang, J. Fritsche, J. Roszik, et al., RNA editing derived epitopes function as cancer antigens to elicit immune responses, Nat. Commun. 9 (1) (2018 Sep 25) 3919.
[35] M.H. Jin, D.Y. Oh, ATM in DNA repair in cancer, Pharmacol. Ther. 203 (2019 Nov), 107391.
[36] J.M. Watters, G. Wright, M.A. Smith, et al., Histone deacetylase 8 inhibition suppresses mantle cell lymphoma viability while preserving natural killer cell function, Biochem. Biophys. Res. Commun. 534 (2021 Jan 1) 773–779.
[37] S.M. Kupchan, Y. Aynehchi, J.M. Cassady, et al., Tumor inhibitors XL. The isolation and structural elucidation of elephantin and elephantopin, two novel sesquiterpenoid tumor inhibitors from Elephantopus elatus, J. Org. Chem. 34 (12) (1969 Dec) 3867–3875.
[38] P.S. Reel, S. Reel, E. Pearson, E. Trucco, E. Jefferson, Using machine learning approaches for multi-omics data analysis: a review, Biotechnol. Adv. 49 (2021 Jul-Aug), 107739.
[39] D.M. Camacho, K.M. Collins, R.K. Powers, J.C. Costello, J.J. Collins, Next-generation machine learning for biological networks, Cell 173 (7) (2018 Jun 14) 1581–1592.
[40] S.J. MacEachern, N.D. Forkert, Machine learning for precision medicine, Genome 64 (4) (2021 Apr) 416–425, https://doi.org/10.1139/gen-2020-0131.
[41] C. Cenci, R. Barzotti, F. Galeano, et al., Down-regulation of RNA editing in pediatric astrocytomas: ADAR2 editing activity inhibits cell migration and proliferation, J. Biol. Chem. 283 (11) (2008 Mar 14) 7251–7260.