

Early 2 factor (E2F) transcription factors contribute to malignant progression and have clinical prognostic value in lower-grade glioma

Haitao Luo^{a,*}, Chuming Tao^{a,b,*}, Xiaoyan Long^{b,*}, Xingen Zhu^{a,c}, and Kai Huang^{a,c}

^aDepartment of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi Province, China; ^bScientific Research Center, East China Institute of Digital Medical Engineering, Shangrao, Jiangxi Province, China; ^cInstitute of Neuroscience, Nanchang University, Nanchang, Jiangxi Province, China

ABSTRACT

Early 2 factor (E2F) genes encoding a family of transcription factors are significantly associated with apoptosis, metabolism, and angiogenesis in several tumor types. However, the biological functions of E2F transcription factors (E2Fs) and their potential involvement in the malignancy of lower-grade glioma (LGG) remain unclear. We explored the effects of the expression of eight E2F family members on the clinical characteristics of LGG based on the Chinese Glioma Genome Atlas (CGGA), The Cancer Genome Atlas (TCGA), and GSE16011 datasets. Two LGG subgroups were identified according to the consensus clustering of the eight E2Fs. We employed the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm for further functional experiments and the development of a potential risk score. Two categories of patients with LGG were identified based on the median risk scores. We then developed a nomogram based on the results of the multivariate analysis. Real-time quantitative polymerase chain reaction (RT-qPCR) and immunohistochemistry were performed to validate the bioinformatics results. Our results indicated that E2F family members were significantly involved in the malignancy of LGG and might serve as effective prognostic biomarkers of the disease.

ARTICLE HISTORY

Received 27 July 2021
Revised 21 September 2021
Accepted
21 September 2021

KEYWORDS

E2F transcription factors; lower-grade glioma; risk model; overall survival; gene expression profile





Introduction

The early 2 factor (E2F) family includes eight critical members (E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, E2F7, and E2F8), which share highly similar DNA-binding domains that interact directly with consensus sequences [1,2]. Members of the E2F family participate significantly in the oscillatory nature of the cell cycle by forming a core transcriptional axis [3,4]. The E2F transcription factors (E2Fs) also exert biological effects on pathways other than cell cycle regulation to contribute to malignant progression, such as apoptosis, angiogenesis, and metabolism [5–8].


The eight E2F family members were classified into three subgroups according to sequence homology and activity (activator protein, atypical repressor, and canonical repressor). Each member displays different expression and functional patterns consistent with their sub-categories [9,10].

The expression of activator proteins (E2F1-3) peaks in the G1-S phase, while atypical repressors (E2F7-8) show increased expression in the late S phase and canonical repressors (E2F4-6) are constitutively expressed during all phases [9]. Several protein complexes mediate the ability of E2Fs to bind DNA and regulate transcription [10]. E2Fs are largely self-regulated, and their subcellular localization is dependent on several factors. Moreover, multiple modifications of E2Fs have been identified that affect their expression, function, stability, and location [11–13].

Accumulating evidence strongly suggests that high expression of E2F family members is significantly associated with malignant progression in several tumor types [14–18]. For example, a strong positive correlation between E2F overexpression and poorer overall survival (OS) in patients with pancreatic tumors has been reported

CONTACT Kai Huang  kaihuang@ncu.edu.cn  Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, No 1 Minde Road, Nanchang, China; Xingen Zhu  ndefy89006@ncu.edu.cn  Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, No 1 Minde Road, Nanchang, China

*These authors contributed equally to this work.

 Supplemental data for this article can be accessed [here](#)

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[14]. E2F1 and E2F2 expression levels were also significantly associated with a pro-angiogenic gene in breast cancer, which potentially conferred a more invasive phenotype [15]. E2F8 was significantly upregulated in lung cancer compared to normal lung tissue and was essential for cancer cell maturation [16,17]. Furthermore, E2F1 down-regulation suppressed tumor growth and invasion in bladder cancer [18]. These findings suggest that E2Fs contribute to the malignant progression of tumors.

While the tumor-promoting roles of E2Fs in multiple cancer types have been reported, limited knowledge exists on their biological functions and potential involvement in the malignancy of lower-grade glioma (LGG). Thus, we systematically investigated the relationship between E2F family members and LGG malignancy based on the clinicopathological characteristics of patients. Patients with LGG in the Chinese Glioma Genome Atlas (CGGA) dataset were sorted into two clusters, cluster1 and cluster2, by consensus clustering analysis based on E2Fs expression. The cluster1/2 subgroups affected the prognosis and clinical characteristics of LGG and showed strong associations with many critical biological processes, signaling pathways, and hallmarks of malignant LGG. We further selected four E2Fs using least absolute shrinkage and selection operator (LASSO) Cox regression analysis to derive risk scores, which facilitated the sorting of patients in both training (CGGA) and validation (The Cancer Genome Atlas (TCGA) and GSE16011) datasets into two categories. Finally, real-time quantitative polymerase chain reaction (RT-qPCR) and immunohistochemistry assays were performed to validate the bioinformatics results. Our collective results implicated the involvement of the eight E2Fs in malignant progression and might serve as prognostic biomarkers for LGG.

Materials and methods

Data collection

The RNA-seq transcriptome data and clinicopathological data of LGG samples were obtained from the CGGA (n = 443) (<https://www.cgga.org.cn/>), TGGA (n = 462) (<https://portal.gdc.cancer.gov/>), and GSE16011 datasets (n = 109) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16011>). The clinicopathological information is presented in Supplementary Table 1. Samples with incomplete data on survival time and survival status were excluded from the survival analysis. We normalized the RNA-seq data obtained from the CGGA, TCGA, and GSE16011 datasets.

Selection of E2F family members

We selected eight E2F genes based on published literature [1,2] and mRNA expression data acquired from the CGGA, TCGA, and GSE16011 datasets. Next, we compared the translation levels of the E2Fs in brain and central nervous system (CNS) cancer samples to those of the corresponding normal control samples based on the ONCOMINE dataset (<https://www.oncomine.org/>) [19]. We also explored the mutation rates of the eight E2Fs and investigated their relationships with tumor immunology.

Bioinformatic analysis

To explore the potential functions of the eight E2Fs in LGG, the R package ‘limma’ was applied to determine the associations between E2Fs expression patterns and the clinical characteristics of samples from the CGGA, TCGA, and GSE16011 datasets [20]. The relationships between the eight E2F genes were examined via Spearman’s correlation tests. A protein-protein interaction (PPI) network of the eight E2Fs was constructed by the STRING database and Cytoscape software [21]. Next, 443 patients with LGG from the CGGA dataset were sorted into two groups using the R package ‘ConsensusClusterPlus’ [22]. The grouping results were further validated using principal component analysis (PCA). Kaplan-Meier curves were plotted to analyze the OS of patients with LGG [23]. The R packages ‘clusterProfiler’, ‘enrichplot’, and ‘ggplot2’ were applied to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of genes showing differential expression between cluster1 and cluster2 [24]. Gene set enrichment analysis (GSEA) was then conducted to validate the functions of these differentially expressed genes.

Univariate Cox regression analysis was performed to determine the prognostic value of the E2F genes. We employed LASSO Cox regression algorithm for further functional experiments and the development of a potential risk score [25]. The minimum criteria were utilized to define four genes and select the optimal penalization coefficient lambda. We calculated the risk scores for all patients in the CGGA, TCGA, and GSE16011 datasets using the following formula: risk score = [E2F4 expression * (0.1612)] + [E2F7 expression * (0.3341)] + [E2F3 expression * (0.0687)] + [E2F2 expression * (0.3455)], in which we determined the coefficients by the minimum criteria and the transformed relative expression value of each selected E2Fs could be found in the datasets. Two categories of patients with LGG were identified based on the median risk scores. We employed a time-dependent receiver operating characteristic (ROC) curve to evaluate the prediction efficiency of our prognostic risk model [26]. We also developed a nomogram based on the results of multivariate analysis using the R packages ‘rms’ and ‘foreign’. The nomogram performance was evaluated by concordance index (C-index) and by comparing nomogram-predicted versus observed Kaplan-Meier estimates of survival probability [27].

Use of RT-qPCR and immunohistochemistry assay to validate bioinformatics results

Normal brain tissues (NBT) and LGG tissues of patients were obtained from the Department of Neurosurgery of the Second Affiliated Hospital of Nanchang University between January 2017 and January 2021. Total RNA was isolated from the frozen tissue specimens using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer’s instructions. RT-qPCR was performed using a LightCycler® 480 Real-Time PCR System. Four selected E2Fs (E2F7, E2F4, E2F3, and E2F2) were assayed by qPCR on an Applied Biosystems real-time instrument using three-step amplification. The gene expression levels were measured using the comparative cycle threshold ($\Delta\Delta C_t$) method. The sequences of the forward and reverse primers for the four E2Fs family members are listed in Supplementary Table 2. All

samples were repeated in triplicate. We also performed an immunohistochemistry assay on human tissues as previously described [28].

Statistical analysis

Wilcoxon tests were applied to compare the risk scores and eight E2Fs between pairs of subtypes with the different clinical characteristics in LGG. Univariate and multivariate Cox regression analyses were used to determine the independent prognostic value of the risk scores and the nomogram, and Kaplan-Meier survival curves were generated to compare the OS of patients with LGG between different categories. Statistical analyses were conducted using SPSS for Windows, version 16.0 (SPSS Inc., Chicago, IL, USA), R software v3.6.3 (<http://www.r-project.org/>), and Prism 8.0 (GraphPad Software, Inc). Data were considered significant at $P < 0.05$.

Results

The present study explored the effects of eight E2Fs on the clinical characteristics of LGG and their potential functions based on the CGGA, TCGA, and GSE16011 datasets. We then employed the LASSO Cox regression algorithm to develop a potential risk score. Two categories of patients with LGG were identified based on the median risk scores. A nomogram was formulated based on the results of the multivariate analysis. RT-qPCR and immunohistochemical analyses were used to validate the bioinformatics results.

Relationships between E2Fs expression and the clinical characteristics of patients with LGG

Accumulating evidence suggests that E2Fs are significantly associated with apoptosis, metabolism, and angiogenesis in several tumor types [14–18]. To explore the potential involvement of E2F members in the malignancy of LGG, we systematically analyzed the associations between the expression patterns of the eight E2Fs and different clinicopathological features. To this end, we examined 443 patients with LGG from the CGGA dataset as the training set, 462 and 109 patients with LGG from the TCGA and GSE16011 datasets,

respectively, as the validation set (Supplementary Table 1). A heatmap was generated to explore the associations between E2Fs expression and WHO grades in the training (CGGA) and validation (TCGA and GSE16011) datasets (Figure 1(a-b); Supplementary Figure 1(a)). Most E2Fs were critically associated with WHO grade. Our results

showed marked upregulation of E2F1, E2F2, E2F3, E2F6, E2F7, and E2F8 in WHO III glioma (Figure 1(c-d); Supplementary Figure 1(b)). We further explored the correlations between E2Fs expression and isocitrate dehydrogenase (IDH) status in the training (CGGA) and validation (TCGA) datasets (Figure 1(e-f)), which revealed

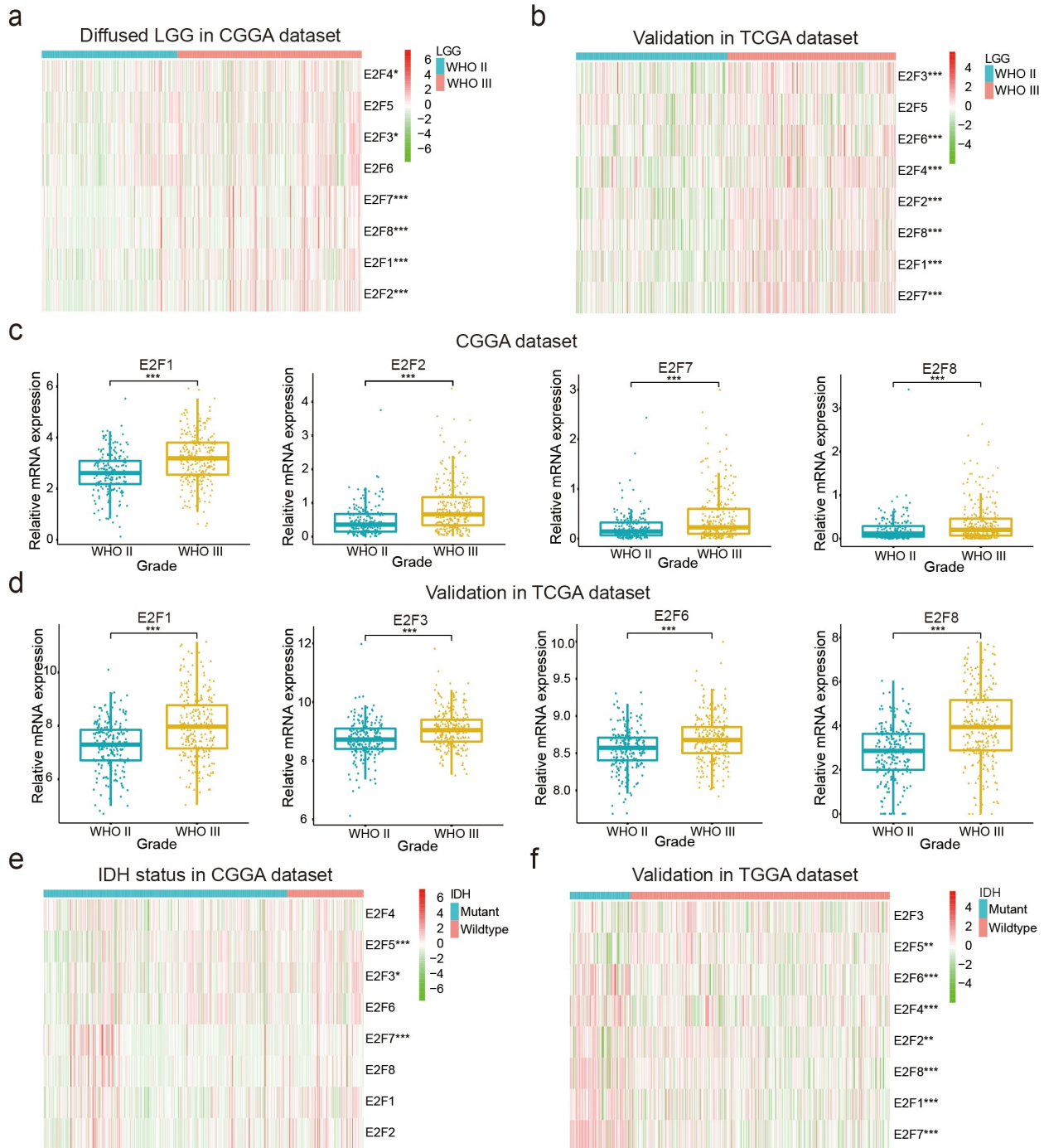


Figure 1. Correlations between the expression levels of E2F members and different clinical characteristics of patients with LGG. (a-d): Expression of eight E2F genes in LGG of different WHO grades. (e-f): Expression of eight E2F genes in patients with differential IDH status from the CGGA (e) and TCGA datasets (f). * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

that wildtype IDH status was positively correlated with E2F3, E3F5, and E2F7 expression in the CGGA dataset, and with E2F1, E2F2, E2F4, E2F5, E2F6, E2F7, and E2F8 expression in the TCGA dataset. The difference in the correlation of wildtype IDH status and E2F expression profile between the two databases may be due to the many samples with missing IDH status in the CGGA dataset (Supplementary Table 1).

To explore whether E2Fs family members could serve as biomarkers for predicting the prognoses of LGG, we compared the transcript levels of the eight E2Fs between brain glioma and normal brain tissue specimens based on the ONCOMINE dataset. E2F5, E2F7, and E2F8 were significantly upregulated in brain and CNS cancers relative to normal brain samples, while E2F1 and E2F3 were more highly expressed in normal brain samples in the ONCOMINE dataset (Supplementary Figure 2a). We also explored the mutation rates of the eight E2Fs based on the TCGA and found very low rates for all E2Fs, indicating that the different expression levels of the eight E2Fs were not caused by genetic alterations (Supplementary Figure 2(b)). We then investigated the relationships between the eight E2Fs and tumor immunology and found that E2F1, E2F3, E2F5, E2F6, and E2F7 expression was positively correlated with many immune cells, which might explain their favorable prognostic value (Supplementary Figure 2(c)).

Categories based on consensus clustering of E2Fs are significantly correlated with LGG malignancy

To explore the interactions of the eight E2Fs, we analyzed their correlations according to their mRNA expression levels in the CGGA dataset. The expression levels of E2F1, E2F2, E2F3, E3F4, E2F5, E2F6, E2F7, and E2F8 were positively correlated with LGG (Figure 2(a)). Meanwhile, the PPI network of eight E2Fs constructed using STRING and Cytoscape showed that the E2F1 was a hub gene in the interactions of the eight E2Fs (Supplementary Figure 1(c)), indicating that E2F family members are linked to the promotion of malignant progression in LGG. The application of consensus clustering analysis led to the classification of patients into two categories based on the expression patterns of the eight E2F family members in the CGGA dataset

($n = 443$). A K value of 2 from among K values of 2–9 was selected using cumulative distribution function (CDF) curves and SigClust analysis (Figure 2(b-c)). The consensus score matrix of the expression levels of the eight E2Fs based on the CGGA dataset was generated to evaluate the influence of correlations within and between groups (Supplementary Figure 3), which led to the identification of two categories according to the different clinical characteristics of patients with LGG (Figure 2(d)). E2Fs transcript levels differed significantly between the two categories, as determined by PCA (Figure 2(e)). Furthermore, Kaplan-Meier analysis disclosed that compared to cluster2, the patients in cluster1 had poorer OS ($P < 0.001$; Figure 2(f)), based on the high levels of E2F1-8 in cluster1 (Supplementary Figure 4). In addition, the prevalence of high grade, old age, mutant-type IDH status, and 1p19q non-codeletion status were higher in cluster1 than those in cluster2 (Supplementary Table 3).

Molecular mechanisms by which E2F family members contribute to malignant progression in LGG

We identified genes that were significantly differentially expressed between cluster1 and cluster2 and further analyzed the malignancy-related biological processes enriched in cluster1 associated with these highly expressed genes were. We observed enrichment of genes involved in mitotic nuclear division, nuclear division, organelle fission, DNA replication, regulation of nuclear division, chromosomal region, spindle and microtubule motor activity (Figure 3a). KEGG pathway analysis further revealed a significant enrichment of genes involved in herpes simplex virus 1 infection and the cell cycle (Figure 3(a)). We also applied GSEA to identify the hallmarks of malignant tumors. Our results showed that Wnt beta-catenin signaling (NES = 1.75; normalized $P = 0.011$), mitotic spindle (NES = 1.84; normalized $P < 0.001$), unfolded protein response (NES = 1.65; normalized $P < 0.05$), and P13 K/Akt signaling pathway (NES = 1.59; normalized $P < 0.05$) were markedly correlated with cluster1 (Figure 3(b)). Thus, cluster1, as identified from consensus clustering, was a positively associated with malignancy progression in LGG.

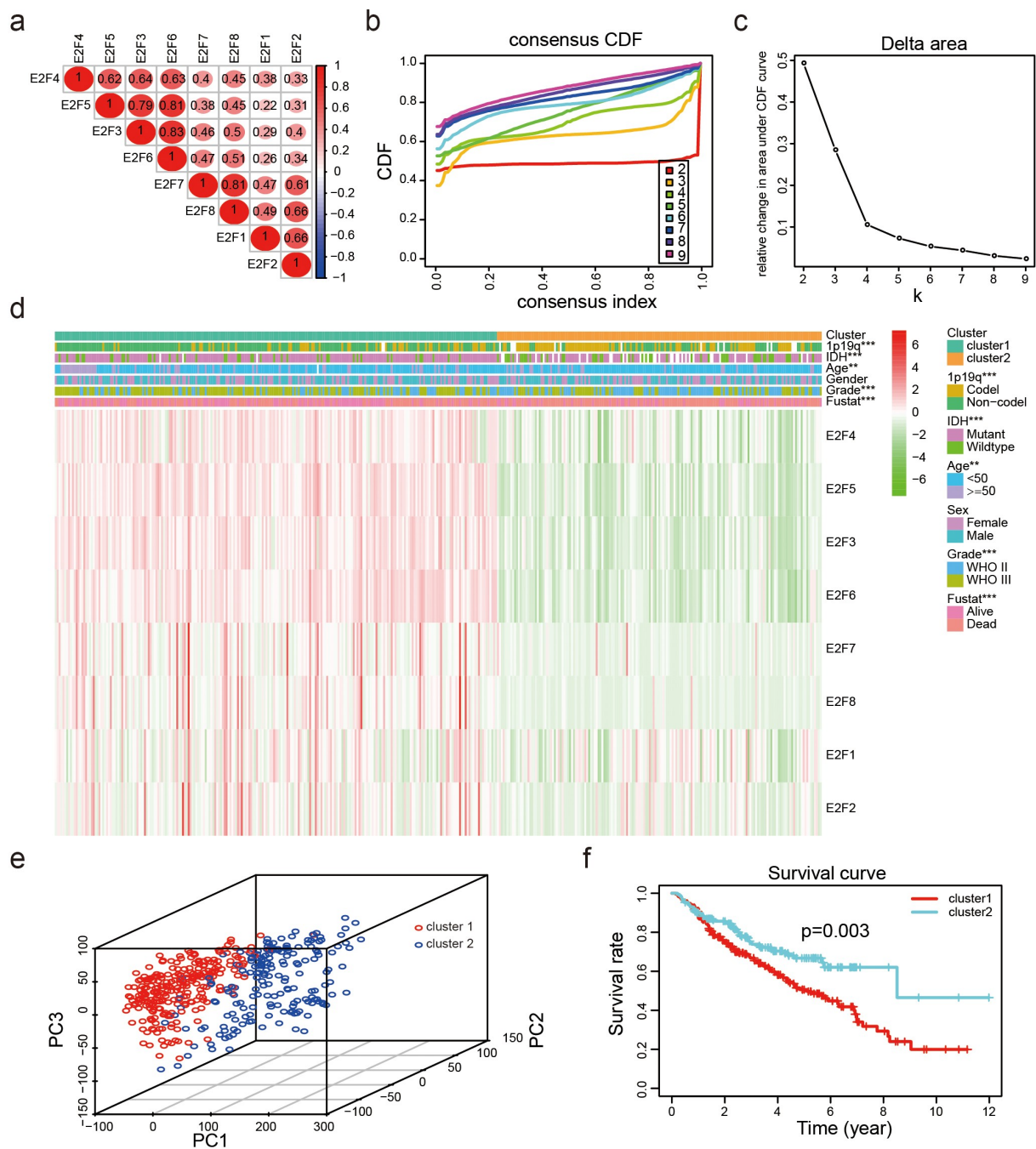


Figure 2. Two categories of patients based on distinct clinical characteristics and OS according to the gene expression of eight E2Fs in the CGGA dataset. (a): Spearman's correlation analysis of the E2F family members. (b): CDFs for $K = 2-9$. (c): Delta areas under the CDF curves for $K = 2-9$. (d): Clinical characteristics of the two clusters defined based on the consensus expression of the eight members of the E2F family. (e): PCA of total RNA expression profile in the CGGA datasets. (f): Kaplan-Meier curves for samples in the CGGA datasets.

Prognostic value of E2Fs and a risk model derived from four selected E2F members

We explored the potential correlations between the eight E2Fs and prognosis using univariate Cox regression analysis of samples from the

CGGA dataset. Expression of the eight E2Fs was strongly correlated with OS ($P < 0.01$) and all eight E2Fs were identified as high-risk genes (hazard ratio > 1 ; Figure 4(a)). Next, LASSO Cox regression analysis was performed using the CGGA

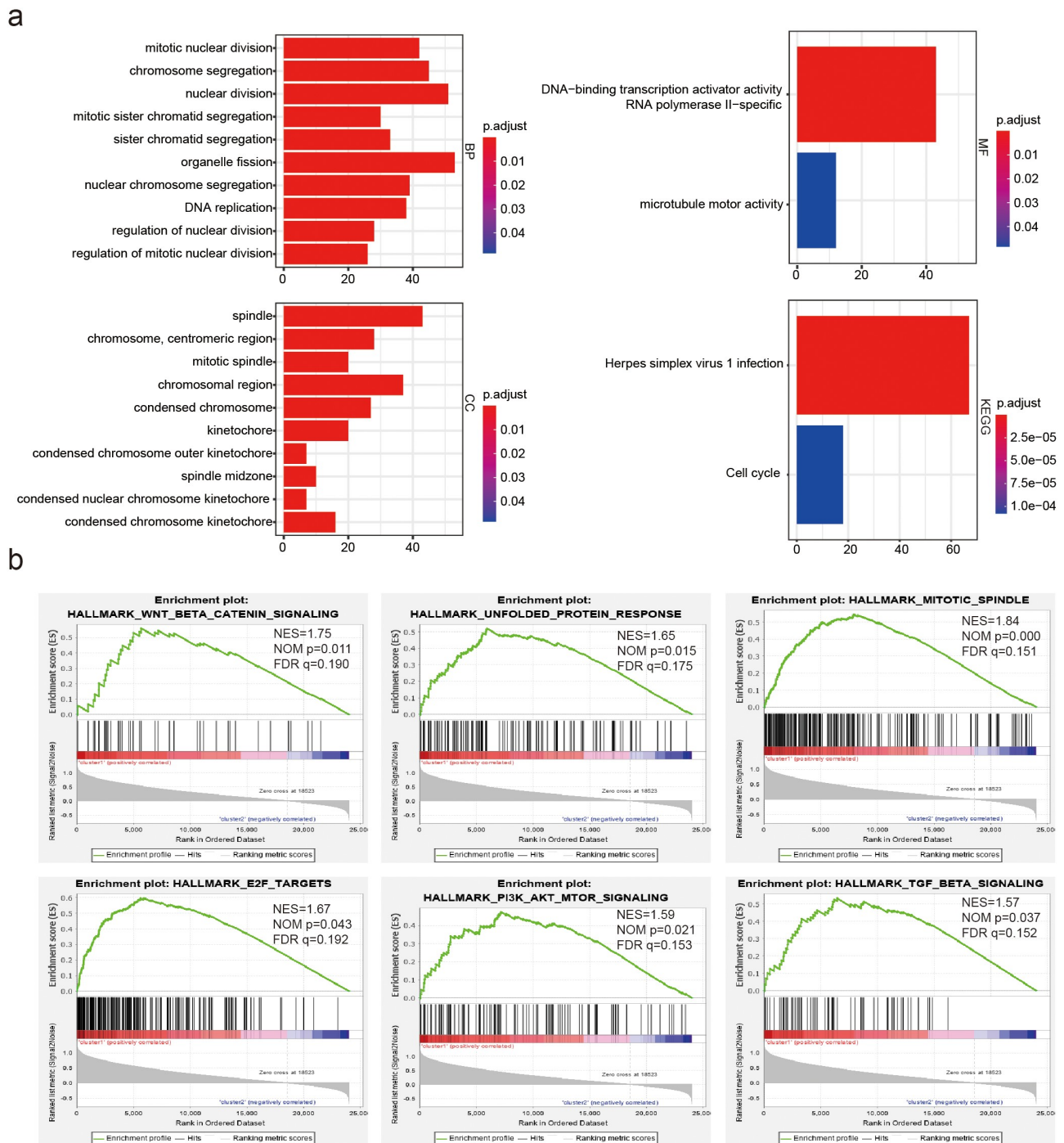


Figure 3. Functional annotations of differentially expressed genes between cluster1 and cluster2. (a): Functional annotations of differentially expressed genes via GO and KEGG pathway analyses. (b): Malignant hallmarks enriched in cluster1 determined using GSEA.

dataset as a training set to construct an improved prognostic risk model according to E2Fs expression patterns (Figure 4(a)). Four genes were selected based on the minimum criteria to establish the risk score and the coefficients derived from the LASSO Cox regression analysis were used to calculate risk scores for all patients with

LGG in both the training (CGGA) and validation (TCGA and GSE16011) datasets. The samples were categorized into low-risk and high-risk subgroups based on median risk scores from both datasets (Supplementary Table 4). Kaplan-Meier survival curves revealed a significant difference in OS between the low-risk and high-risk categories

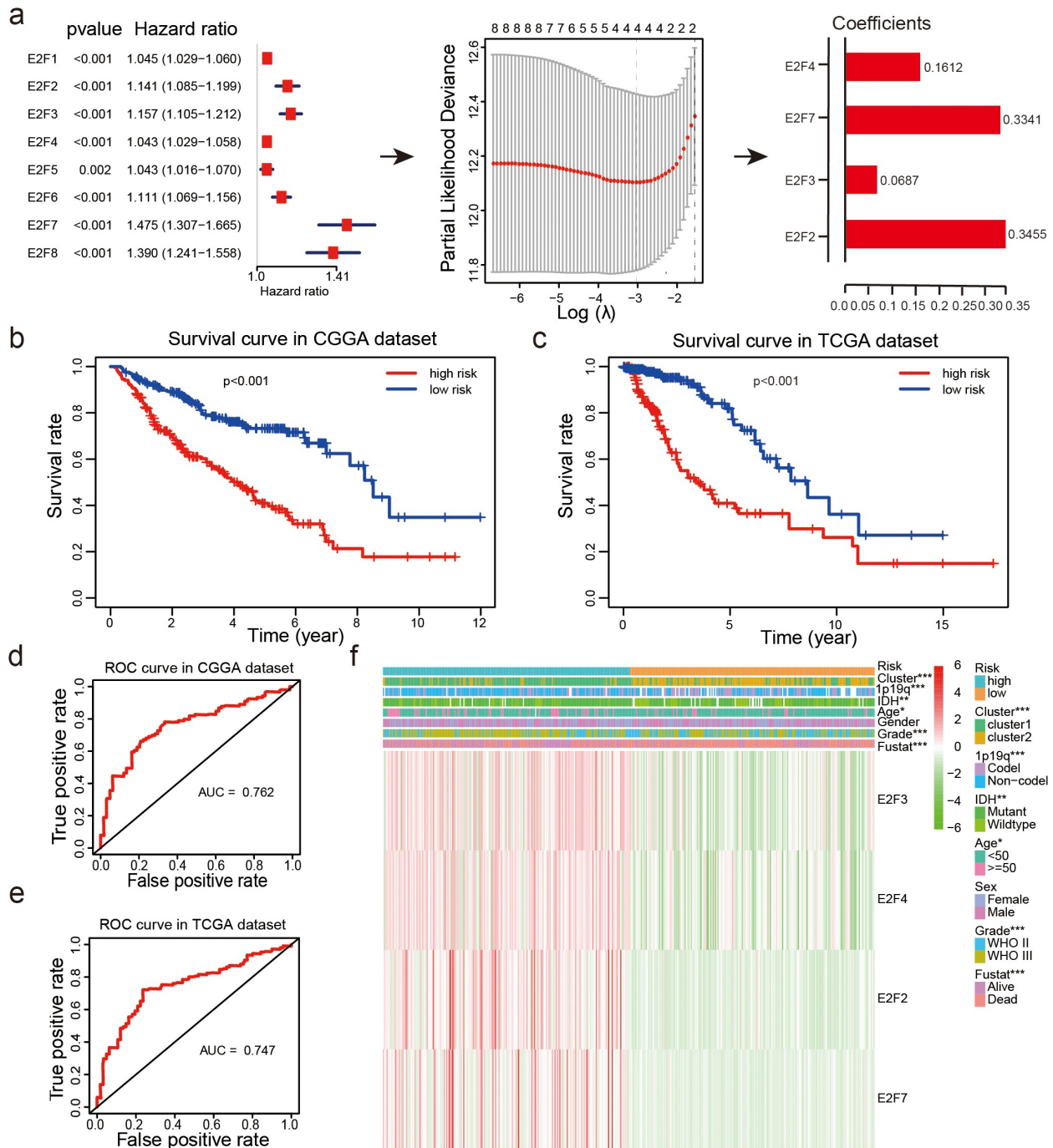


Figure 4. Risk models derived from expression patterns of four E2F genes. (a): The process of risk model construction. (b-c): Kaplan-Meier OS curves for patients from the CGGA (b) and TCGA (c) datasets categorized into two groups based on the median risk scores. (d-e): ROC curve analysis of the predictive efficiency of our risk model in the CGGA (d) and TCGA (e) datasets. (f): Heatmap of genes corresponding to four E2F genes and the distributions of clinical characteristics in the two subgroups. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

in both the training (CGGA) and validation (TCGA and GSE16011) datasets ($P < 0.001$; Figure 4b-c; Supplementary Figure 1(d)). We further evaluated the predictive value of our risk

model with the aid of ROC curve analysis. The areas under the ROC curves (AUCs) for OS of patients was 0.762 in the CGGA dataset, 0.747 in the TCGA dataset, and 0.679 in the GSE16011

dataset, indicating that our risk model could accurately predict the OS of patients with LGG (Figure 4(d-e); Supplementary Figure 1(e)).

The risk model has good prognostic performance and is significantly associated with the clinicopathological features of LGG

Examination of the relationships between the risk scores and clinicopathological features of LGG from the CGGA datasets revealed significant differences between risk scores in groups stratified by cluster ($P < 0.001$), WHO grade ($P < 0.001$), age ($P < 0.05$), 1p/19q status ($P < 0.001$), survival status ($P < 0.001$), and IDH status ($P < 0.01$) (Figure 4(f); Supplementary Table 4; Figure 5(a-f)). Overall, our risk prognostic model accurately predicted the OS and clinical characteristics of patients with LGG. ROC curves were generated to validate the ability of our risk model to reliably predict the survival rate (AUC = 0.709), cluster1 category (AUC = 0.874), IDH mutant status (AUC = 0.674), and 1p19q codel status (AUC = 0.634; Figure 5(g-j)).

To establish whether the risk scores represented an independent prognostic index, we performed univariate and multivariate Cox regression analyses on both the training (CGGA) dataset and validation (TCGA and GSE16011) datasets. Univariate Cox regression analysis showed that WHO grade, age, IDH status, 1p/19q status, and risk score were strongly correlated with the OS of patients in the CGGA dataset. Multivariate Cox regression analysis revealed that WHO grade ($P < 0.001$), age ($P < 0.05$), IDH status ($P < 0.01$), 1p/19q status ($P < 0.01$), and risk score ($P < 0.001$) remained significantly correlated with OS (Figure 5(k)). Similar conclusions were reached from multivariate Cox regression analyses of the TCGA and GSE16011 datasets (Figure 5(l), Supplementary Figure 1(f)), which showed significant correlations between WHO grade ($P < 0.05$), age ($P < 0.01$), 1p19q ($P < 0.01$), and risk scores ($P < 0.01$) and OS. Therefore, the risk scores derived from the four selected E2Fs genes showed strong prognostic value in LGG.

To explore the prognostic value of our risk model in LGG of different WHO grades, we analyzed the correlations between risk scores and OS

via the Kaplan -Meier method, which revealed poor OS in the high-risk subgroups compared to those in the low-risk subgroups of different WHO grades from the CGGA dataset (Figure 5(m-n)). These collective findings indicated that our risk model had significant prognostic value for LGG of different WHO grades.

Development of a nomogram based on clinicopathological features and risk score for predicting the prognosis of LGG

Then a nomogram was constructed for predicting LGG prognosis, integrating grade, age, IDH status, 1p19q status and risk scores based on the CGGA dataset (Figure 6(a)), and showed a C-index for survival prediction of 0.735. The calibration curve for the probabilities of survival at 2-, 3-, and 5- years showed optimal agreement between nomogram prediction and the actual observed outcomes (Figure 6(b-d)). The AUC for OS was 0.692 at 2 years, 0.642 at 3 years, and 0.663 at 5 years, showing the reliable predictive ability in the CGGA dataset (Figure 6(e-g)).

The mRNA and protein expression of four selected E2F members in LGG

We performed RT-qPCR and immunohistochemistry assays to assess the four E2F members to construct the risk scores. The RT-qPCR assay showed different mRNA expression level among the four E2F members (E2F2, E2F3, E2F4, and E2F7) between NBT and LGG (Figure 7(a-d)). The results of the immunohistochemistry assay also showed different protein expression levels of the four E2Fs between NBT and LGG tissue samples, consistent with the results of the bioinformatics analysis (Figure 7(e-h)).

Discussion

Glioma is the most common lethal CNS tumor type in adults [29–31]. Precise surgical resection with preoperative imaging, adjuvant chemotherapy and postoperative radiotherapy constitute the standard treatment strategies for gliomas [32–35]. However, the OS rate of patients with glioma remains poor

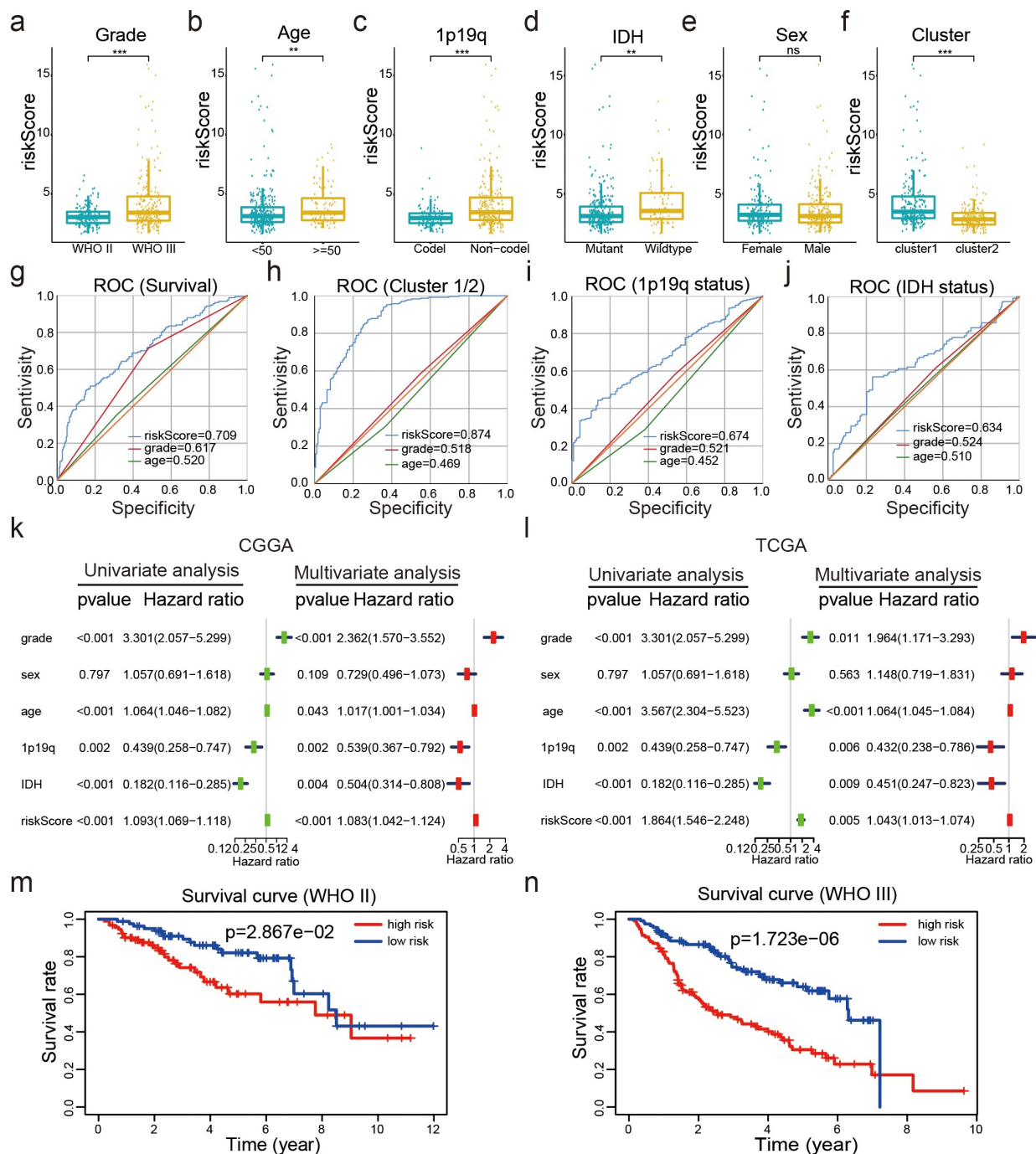


Figure 5. Correlations between risk scores, clinical characteristics and clusters. (a-f): Distributions of risk scores stratified by WHO grade (a), age (b), 1p/19q status (c), IDH status (d), sex (e), and cluster (f). (g-j): Predictive efficiency of risk score, WHO grade and age relative to the survival rate (g), cluster1 group (h), IDH status (i) and 1p/19q codon status (j). (k-l): Relationships between clinical characteristics and OS of patients in the CGGA (k) and TCGA (l) datasets determined via univariate and multivariate Cox regression analyses. (m-n): Kaplan-Meier analysis of gliomas different WHO grades from the CGGA dataset. ns $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

following standard therapy. LGG can evolve into high-grade gliomas and develop resistance to chemotherapy. Accumulating evidence suggests that E2Fs have significant prognostic value in multiple

cancer types and may serve as potential diagnostic biomarkers [4,16-18].

Among the eight E2F family members, up-regulated E2F1 expression gene reportedly

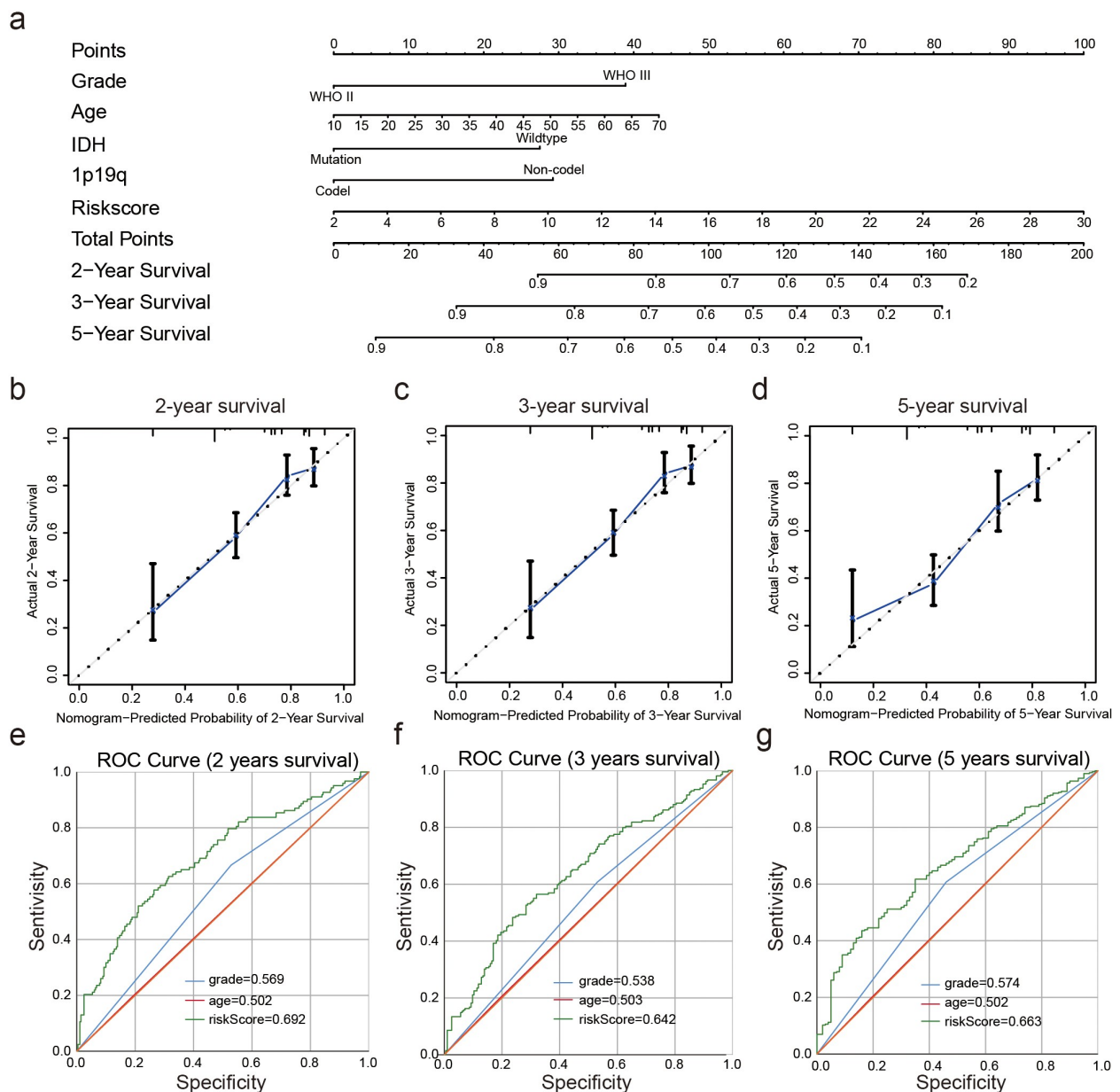


Figure 6. Construction and assessment of a nomogram to predict patient' OS. (a): Nomogram based on the clinical characteristics and risk scores for predicting patient survival. (b-d): Calibration curve for predicting patient survival at 2 years (b), 3 years (c), and 5 years (d). (e-g): The predictive efficiency of the risk scores, WHO grade, and age showed by ROC curves based on 2- (e), 3- (f), and 5- (g) year survival rates.

induced apoptosis in glioma cell lines [36,37], while E2F3 knockdown inhibited proliferation and migration in glioma [38]. Yu et al. investigated the expression of E2Fs and their prognostic value in high-grade glioma based on the TCGA dataset and identified E2F7 and E2F8 as novel potential prognostic markers linked with aggressive oncogenic processes [39]. Yin et al. also showed that E2F7 expression was significantly associated with poor prognosis in patients with

gliomas [40]. However, conflicting data suggesting that atypical repressor E2Fs may have tumor suppressive and oncogenic functions in different tumor types have been reported [41–44]. For instance, E2F8 knockdown inhibited tumor formation in xenograft models of hepatocellular carcinoma and lung cancer in one study but significantly suppressed stress-induced skin cancer in another study using knockout mice [41]. E2F3B, an isoform of E2F3, usually acts as

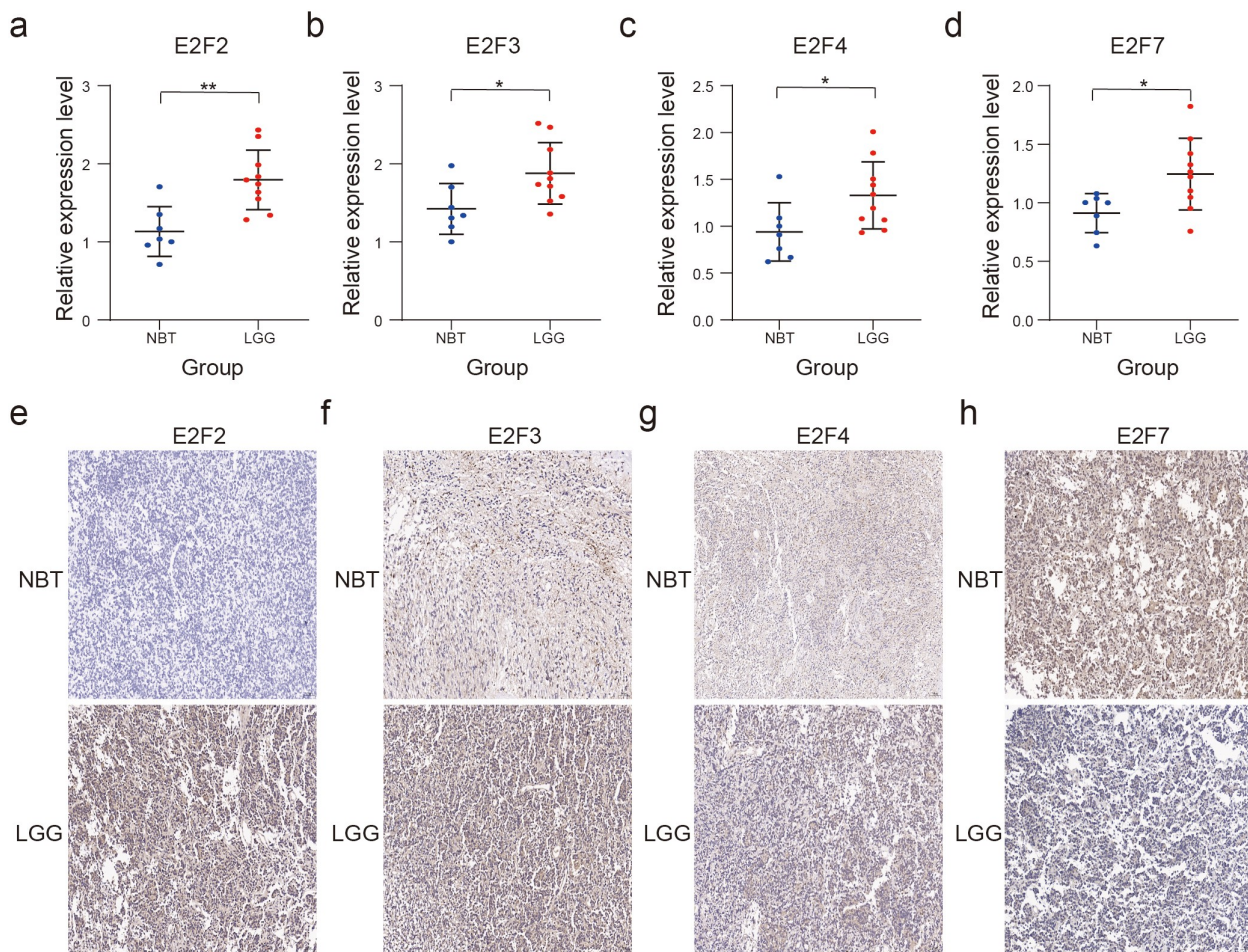


Figure 7. Validation of four selected E2F members by RT-qPCR and immunohistochemistry analysis. (a-d): Comparative E2F2 (a), E2F3 (b), E2F4 (c), and E2F7 (d) mRNA expression levels in NBT and LGG tissues. (e-h): Comparative E2F2 (e), E2F3 (f), E2F4 (g), and E2F7 (h) protein expression levels in NBT and LGG tissues by immunohistochemistry assay. ns $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

a canonical repressor but has also been reported as an oncogene in some tumor types [42–44].

Some studies have explored the mechanism of E2F transcription factors in the malignancy of glioma. Manzano et al. demonstrated that E2F1 modulated the expression of the anti-apoptotic molecule Bcl-2 and suggested that up-regulation of Bcl-2 might favor the oncogenic role of E2F1 and the other E2F family members in human glioma cells [45]. Zhang et al. reported that E2F2 enhanced PFKFB4 expression and regulated P13K/AKT phosphorylation to promote glioma malignancy [46].

In the present study, we evaluated the associations between the expression levels of eight E2F family members and the clinical characteristics of patients with LGG in training (CGGA) and validation (TCGA and GSE16011) datasets. As expected,

most E2Fs (E2F1, E2F2, E2F3, E2F4, E2F7, and E2F8) were strongly associated with increasing LGG WHO grade, indicating that E2F gene expression is closely related to the clinical characteristics of malignancy. Our data based on samples from the ONCOMINE dataset, and the correlation with tumor immunology further support the utility of these eight E2F genes as biomarkers for predicting the malignancy of LGG.

We used consensus clustering analysis to sort patients with LGG into two clusters and found that cluster1 was strongly positively associated with poor OS, WHO grade III, IDH wild-type, older age, and 1p/19q non-codeletion of LGG. Multiple biological processes and signaling pathways involving E2F family members have been identified, including cell cycle [47,48], DNA repair [49], DNA replication [50,51], apoptosis [52], and

p53 signaling [53]. In our study, E2Fs were associated with critical processes in LGG, including DNA replication, nuclear division, organelle fission, mitotic spindle, kinetochore, spindle midzone, and DNA-binding transcription activator, as well as signaling pathways underlying malignancy progression such as herpes simplex virus 1 infection and the cell cycle. Furthermore, GSEA revealed associations with several hallmarks of malignant tumors such as P13K-AKT-mTOR signaling, TGF-beta signaling, and unfolded protein response.

To ascertain whether E2F expression can be utilized as an independent prognostic index for the diagnosis of LGG, we derived a prognostic model using four selected E2Fs. Our results showed that a poor OS was strongly associated with the high-risk subgroup in both the training (CGGA) and validation (TCGA and GSE16011) datasets. We also observed a significant correlation between the risk scores of our prognostic model and the clinical characteristics of patients with LGG. Furthermore, our risk scores could serve as an independent prognostic index for the diagnosis of LGG, as determined by univariate and multivariate Cox regression analyses of both training (CGGA) and validation (TCGA and GSE16011) datasets. The risk scores based on the four selected E2Fs showed prognostic value for WHO grade II and III LGG. In addition, we constructed a nomogram to predict the prognosis of LGG, which integrated grade, age, IDH status, 1p19q status, and risk scores. The C-index for survival prediction and the AUC for OS showed the reliable ability of our nomogram in the CGGA dataset. Our results further indicated that the risk scores had clinical prognostic value in LGG.

Conclusion

The expression levels of E2Fs were correlated with malignant clinical characteristics and cancer-related biological processes in LGG. The risk signature of the four selected E2Fs was an independent prognostic index for the diagnosis of LGG. Our results provide significant insights that offer a platform for further in-depth exploration of the specific roles of individual E2Fs in LGG.

Highlights

1. The correlations between E2Fs expression and LGG malignancy were explored.
2. A risk signature of four E2Fs was an independent prognostic index for LGG prognosis.
3. The bioinformatics results were validated by RT-qPCR and immunohistochemistry assays.

Data availability statement

Data for this work were obtained from the CGGA (<http://www.cgga.org.cn/>), TCGA (<https://portal.gdc.cancer.gov/>), GSE16011 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16011>), ONCOMINE (<https://www.oncomine.org/>), and Human Protein Atlas datasets (<https://www.proteinatlas.org/>).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by National Natural Science Foundation of China (Grant no: 82002660 and 81660420); and Jiangxi Province Department of Education Science and technology research project (Grant no. GJJ190018). National Natural Science Foundation of China [82002660]; National Natural Science Foundation of China [Grant no: 81660420]; Key Science and Technology Research Project in Jiangxi Province Department of Education [Grant no. GJJ190018].

Ethics approval

Our study was approved by the Ethics Committee of the second affiliated hospital of Nanchang University. The Examination and Approval No. Review [2016] No. (122).

Author's contributions

Haitao Luo was involved in conception and design and Chuming Tao in data acquisition, Xiaoyan Long performed the RT-qPCR and immunohistochemistry assays. Kai Huang and Xingen Zhu drafted the manuscript and contributed to the revision. All authors edited and approved the final version of manuscript.

References

- [1] Chen HZ, Tsai SY, Leone G. Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat Rev Cancer*. 2009;9(11):785–997.

- [2] Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer. *Nat Rev Cancer*. 2019;19(6):326–338.
- [3] Lammens T, Li J, Leone G, et al. Atypical E2Fs: new players in the E2F transcription factor family. *Trends Cell Biol*. 2009;19(3):111–118.
- [4] Schaal C, Pillai S, Chellappan SP. The Rb-E2F transcriptional regulatory pathway in tumor angiogenesis and metastasis. *Adv Cancer Res*. 2014;121:147–182.
- [5] Shats I, Deng M, Davidovich A, et al. Expression level is a key determinant of E2F1-mediated cell fate. *Cell Death Differ*. 2017;24(4):626–637.
- [6] Teixeira LK, Wang X, Li Y, et al. Cyclin E deregulation promotes loss of specific genomic regions. *Curr Biol*. 2015;25(10):1327–1333.
- [7] Wu M, Seto E. E2f ZJ. 1 enhances glycolysis through suppressing Sirt6 transcription in cancer cells. *Oncotarget*. 2015;6(13):11252–11263.
- [8] Araki K, Kawachi K, Sugimoto W, et al. Mitochondrial protein E2F3d, a distinctive E2F3 product, mediates hypoxia-induced mitophagy in cancer cells. *Commun Biol*. 2019;2(1):3.
- [9] Kent LN, Rakijas JB, Pandit SK, et al. E2f8 mediates tumor suppression in postnatal liver development. *J Clin Invest*. 2016;126(8):2955–2969.
- [10] Logan N, Graham A, Zhao X, et al. E2F-8: an E2F family member with a similar organization of DNA-binding domains to E2F-7. *Oncogene*. 2005;24(31):5000–5004.
- [11] Liban TJ, Thwaites MJ, Dick FA, et al. Structural Conservation and E2F Binding Specificity within the Retinoblastoma Pocket Protein Family. *J Mol Biol*. 2016;428(20):3960–3971.
- [12] Liu H, Tang X, Srivastava A, et al. Redeployment of Myc and E2f1-3 drives Rb-deficient cell cycles. *Nat Cell Biol*. 2015;17(8):1036–1048.
- [13] Saenz-Ponce N, Pillay R, De Long LM, et al. Targeting the XPO1-dependent nuclear export of E2F7 reverses anthracycline resistance in head and neck squamous cell carcinomas. *Sci Transl Med*. 2018;10(447):eaar7223.
- [14] Lan W, Bian B, Xia Y, et al. E2F signature is predictive for the pancreatic adenocarcinoma clinical outcome and sensitivity to E2F inhibitors, but not for the response to cytotoxic-based treatments. *Sci Rep*. 2018;8(1):8330.
- [15] Hollern DP, Honeysett J, Cardiff RD, et al. The E2F transcription factors regulate tumor development and metastasis in a mouse model of metastatic breast cancer. *Mol Cell Biol*. 2014;34(17):3229–3243.
- [16] Jin DH, Kim Y, Lee BB, et al. Metformin induces cell cycle arrest at the G1 phase through E2F8 suppression in lung cancer cells. *Oncotarget*. 2017;8(60):101509–101519.
- [17] Park SA, Platt J, Lee JW, et al. E2F8 as a Novel Therapeutic Target for Lung Cancer. *J Natl Cancer Inst*. 2015;107(9):djv151.
- [18] Lee SR, Roh YG, Kim SK, et al. Activation of EZH2 and SUZ12 Regulated by E2F1 Predicts the Disease Progression and Aggressive Characteristics of Bladder Cancer. *Clin Cancer Res*. 2015;21(23):5391–5403.
- [19] Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007;9(2):166–180.
- [20] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- [21] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504.
- [22] Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–1573.
- [23] Palazzo M, Templeton M. Statistical interpretation of Kaplan–Meier curves. *Intensive Care Med*. 2007;32(12):2235.
- [24] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS*. 2012;16(5):284–287.
- [25] Hu X, Martinez-Ledesma E, Zheng S, et al. Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma. *Neuro Oncol*. 2017;19(6):786–795.
- [26] Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol*. 2018;63(7):07tr1.
- [27] Qi-Dong X, Yang X, Lu JL, et al. Development and validation of a nine-redox-related long noncoding RNA Signature in renal clear cell carcinoma. *Oxid Med Cell Longevity*. 2020;2020:6634247.
- [28] Hu LP, Zhang XX, Jiang SH, et al. Targeting Purinergic Receptor P2Y2 Prevents the Growth of Pancreatic Ductal Adenocarcinoma by Inhibiting Cancer Cell Glycolysis. *Clin Cancer Res*. 2019;25(4):1318–1330.
- [29] Bates A, Gonzalez-Viana E, Cruickshank G, et al. Primary and metastatic brain tumours in adults: summary of NICE guidance. *BMJ*. 2018;362:k2924.
- [30] Chang SM, Messersmith H, Ahluwalia M, et al. Anticonvulsant Prophylaxis and Steroid Use in Adults With Metastatic Brain Tumors: ASCO and SNO Endorsement of the Congress of Neurological Surgeons Guidelines. *J Clin Oncol*. 2019;37(13):1130–1135.
- [31] Wen PY, Weller M, Lee EQ, et al. Glioblastoma in Adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) Consensus Review on Current Management and Future Directions. *Neuro Oncol*. 2020;24:noaa106.
- [32] Brandner S, Jaunmuktane Z. Neurological update: gliomas and other primary brain tumours in adults. *J Neurol*. 2018;265(3):717–727.

- [33] Cheng M, Zhang ZW, Ji XH, et al. Super-enhancers: a new frontier for glioma treatment. *Biochim Biophys Acta Rev Cancer*. 2020;1873(2):188353.
- [34] Lin AJ, Campian JL, Hui C, et al. Impact of concurrent versus adjuvant chemotherapy on the severity and duration of lymphopenia in glioma patients treated with radiation therapy. *J Neurooncol*. 2018;136(2):403–411.
- [35] Patil CG, Walker DG, Miller DM, et al. Phase 1 Safety, Pharmacokinetics, and Fluorescence Imaging Study of Tozuleristide (BLZ-100) in Adults With Newly Diagnosed or Recurrent Gliomas. *Neurosurgery*. 2019;85(4):641–649.
- [36] Shu HK, Julin CM, Furman F, et al. Overexpression of E2F1 in glioma-derived cell lines induces a p53-independent apoptosis that is further enhanced by ionizing radiation. *Neuro Oncol*. 2000;2(1):16–21.
- [37] Donaires FS, Godoy PR, Leandro GS, et al. E2F transcription factors associated with up-regulated genes in glioblastoma. *Cancer Biomark*. 2017;18(2):199–208.
- [38] Shen ZG, Liu XZ, Chen CX, et al. Knockdown of E2F3 Inhibits Proliferation, Migration, and Invasion and Increases Apoptosis in Glioma Cells. *Oncol Res*. 2017;25(9):1555–1566.
- [39] Yu H, Zhang D, Li Z, et al. E2F transcription factor 8 promotes proliferation and radioresistance in glioblastoma. *Pathol Res Pract*. 2020;216(8):153030.
- [40] Yin W, Wang B, Ding M, et al. Elevated E2F7 expression predicts poor prognosis in human patients with gliomas. *J Clin Neurosci*. 2016;33:187–193.
- [41] Thurlings I, Martinez-Lopez LM, Westendorp B, et al. Synergistic functions of E2F7 and E2F8 are critical to suppress stress-induced skin cancer. *Oncogene*. 2017;36(6):829–839.
- [42] Chong JL, Tsai SY, Sharma N, et al. E2f3a and E2f3b contribute to the control of cell proliferation and mouse development. *Mol Cell Biol*. 2009;29(2):414–424.
- [43] Danielian PS, Friesenhahn LB, Faust AM, et al. E2f3a and E2f3b make overlapping but different contributions to total E2f3 activity. *Oncogene*. 2008;27(51):6561–6570.
- [44] Kent LN, Bae S, Tsai SY, et al. Dosage-dependent copy number gains in E2f1 and E2f3 drive hepatocellular carcinoma. *J Clin Invest*. 2017;127(3):830–842.
- [45] Gomez-Manzano C, Mitlianga P, Fueyo J, et al. Transfer of E2F-1 to human glioma cells results in transcriptional up-regulation of Bcl-2. *Cancer Res*. 2001;61(18):6693–6697.
- [46] Zhang L, Liu Z, Dong Y, et al. E2F2 drives glioma progression via PI3K/AKT in a PFKFB4-dependent manner. *Life Sci*. 2021;276:119412.
- [47] Clijsters L, Hoencamp C, Calis JJA, et al. Cyclin F Controls Cell-Cycle Transcriptional Outputs by Directing the Degradation of the Three Activator E2Fs. *Mol Cell*. 2019;74(6):264–277.
- [48] Yuan R, Vos HR, Van Es RM, et al. Chk1 and 14-3-3 proteins inhibit atypical E2Fs to prevent a permanent cell cycle arrest. *EMBO J*. 2018;37(5):e97877.
- [49] Polager S, Kalma Y, Berkovich E, et al. E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis. *Oncogene*. 2002;21(3):437–446.
- [50] Gujar AD, Le S, Mao DD, et al. An NAD⁺-dependent transcriptional program governs self-renewal and radiation resistance in glioblastoma. *Proc Natl Acad Sci USA*. 2016;113(51):8247–8256.
- [51] Shibata E, Kiran M, Shibata Y, et al. Two subunits of human ORC are dispensable for DNA replication and proliferation. *Elife*. 2016;5:e19084.
- [52] Muller H, Bracken AP, Vernell R, et al. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev*. 2001;15(3):267–285.
- [53] Carvajal LA, Hamard PJ, Tonnessen C, et al. E2F7, a novel target, is up-regulated by p53 and mediates DNA damage-dependent transcriptional repression. *Genes Dev*. 2012;26(14):1533–1545.