A novel six-methylated pseudogenes signature predicts prognosis of glioma patients

Zhigang Chen
Department of Neurosurgery, The Second Affiliated Hospital of Anhui Medical University
Jialin Zhou
Department of Clinical Medicine, The Second School of Clinical Medical , Anhui Medical University
Bingran Wang
Department of Clinical Medicine, The Second School of Clinical Medical , Anhui Medical University
Jiahui Li
Department of Clinical Medicine, The Second School of Clinical Medical , Anhui Medical University
Han Xie
Department of Neurosurgery, The Second Affiliated Hospital of Anhui Medical University
JiaJia Zhao
Department of Neurosurgery, The Second Affiliated Hospital of Anhui Medical University
Jun Liu
Department of Orthopaedics, The Second Affiliated Hospital of Anhui Medical University
Dasheng Tian
Department of Orthopaedics, The Second Affiliated Hospital of Anhui Medical University
Erbao Bian (bianerbao@ahmu.edu.cn)
Department of Neurosurgery and Orthopaedics, The Second Affiliated Hospital of Anhui Medical University

Research Article

Keywords: Methylated pseudogene, Prognosis, Glioma, Risk signature, Overall survival

Posted Date: April 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1536640/v1

Abstract

Background: Gliomas are the most common malignant tumor from the central nervous system(CNS). Dysregulated pseudogene expression was significantly associated with the prognosis of glioma patients. However, the role of abnormal methylation of pseudogenes in glioma prognosis has not yet been studied. This study aimed to develop a novel six-methylated pseudogenes signature to predict prognosis of glioma patients.

Methods: Based on multiple screening, a risk signature for six-methylated pseudogenes was constructed, and then classified glioma patients into high-risk and low-risk groups. Next, a prognostic nomogram including grade, age, gender, and radiation was constructed. Besides, the immune cell infiltration analyses of patients based on the six-methylated pseudogenes in two risk groups were performed. Meanwhile, consensus cluster analysis of six methylated pseudogenes identified two glioma patient subgroups (cluster1/2). Furthermore, Gene Ontology (GO) , Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene set enrichment analysis (GSEA) were used to analyzed related genes. Finally, the ability of glioma to proliferate, migrate and invade were used to verify subsequent functions.

Results: In this study, six gene models consisting of methylated pseudogenes were identified and validated, and showed strong prognostic power in the training dataset, validation dataset, and entire dataset. The calibration diagram showed good predictive performance. In addition, the proportion of B cells and CD4+T cells was significantly higher in the high-risk group, while the proportion of Mono cells was lower. By silencing the expression of SBF1P1 and SUMO1P1, the ability of glioma proliferation, migration and invasion can be inhibited.

Conclusion: The six-methylated pseudogenes signature may be a novel predictor for prognostic assessment of glioma patients, which could accurately predict patient prognosis.

Background

Gliomas are most frequently occurring primary malignancies of the CNS [1]. In 2016, the classification of central nervous system tumors was defined by World Health Organization (WHO) for the first time which combined molecular biomarkers with typical histological characteristics to differentiate the different glioma subtypes [2]. Glioma has the characteristics of "three high, one low", high recurrence rate, high morbidity, high mortality and low curative ratio [3]. Hitherto, great progress has been made in etiological exploration and clinical treatment [4]. However, the trend of wide spread of glioma cells in normal brain parenchyma has seriously limited the therapeutic effect, and the efficacy of chemotherapy and biological modulators has not been confirmed. Although there has been dendritic cell (DC) therapy and checkpoint blockade, but the effectiveness of these new therapies is still not ideal in glioma [5]. In addition, systemic and local immunosuppression, which plays an important role in glioma progression, has not been solved thoroughly. Immune cell dysfunction may be associated with glioma prognosis [6]. The prognosis of gliomas of the same grade varies significantly due to the heterogeneity of gliomas [7,8]. Consequently, it

is to some extent impossible to predict the survival of patients accurate with the current prognostic methods. Therefore, the new method to predict the prognosis should be investigated.

Pseudogenes are long noncoding RNAs derived from protein encoding genes. They are defined as genomic sites that are similar with their coding homologous genes but lack the ability to translate into proteins [9]. Pseudogenes are abundant in the human genome, among which there are about 11,000 pseudogenes, occupying more than half the number of protein-coding genes [10]. The previous studies showed that pseudogenes are genomic loci similar with real genes, yet they are biologically inconsequential [11]. Therefore, pseudogenes are once regarded as “junk gene” as a result of they are not transcribed or can’t be translated into functional proteins[12] . However, several pseudogenes retain or regain their protein-coding capacity, and the resulting proteins/peptides reflects or interfere the function of their homologous proteins/polypeptides in tumorigenesis[12]. Nevertheless, with the development of next-generation sequencing, multiple functions of pseudogene have been found in a variety of cancers. Liu et al. had shown that pseudogenes may be involved in the development of tumor through crosstalk with parent genes associated with tumor [13]. Recently, pseudogenes have also been applied for identifying cancer subtypes. A recent study found that many pseudogenes function differently in different tumor subtypes [14]. Moreover, Gao et al.demonstrated the associations between the high expressions of pseudogenes and shortened overall survival in glioma[15]. In addition, overexpression of the pseudogene PTENP1 inhibited the proliferation, migration and invasion of glioma cells [16]. However, methylation of pseudogenes promoter in glioma prognosis is still largely unknown.

DNA methylation, a well-established and the most commonly studied epigenetic phenomenon in humans[17], can alter the chromosomal stability and gene expression of organisms. It is one of the major DNA modifications that regulate gene transcription and genome function [18]. DNA methylation can abate the liveness of certain genes, whereas demethylation leads to gene re-activation and expression [19]. In addition, in some malignant tumors, abnormal DNA methylation was considered to be one of the most important causes of tumorigenesis, and different patterns of DNA methylation subsets have been found in gliomas. One such subgroup is CpG island methylation phenotype of glioma (GCIMP), which is characterized by high levels of DNA methylation [20]. In gliomas, GCIMP is associated with IDH mutations and higher overall survival, which is more common in low-grade gliomas. Additionally, IDH1 mutations are likely to be early events leading to tumor development. This is an independent good prognostic marker in human gliomas [21]. Thus, DNA methylation may provide a vital foundation for clinical treatment and outcome of glioma [22]. A few total methylation status of single genes associated with their expression levels and glioma prognosis have already been recognized. In addition, DNA methylation gene biomarkers that can reveal the peculiarity of glioma and reflect the prognosis of glioma patients that have been extensively studied [23-25]. Besides, the lncRNA methylation model can better reflect the prognosis of glioma patients, which has been previously demonstrated [26]. However, the pseudogene methylation model has not been reported in glioma. Here we assume that the prognostic value would be reliable to a certain extent by integrating multiple methylated pseudogenes biomarkers.

In the present study, a risk signature was established containing the methylation level of the selected six pseudogenes which were prominent relevant to the overall survival (OS) in patients with glioma. Furthermore, our study revealed that the potentially biology functions of six methylated pseudogene signature in glioma patients and their relationship to tumor infiltrating immune cells. Finally. we found that knockdown of SBF1P1 and SUMO1P1 can restrain the ability of glioma cells to proliferate, migrate and invade. Together, our study revealed that six-methylated pseudogene signatures may be a good prognostic model and provide an effective therapeutic target for glioma patients.

Methods

Pre-processing and preliminture ary screening of pseudogene methylation site data in glioma

The clinicopathological feature of patients with glioma was downloaded from TCGA, while the methylation data was obtained from 450K methylation (https://xena.ucsc.edu) . To generate the methylation of pseudogenes-related prognostic signature, we enrolled TCGA public databases with the clinical data and methylation level of patients. The β-value (0 to 1) represented the methylation level of each probe. Finally, 649 glioma samples were classified into a training and a validation dataset randomly, in which there are 325 samples in training dataset while the validation dataset contained 324 samples.

Identification of methylated pseudogenes and construction of signature of six-methylated pseudogenes

The differentially expressed heatmaps were drawn to systematically compare the levels of pseudogene methylation in LGG and GBM. Under the filtering conditions of P <0.05, 22 prognostic pseudogenes were obtained. Subsequently, the univariate Cox regression analysis was put into to filtrate methylated pseudogenes, with the cut off P-value setting as 0.05. Next, the multivariate stepwise Cox regression analysis was carried out to build the risk prognostic model. Ultimately, six-methylated pseudogenes were used as candidates to construct the prognosis predictive model. The risk score was figured up using the following formula of the model. Risk score = coef (AZGP1P1) × promoter methylation level of AZGP1P1 + coef (SUMO1P1) × promoter methylation level of SUMO1P1 + coef (INGX) × promoter methylation level of INGX + coef (KRT19P2) × promoter methylation level of KRT19P2 + coef (SBF1P1) × promoter methylation level of SBF1P1 + coef (CES1P1) × promoter methylation level of CES1P1. Glioma patients were ranked according to risk scores and divided into high and low risk groups using the median risk score of the training dataset as a tipping point.

Evaluation of the Prognostic Model

Kaplan-Meier survival curve was employed to compare the overall survival（OS）of high and low risk groups. To determine whether this model can be an independent prognostic factor for patients with glioma, univariate and multivariate Cox regression analyses were performed for these prognostic factors. Then ROC curves were employed to estimate the predictive power of the prognostic model.

Consensus clustering based on pseudogene methylation status to analyze prognostic molecular subtypes

Consensus clustering was performed by the ConsensusClusterPlus package in R to obtain certain glioma subgroups. The samples were clustered and divided into different subtypes based on 6 methylated pseudogenes signature, so as to further explore the impact of different subtypes on the prognosis of glioma patients. The resampling method was performed to extract the data datasets of a certain sample. Different cluster numbers are specified and the rationality of k was calculated. The number of subclasses were determined based on the following standards: First, cluster in a relatively high consistency and no significant rise in the area under the cumulative distribution function (CDF) curve. Then, the heatmap of the sample consensus matrix made by rearranging the final clustering results was distinct. The heatmap was easy to identify the quality of the sample clustering results and the internal structure of the sample similarity matrix. Besides, we performed survival analysis to obtain the survival circumstance. To examine patient's survival status among different glioma subgroups, K-M curve was performed to compare the difference in overall survival between high and low risk groups. Besides, to generate a heatmap containing the pseudogene methylation level, sample type, radiation, gender, age, grade and fustat, as is to show the difference of pseudogenes methylation levels.

Analysis by Bioinformatics Methods

Nomogram was constructed using R software (version 4.1.0) package "rms" to evaluate the 1-year, 3-year, and 5-year patient's overall survival. Then, calibration plots were also used to graphically evaluate the discriminative ability of the nomogram. Finally, the prognostic nomogram was externally demonstrated in the validation dataset and entire dataset. Differentially methylated sites were investigated by Gene Ontology (GO), Kyoto Encyclopedia of Genes (KEGG) analyses and GSEA, which we employed to assess the cell functions related to the risk factors based on the prognostic characteristics of six-methylated pseudogenes.

Tumor Immune cell infiltration

In order to confirm whether and how pseudogene methylation influenced the tumor-immune microenvironment, related plots and a violin plot were drawn to show the contribution of immune cell infiltration in two datasets [27]. EpiDISH package was used to deduce the proportions of prior known certain cells in a sample containing a mixture of such cell types. In this study, it was performed to infer the proportions of 6 immune cells (Nature Killer(NK) cells, B cells, Monocytes, CD4+ T cells, CD8+ T cells and Neutrophils) in samples based on the different pseudogene methylation levels.

Cells Culture

H4, SF126, LN18, SNB19, T98G, U251 and SW1088 glioma cells, and one normal human astrocyte cell line (HEB) were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) and the cell bank of Sun Yat-Sen University (Guangzhou, China), respectively. All cells were added with

Dulbecco's modified Eagle medium (DMEM, 319-00518, Gibco, USA) containing 10% fetal bovine serum (FBS, 10,270-106, Gibco; USA) and placed in an incubator at temperature of 37.0℃, humidity saturation and CO2 concentration of 5%. When the adhesion rate reaches 80% or more, the cells are digested by trypsin (C0201, Beyotime, China) and passed from generation to generation.

Silencing and Transfection of pseudogenes

SiRNAs were used to silence SBF1P1 and SUMO1P1. SBF1P1 siRNA (U0804, RiboBio, Guangzhou, China) included three sequences (siSBF1P1-1 CCTCTCAGATACAGCTTCA, siSBF1P1-2 GCAAGAACAAGAACCTGTA, and siSBF1P1-3 CACATTCCAGCTGCTGAAA), while SUMO1P1 siRNA (U0804, RiboBio, Guangzhou, China) included another three sequences (siSUMO1P1-1 CCTCTCAAGAAACTCAAGA, siSUMO1P1-2 GAGGTCATTCAACAGTTTA, and siSUMO1P1-3 TAACGACTAACTCCAAAGA). Firstly, cells at logarithmic growth stage were choosed and uniformly inoculated into 6 well plates, and transfections were prepared when cell growth compatibility reached 40 to 50% density. Secondly, 10μL siRNA/con-RNA was diluted with 120μL 1X riboFECTTM CP Buffer and 12μL riboFECTTM CP Reagent, respectively. The reagents were fully mixed and incubated at room temperature for 15 minutes. Add riboFECTTM CP mixture to 6 well plates. Finally, 24 h after transfection cell culture, the original medium in the 6-well plate was discarded and trypsin was washed twice with PBS for use.

RT-qPCR

RT-qPCR primers applied are shown below: GAPDH, forward5'-AATGGGCAGCCGTTAGGAAA-3', reverse5'-GCGCCCAATACGACCAAATC-3'. SBF1P1, forward5'-ATTCCCCCAGCTGTTTTGCC-3', reverse5'-TTTCCTGCTCCCAGAAGGTCAAG-3'. SUMO1P1, forward5'-TGAGGCATAGCGGAAGTGAC-3', reverse5'-CAGACATGGTGATGGGGCAT-3'. Total RNA was extracted from LN18 and T98G cells by TRIzol reagent of Takara Company. Cell RNA was reversely transcribed into cDNA according to PrimeScript RT Master Mix (AK51812A, Takara, Japan) reverse transcription kit instructions. Secondly, RT-qPCR analyses were carried out with TB Green Premix Ex Taq II (Tli RNaseH Plus) (AK51812A, Takara, Japan). PCR using cDNA as template, substrate 1.6μ, primer 3.2μ, TB Green 5.2μ. The step of RT-qPCR was as follows: pre-denatured 95.0◦C, 30s, 1 cycle of 95.0◦C for 10 s, 60.0◦C for 31 s, and 40 cycles. The solubility curve was observed in 95.0°C for 15 S, 60.0◦C for 1 minute, 95.0◦C for 15 s, a cycle. The relative expression of other reference genes should be as follows: 2-ΔΔCt.

Cell Viability Assay

The MTT kit (Biyuntian, Shanghai) was used to counted cell viability(Biyuntian, Shanghai). Transfection of T98G and U251 cells was performed using si-SBF1P1 and si-SUMO1P1, respectively. After 24 hours, the cells were digested and counted, and 6000 cells were planted in each well on a 96-well plate for culture. 10 μl MTT stain was added to each well, and then the tablet was kept for 4 hours in the dark. Next, carefully discard the medium and add 140μ L DMSO to each well. After that, the 96-well plate was

gently shaken in the dark environment for 10min. Finally, the tablet was tested with the microedition reader.

Cell Migration and Invasion Assay

24 h after transfection, T98G and U251 cells were collected from 6-well plates. A transwell chamber was hydrated with 200 μl of serum-free DMEM for 1h. Next, 2×10^4 cells were inoculated in each chamber after the cells were counted using the cell counting plates carefully. Add 600 μl DMEM containing 30% fetal bovine serum to overlay the bottom of chambers. After 24 hours of culture, cells with 4% paraformaldehyde soak for 30 min, then use 0.1% crystal violet staining for 30 min. Next, Matrigel R (Gibco, USA) was added into the upper chamber, and 10×10^4 cells were inoculated in the upper chamber and cultured for 48 hour under the same culture conditions to evaluate the invasion ability of the cells. Then, the matrix R was carefully wipe on matrix R chamber. Finally, an inverted contrast phase microscope (Olympus, Japan, , magnifcation, 100×) was used to measure the number of cells which passed through the membrane at the bottom of the compartment.

Colony Formation Assay

Colony formation experiment was performed to measure glioma cell proliferation ability. Experimental group and control group cells were collected first, then inverted phase contrast microscope was used and cell count plate was used to calculate the number of cells. Add 800-900 cells to each well in 6-well plates, and then incubate the cells in a humidified incubator at 37℃ with 5% CO2 (Model no. :370, The American Thermoscientific). Culture for a week, the cell colonies suspended in the culture chamber gradually formed. 0.1% crystal violet (Shanghai, Biyuntian) was dyed for 20minture, and 4% paraformaldehyde (Shanghai, Biyuntian) was fixed for 30 minutes. White light was used to take photos and the number of cloning per well was calculated.

Statistical Analysis

All data were statistically analyzed using GraphPad Prism 9.0 (La Jolla, USA). Results were expressed by means of at least 3 times ±SD. Unpaired T test was used to compare expression differences between the two groups, or one-way analysis of variance (ANOVA) was used to assess the mean expression between the different groups. $P < 0.05$ was considered statistically significant.

Results

Selection of the different pseudogene methylation

We obtained methylation 450K and clinicopathological data from 685 glioma samples from TCGA, of which 649 had a survival value. The differential methylation of pseudogenes between Brain Lower Grade Glioma (LGG) and Glioblastoma Multiforme (GBM) were calculated (**Fig 1A**).

Prognostic value of pseudogene promoter region methylation and risk profiles established using six methylated pseudogenes

Subsequently, we used Univariate Cox regression analysis to compare methylation expression levels of different pseudogene promoters of selected samples (P□0.001). In the end, six-methylated pseudogenes (AZGP1P1, SUMO1P1, INGX, KRT19P2, SBF1P1, CES1P1) were found to be closely associated with patient prognosis **(Figure 2A,B)**. The regression coefficients of the six-methylated pseudogenes were acquired based on the multivariate Cox proportional hazard regression model, and linear combination method was used to deal with methylation expression level and corresponding coefficient of promoter region of each pseudogene selected. Finally, we obtain the following risk prediction model: risk score = (1.17854840586748 × promoter methylation level of AZGP1P1) + (-1.71361948514048 × promoter methylation level of SUMO1P1) + (-0.558409311604914 × promoter methylation level of INGX) + (-1.37625870859091 × promoter methylation level of KRT19P2) + (-1.096712728635 × promoter methylation level of SBF1P1) + (-1.84088948280091 × promoter methylation level of CES1P1). Then, based on the median risk score in the training cohort, we divided the glioma samples into two groups: the low-risk group and the high-risk group **(Figure 2C)**. We found that significant differences between high-risk and low-risk groups for WHO class **(Figure 2D)**, patient age **(Figure 2E)**, and survival **(Figure 2F)**, but not by sex **(Figure 2G)**. Similarly, the results were observed in the validation dataset **(Supplementary Figure 1).**

Construction of Survival Predicting Model using Pseudogene Promoter Region Methylation and its Prognostic Value

For 326 glioma patients from training dataset, they were classified into low- and high-risk subgroups based on the median risk score. Then, the K-M analysis was performed to compare the survival difference between high-risk and low-risk groups, and we found that for those high-risk patients, they mostly had poorer survival outcome, while better survival outcome was revealed for patients with lower risk (**Fig 3A**). There was also considerable difference between the two groups in the validation dataset (**Fig 3B**) and the entire dataset (**Fig 3C**). Subsequently, the predictive accuracy and stability were suggested by the ROC curve, with area under the receiver operating characteristic (AUC) being 0.889, indicating the model had high specificity and sensitivity (**Fig 3D**). Meanwhile, the AUC in the validation dataset (**Fig 3E**) and the entire dataset is 0.830 and 0.822, respectively (**Fig 3F**). Moreover, the methylation of six candidate pseudogenes promoter region was shown in the heatmap (**Fig 4A)**. In order of the increasing risk score, the risk score distribution and survival status of glioma patients was presented in the training dataset **(Fig 4B,C).** Subsequently, we observed similar results in the test dataset **(Fig 4D-F)** and the entire dataset **(Fig 4G-I)**. Our results suggest that the survival rate of high-risk patients was significantly lower than that of low-risk patients.

Pseudogenes Methylation Levels showed Strong Power for Prognosis Assessment

Univariate and multivariate Cox analyses were performed to estimate the independent predictive value of pseudogenes methylation for overall survival in glioma patients. The results indicated that the risk score

determined by pseudogenes methylation in promoter region was risky factor, whose hazard ratio (HR) was 1.421. WHO grade and age were also suggested to significantly affect the prognosis **(Fig 5A)**. Therefore, we integrated these risk factors into multivariate cox regression analysis. Our results demonstrated that WHO grade and risk score were significant (HR > 1, P < 0.001). The risk score was independently relevant to the prognosis **(Fig 5B)** and may serve as an independent predictive factor of glioma[28]. In the validation dataset **(Fig 5C,D)** and the entire dataset **(Fig 5E,F)**, the coincident results were presented.

Construction of a predictive nomogram and its predictive value

We employed univariate and multivariate Cox regression to evaluate whether these six methylated pseudogenes can be used as independent predictors of glioma patients. Based on multivariate analysis, a prognostic nomogram including independent prognostic factors (grade, age, gender, and radiation) was constructed. It graphically shows the 1, 3, and 5-year overall survival rates of individuals **(Fig 6A)**. "Points" indicates the score corresponding to the methylation level of a single pseudogene. "Total Points" is the sum of "points" obtained by six-methylated pseudogenes, indicating the survival probability of each glioma patient. The risk of death increased as the "total points" increased. The calibration plots **(Figure 6B)** presented that nomogram prediction is consistent with the observation in terms of the 1-, 3- and 5-year survival rates in the training dataset. These findings suggest the appreciable reliability of the nomogram. In addition, the calibration plots also demonstrated markedly different between prediction and observation for the 1-, 3- and 5- year OS probabilities of the patients in the validation dataset **(Figure 6C)** and entire dataset **(Figure 6D)**.

Consensus Clustering of Glioma Patients and Different Characteristics of pseudogene Methylation Clustering

The pseudogene methylation profiles were applied for the consensus clustering analysis to classify the primary gliomas. In order to determine the optimum number of datasets, we comprehensively consider the criteria as follows: relatively high consistency in different clusters, no significant increase in the area under the CDF curve and clear classification in the consensus matrix. Taken together, we decided to apply 2 as the optional cluster number of pseudogene methylation categories for further analysis **(Fig 7A-C)**. The 326 glioma patients from the training dataset were divided into cluster1 and cluster2. Heatmap noted by two subgroups of prognosis, sample type, radiation, gender, age, grade and fustat was drawn by heat map R software package. We analyzed the survival of each sample in the two clusters. It was also noted that patients with low pseudogene methylation were generally patients older than 60 years. We found that high grade glioma was associated with hypomethylation of pseudogene **(Fig 7D)**. Besides, Kaplan–Meier and log-rank tests were performed, and there were significant variations between these two clusters. The curve reveals that the prognosis of low pseudogene methylated samples was poorer than that of high pseudogene methylated samples **(Fig 7E)**. Additionally, we observed similar results in the validation set **(Supplementary Figure 2)**.

The Methylation Levels of pseudogenes in the promoter region influenced the Distribution of Immune Cell Infiltration

We studied the correlation between the two risk groups and immune infiltration levels, aiming to reveal the possible mechanism of the six-methylated pseudogenes signature on the prognosis of glioma. The expression of six immune cells (B cells, CD4+ T cells, CD8+ T cells, NK cells, Monocytes, and Neutrophils) were shown in the heatmap **(Fig 8A)**. Violin plot revealed the proportions of infiltrating cells, including B and CD4T cells, were significantly increased in the high-risk group, while the proportions of monocytes exerted the opposite result **(Fig 8B)**. The correlation analysis revealed the correlation between six immune cells **(Fig 8C)**. The samples revealed a significant association with an immunosuppressive tumor microenvironment in high-risk group, which coincides with the high-risk population's poor prognosis [29]. The proportions of the immune cells in the samples were presented in the histogram and monotypes took up a large proportion **(Fig 8D)**. Our results suggest that the six selected pseudogenes signature might play a vital part in the formation of immune microenvironment and distribution of the immune cell infiltration.

Potential Biological Processes Related to the High-Risk Group

To investigate potential functional characteristics associated with the six-methylated pseudogenes signature, GO analysis was used to explore the association between high-risk and low-risk groups according to the potential functional changes of six-methylated pseudogenes. Of note, we found these genes in the high-risk group were mainly enriched in extracellular matrix, organelle fission and nuclear division, which were closely relevant to the malignant phenotypes of glioma **(Fig 9A)**. Next, KEGG enrichment analysis validated that cell cycle, focal adhesions and ECM receptor interactions were strongly linked to high-risk groups according to the signature **(Fig 9B)**. Additionally, gene set enrichment analysis (GSEA) was carried on between high-risk and low-risk groups. We found that the high-risk group was strongly associated with the E2F targets, G2M checkpoint and mitotic spindle **(Fig 9C-E).** In conclusion, we believe that these genes and pathways associated with the six-methylated pseudogenes signature may be involved in the malignant phenotypes of glioma.

Inhibiting SBF1P1 or SUMO1P1 expression could significantly control migration, invasion and proliferation of glioma

Based on the literature reports on six-methylated pseudogenes, we found that SBF1P1 and SUMO1P1 were rarely investigated in gliomas and have not been functionally tested. Therefore, we selected them for the following experiment. RT-qPCR was carried out using seven glioma cell lines. The results revealed that SBF1P1 and SUMO1P1 were highly expressed in U251 and T98G cells **(Figure 10A,B)**. To validate the influence of SBF1P1 and SUMO1P1 on glioma cell proliferation, migration and invasion, we silenced SBF1P1 and SUMO1P1 with siRNA. Four sequences of SBF1P1-1, SBF1P1-2,SUMO1P1-1 and SUMO1P1-2 were selected to silence genes in T98G and U251 cells (**Figure 10C,D)**. Then, functional experiments were carried out using SBF1P1 and SUMO1P1. After transfection, MTT assay revealed that cell proliferation viability was significantly suppressed (**Figure 10E,F)**. About 2 weeks after colony formation,

we found that the colony formation viability of T98G cells was significantly suppressed, while U251 cells had trouble forming colonies **(Figure 10G-J).** Transwell and Matrigel invasion experiments showed that migration and invasion ability of glioma cells were significantly suppressed after inhibiting SBF1P1 or SUMO1P1 expression **(Figure 11A-H).** Our results suggest that Inhibiting SBF1P1 or SUMO1P1 expression can significantly decrease the ability of glioma cells to proliferate, migrate and invade.

Discussion

Glioma is the most frequent primary intracranial tumor with a poor prognosis [30]. Identification of markers predicting the survivals of gliomas is required for appropriate follow-up and treatment. Over the past few decades, varieties of molecular markers, including microRNAs [29], lncRNAs [31] and mutations of unique genes [32] were introduced in predicting survival status.

Pseudogene, though known for useless before, has been proved to be an important part of the genome in the true nuclear organism [33]. Besides, pseudogene, similar to the sequence of functional gene, is a defective copy of a functional gene, but it has undergone numerous mutations in the sequence that have lost its original function [34]. However, recently more and more studies have confirmed that pseudogenes have important biological functions. And some studies have suggested that the functions of certain pseudogenes play an important regulatory role in the development of certain diseases [35]. For example, the pseudogene CYP4Z2P was reported to correlate breast cancer [36]. Moreover, there have also been studies about the effect of pseudogenes on the prognosis prediction in cancer. The prognosis of renal cell carcinoma and breast cancer has been found to be associated with pseudogenes [37,38], which may provide a new biomarker for cancer prognosis and targeted treatment. Similar studies have also been reported in glioma. Liu et al. identified five new pseudogenes that predict the survival of LGG patients by constructing risk prognostic model, which provide new idea for the biological role of pseudogenes in glioma and identify the important role of pseudogenes in prognosis [13]. However, pseudogene methylated genes signature in glioma has not been investigated.

DNA methylation is an epigenetic event altering the activity of genes without changing their structure [39]. In addition, it can regulate genomic function and mediate carcinogenesis [40]. Abnormal DNA methylation is considered to be an important part of tumorigenesis in several malignant tumors [41]. There are already some researches about DNA methylation in tumors, of which O6-methylguanine-DNA methyltransferase (MGMT) methylation is particularly significant. For predicting the outcome of GBM (the grade IV glioma), MGMT has been widely studied as a biomarker [42-44]. And 10 genes associated with MGMT promoter methylation were used to predict the outcome of glioma patients [45]. However, this is not an experimental study. Besides, data heterogeneity and platform differences between data sets may affect the accuracy of the results [46], so comprehensive experiments should be designed and performed in the future to validate these findings.

As a result, constructing a pseudogene methylated signature so as to estimate the outcome of patients with glioma of diverse grades is essential. Firstly, we built a prognostic model utilizing six glioma-

associated differentially methylated pseudogenes, which were derived by layers of screening. Next, glioma samples were divided into low- and high-risk group based on risk predictive signature. Survival analysis revealed that high-risk group had a worse outcome compared to the low-risk group. More importantly, the hypomethylation of those pseudogenes are basically tumor-promoting, which contributes to the development of high-grade gliomas. Futhermore, ROC curve is also drawn to assess the accuracy of the signature. Moreover, univariate and multivariate Cox analyses as well as nomogram were performed to verify the risk signature. In the end, the results showed that the predictive ability of the six-methylated pseudogenes prognosis prediction risk model was similar to that of the Nomogram integrating risk scoring models and clinical features, which demonstrates that the signature has a superior performance for prediction and has potential clinical application[47]. Because there might be a false positive in the six-methylated pseudogenes signature, we further checked the outcome value of the model in the validation dataset and entire dataset[48]. The results revealed that this model has stable prediction ability and accuracy.

Different immune infiltration degree can partly explain why patients with the same histology type of cancer may have distinct clinical outcome [49]. Therefore, this study inferred the proportions of six immune cell subsets from glioma samples using EpiDISH analysis. Moreover, there is a high correlation between infiltrating immune cells and tumor cell heterogeneity in different parts of immunotherap [50]. Therefore, this research inferred the proportions of six immune cell subsets from glioma samples using EpiDISH analysis. The results indicated the high-risk group had increased proportions of B cells and $CD4^{+}$T cells, whereas monocytes were upregulated in the low-risk group. Previous studies have shown that high B cells and monocyte infiltration is associated with poor prognosis and high tumor grade in glioma patients [51]. Besides, another study had shown that the high ratio of CD4 + / CD8 + tumor-infiltrating lymphocytes is related to poor prognosis, high tumor grade and malignancy of GBM [52]. Considered together, our findings are consistent with findings that cell infiltration in the tumor immune microenvironment is associated with poor survival in glioma patients.

Our GO analysis revealed that extracellular matrix, organelle fission and nuclear division in cancer significantly affected glioma occurrence and development. Some studies have also shown that the ECM interacts with cells to regulate diverse functions, including proliferation, migration and differentiation [53]. In addition, ECM has also been found to correlate with prognosis and therapeutic efficacy in patients with cancer [54]. Furthermore, the result of our KEGG analysis supported this again. A previous study showed that disorder of cell cycle and mitotic nuclear diversion is feature of cancer cells and leads to uncontrolled cell proliferation [55]. Besides, Cell adhesion and cell cycle play a central role in glioma migration and development [56]. Zou et al. had suggested that CDR1as mediate in alteration of the tumor microenvironment through regulating ECM-receptor interaction [57]. Similarly, GSEA results revealed that selected methylated pseudogenes are closely relevant to cancer progression and involve E2F targets, G2M checkpoint and mitotic spindle [58-60]. Subsequently, we performed functional experiments on the glioma cell lines. Our study showed that the knockdown of SBF1P1 and SUMO1P1 inhibited glioma cell

proliferation, migration, and invasion in vitro. These findings showed that SBF1P1 and SUMO1P1 may play a tumorigenic role.

The current WHO classification and histopathological classification cannot further predict accurate prognosis for glioma patients [61], so the appearance of molecular markers and molecular typing technology of glioma is important. In our study, consensus clustering based on the methylation degree of six selected pseudogenes was utilized to divide samples into two glioma subtypes. The consequences confirm that our molecular classification could steadily identify the two subtypes in glioma patients and there is significant difference between them.

However, the present study has several limitations, how the pseudogene methylation influences the glioma and which factors induce methylation in the pseudogene promoter region have not been thoroughly investigated. Exploring these factors will be our research direction in the next study period. Although the six-methylated pseudogenes risk prognostic signature was validated in clinical glioma patients in the validation dataset and entire dataset, the acquired data lacked some pathological features closely related to glioma, such as molecular typing, characteristic mutation and chemoradiotherapy of the glioma. And our present study focused merely on microarray expression datasets. Additionally, we need to conduct further studies to illustrate the underlying mechanisms behind the functions of these prognostic pseudogenes. Nevertheless, unlike previous studies of promoter region DNA methylation, our research confirms its important role in the development and progression of gliomas. It provides a candidate as prognostic biomarkers for identifying glioma prognosis.

Conclusion

In summary, we developed a novel six-methylated pseudogenes signature to predict prognosis of glioma patients. The signature was an independent prognostic factor, and generated good risk stratification capabilities. More importantly, It can be used as a prognostic classifier for clinical patient individualization and follow-up.

Abbreviations

CNS, central nervous system; TCGA, The Cancer Genome Atlas; OS, overall survival; GBM, glioblastoma multiforme; LGG, lower grade glioma; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; WHO, world health organization; K-M, Kaplan Meier; CDF, cumulative distribution function.

Declarations

Ethics approval and consent to participate

This study was carried out following relevant guidelines and regulations (Declaration of Helsinki). Ethical approval of the manuscript was obtained from The Second Affiliated Hospital of Anhui Medical

University Research Ethics Committee (approval no). All data are freely available and this study did not involve any human or animal experiments.

Consent for publication

Not applicable.

Availability of data and material

The datasets analyzed for this study can be found in TCGA database (https://xena.ucsc.edu). The literature survey data from this study can be request from the corresponding author Erbao Bian.

Competing interests

All authors state that there are no potential competing interests.

Funding

This research was funded by the National Natural Science Foundation of China (No. 81972348), College Excellent Youth Talent Support Program in Anhui Province (No.gxypZD2019019), Key Projects of Natural Science Research in Anhui Province (KJ2019A0267), Academic Funding Project for Top Talents in Colleges and Universities in Anhui Province (No.gxbjZD10).

Author contributions

EB and DS conceived and designed the experiments. ZG performed the experiments. JL and BR analyzed the data. JH and JL drafted the manuscript. HX and JJ drafted the article. All authors contributed to the article and approved the submitted version.

Acknowledgements

The author thanks the TCGA network for their contributions.

References

1. Lin J, Ding S, Xie C, Yi R, Wu Z, Luo J, et al. MicroRNA-4476 promotes glioma progression through a miR-4476/APC/beta-catenin/c-Jun positive feedback loop. Cell Death Dis 2020;11(4):269. https://doi.org/ 10.1038/s41419-020-2474-4.
2. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. Acta Neuropathol 2016;131(6):803-820. https://doi.org/ 10.1007/s00401-016-1545-1.
3. Sun Y, Jing Y, Zhang Y. Serum lncRNA-ANRIL and SOX9 expression levels in glioma patients and their relationship with poor prognosis. World J Surg Oncol 2021;19(1):287. https://doi.org/ 10.1186/s12957-021-02392-2.

4. Feng L, Rao M, Zhou Y, Zhang Y, Zhu Y. Long noncoding RNA 00460 (LINC00460) promotes glioma progression by negatively regulating miR-320a. J Cell Biochem 2019;120(6):9556-9563. https://doi.org/ 10.1002/jcb.28232.
5. Long S, Li M, Liu J, Yang Y, Li G. Identification of immunologic subtype and prognosis of GBM based on TNFSF14 and immune checkpoint gene expression profiling. Aging (Albany NY) 2020;12(8):7112-7128. https://doi.org/ 10.18632/aging.103065.
6. Qi Y, Chen D, Lu Q, Yao Y, Ji C. Bioinformatic profiling identifies a fatty acid Metabolism-Related gene risk signature for malignancy, prognosis, and immune phenotype of glioma. Dis Markers 2019;2019:3917040. https://doi.org/ 10.1155/2019/3917040.
7. Chien Y, Chen J, Chen Y, Chou R, Lee H, Yu Y. Epigenetic silencing of miR-9 promotes migration and invasion by EZH2 in glioblastoma cells. Cancers 2020;12(7):1781. https://doi.org/ 10.3390/cancers12071781.
8. Alessandra Santangelo MRGL, Gian Luca De Salvo SIMC, Gambari CSGD, Ariela Brandes TISR, Vittorina Zagonel AGC. A Molecular Signature associated with prolonged survival in Glioblastoma patients treated with Regorafenib. Neuro-Oncology 2020.
9. Brent MR. Genome annotation past, present, and future: How to define an ORF at each locus. Genome Res 2005;15(12):1777-86. https://doi.org/ 10.1101/gr.3866105.
10. Consortium TEP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489(7414):57-74. https://doi.org/ 10.1038/nature11247.
11. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010;465(7301):1033-1038. https://doi.org/ 10.1038/nature09144.
12. Lu X, Gao A, Ji L, Xu J. Pseudogene in cancer: Real functions and promising signature. J Med Genet 2015;52(1):17-24. https://doi.org/ 10.1136/jmedgenet-2014-102785.
13. Liu B, Liu J, Liu K, Huang H, Li Y, Hu X, et al. A prognostic signature of five pseudogenes for predicting lower-grade gliomas. Biomed Pharmacother 2019;117:109116. https://doi.org/ 10.1016/j.biopha.2019.109116.
14. Han L, Yuan Y, Zheng S, Yang Y, Li J, Edgerton ME, et al. The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. Nat Commun 2014;5(1). https://doi.org/ 10.1038/ncomms4963.
15. Gao K, Chen X, Zhang J, Wang Y, Yan W, You Y. A pseudogene-signature in glioma predicts survival. J Exp Clin Canc Res 2015;34(1). https://doi.org/ 10.1186/s13046-015-0137-6.
16. Hu S, Xu L, Li L, Luo D, Zhao H, Li D, et al. Overexpression of lncRNA PTENP1 suppresses glioma cell proliferation and metastasis in vitro. Oncotargets Ther 2019;Volume 12:147-156. https://doi.org/ 10.2147/OTT.S182537.
17. Zafon C GJPB. DNA methylation in thyroid cancer. 2019.
18. Ghoshal K, Bai S. DNA methyltransferases as targets for cancer therapy. Drugs Today (Barc) 2007;43(6):395-422. https://doi.org/ 10.1358/dot.2007.43.6.1062666.

19. Mirfattah B, Herring J, Tang H, Zhang K. Probes and targets of DNA methylation and demethylation in drug development. Curr Top Med Chem 2017;17(15):1727-1740. https://doi.org/ 10.2174/1568026617666161116143828.
20. Dahlin AM, Wibom C, Ghasimi S, Brännström T, Andersson U, Melin B. Relation between established glioma risk variants and DNA methylation in the tumor. Plos One 2016;11(10):e0163067. https://doi.org/ 10.1371/journal.pone.0163067.
21. Waitkus MS, Diplas BH, Yan H. Biological role and therapeutic potential of IDH mutations in cancer. Cancer Cell 2018;34(2):186-195. https://doi.org/ 10.1016/j.ccell.2018.04.011.
22. Tan Y, Zhang S, Xiao Q, Wang J, Zhao K, Liu W, et al. Prognostic significance of ARL9 and its methylation in low-grade glioma. Genomics 2020;112(6):4808-4816. https://doi.org/ 10.1016/j.ygeno.2020.08.035.
23. Gusyatiner O, Hegi ME. Glioma epigenetics: From subclassification to novel treatment options. Semin Cancer Biol 2018;51:50-58. https://doi.org/ 10.1016/j.semcancer.2017.11.010.
24. Klughammer J, Kiesel B, Roetzer T, Fortelny N, Nemc A, Nenning KH, et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. Nat Med 2018;24(10):1611-1624. https://doi.org/ 10.1038/s41591-018-0156-x.
25. Wang W, Zhao Z, Wu F, Wang H, Wang J, Lan Q, et al. Bioinformatic analysis of gene expression and methylation regulation in glioblastoma. J Neurooncol 2018;136(3):495-503. https://doi.org/ 10.1007/s11060-017-2688-1.
26. Tan Y, Zhang S, Xiao Q, Wang J, Zhao K, Liu W, et al. Prognostic significance of ARL9 and its methylation in low-grade glioma. Genomics 2020;112(6):4808-4816. https://doi.org/ 10.1016/j.ygeno.2020.08.035.
27. Klughammer J, Kiesel B, Roetzer T, Fortelny N, Nemc A, Nenning K, et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. Nat Med 2018;24(10):1611-1624. https://doi.org/ 10.1038/s41591-018-0156-x.
28. Gusyatiner O, Hegi ME. Glioma epigenetics: From subclassification to novel treatment options. Semin Cancer Biol 2018;51:50-58. https://doi.org/ 10.1016/j.semcancer.2017.11.010.
29. Yu S, Hu C, Cai L, Du X, Lin F, Yu Q, et al. Seven-Gene signature based on glycolysis is closely related to the prognosis and tumor immune infiltration of patients with gastric cancer. Front Oncol 2020;10. https://doi.org/ 10.3389/fonc.2020.01778.
30. Cheng M, Sun L, Huang K, Yue X, Chen J, Zhang Z, et al. A signature of nine lncRNA methylated genes predicts survival in patients with glioma. Front Oncol 2021;11. https://doi.org/ 10.3389/fonc.2021.646409.
31. Lin W, Wu S, Chen X, Ye Y, Weng Y, Pan Y, et al. Characterization of hypoxia signature to evaluate the tumor immune microenvironment and predict prognosis in glioma groups. Front Oncol 2020;10. https://doi.org/ 10.3389/fonc.2020.00796.
32. Nørøxe DS, Yde CW, østrup O, Michaelsen SR, Schmidt AY, Kinalis S, et al. Genomic profiling of newly diagnosed glioblastoma patients and its potential for clinical utility – a prospective, translational

study. Mol Oncol 2020;14(11):2727-2743. https://doi.org/ 10.1002/1878-0261.12790.

33. De Martino M, Forzati F, Arra C, Fusco A, Esposito F. HMGA1-pseudogenes and cancer. Oncotarget 2016;7(19):28724-35. https://doi.org/ 10.18632/oncotarget.7427.
34. Li H, Jiang F, Wu P, Wang K, Cao Y. A High-Quality Genome Sequence of Model Legume Lotus japonicus (MG-20) Provides Insights into the Evolution of Root Nodule Symbiosis. Genes (Basel) 2020;11(5). https://doi.org/ 10.3390/genes11050483.
35. Zhu Q, Wang J, Zhang Q, Wang F, Fang L, Song B, et al. Methylationdriven genes PMPCAP1, SOWAHC and ZNF454 as potential prognostic biomarkers in lung squamous cell carcinoma. Mol Med Rep 2020;21(3):1285-1295. https://doi.org/ 10.3892/mmr.2020.10933.
36. Adamson C, Kanu OO, Mehta AI, Di C, Lin N, Mattox AK, et al. Glioblastoma multiforme: A review of where we have been and where we are going. Expert Opin Inv Drug 2009;18(8):1061-1083. https://doi.org/ 10.1517/13543780903052764.
37. Li R, Gao K, Luo H, Wang X, Shi Y, Dong Q, et al. Identification of intrinsic subtype-specific prognostic microRNAs in primary glioblastoma. J Exp Clin Cancer Res 2014;33:9. https://doi.org/ 10.1186/1756-9966-33-9.
38. Zhang X, Sun S, Lam K, Kiang KM, Pu JK, Ho AS, et al. A long non-coding RNA signature in glioblastoma multiforme predicts survival. Neurobiol Dis 2013;58:123-131. https://doi.org/ 10.1016/j.nbd.2013.05.011.
39. Yoo JY, You YA, Kwon EJ, Park MH, Shim S, Kim YJ. Differential expression and methylation of integrin subunit alpha 11 and thrombospondin in the amnion of preterm birth. Obstet Gynecol Sci 2018;61(5):565-574. https://doi.org/ 10.5468/ogs.2018.61.5.565.
40. Aref-Eshghi E, Schenkel LC, Ainsworth P, Lin H, Rodenhiser DI, Cutz JC, et al. Genomic DNA Methylation-Derived algorithm enables accurate detection of malignant prostate tissues. Front Oncol 2018;8:100. https://doi.org/ 10.3389/fonc.2018.00100.
41. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010;17(1):98-110. https://doi.org/ 10.1016/j.ccr.2009.12.020.
42. Rao AM, Quddusi A, Shamim MS. The significance of MGMT methylation in Glioblastoma Multiforme prognosis. J Pak Med Assoc 2018;68(7):1137-1139.
43. Butler M, Pongor L, Su YT, Xi L, Raffeld M, Quezado M, et al. MGMT status as a clinical biomarker in glioblastoma. Trends Cancer 2020;6(5):380-391. https://doi.org/ 10.1016/j.trecan.2020.02.010.
44. Zhang HW, Lyu GW, He WJ, Lei Y, Lin F, Wang MZ, et al. DSC and DCE histogram analyses of glioma biomarkers, including IDH, MGMT, and TERT, on differentiation and survival. Acad Radiol 2020;27(12):e263-e271. https://doi.org/ 10.1016/j.acra.2019.12.010.
45. Zhang Y, Zhu J. Ten genes associated with MGMT promoter methylation predict the prognosis of patients with glioma. Oncol Rep 2019;41(2):908-916. https://doi.org/ 10.3892/or.2018.6903.
46. Zhang Y, Zhu J. Ten genes associated with MGMT promoter methylation predict the prognosis of patients with glioma. Oncol Rep 2019;41(2):908-916. https://doi.org/ 10.3892/or.2018.6903.

47. Zhao X, Liu X, Cui L. Development of a five-protein signature for predicting the prognosis of head and neck squamous cell carcinoma. Aging (Albany NY) 2020;12(19):19740-19755. https://doi.org/ 10.18632/aging.104036.
48. Li X, Ren Q, Weng Y, Cai H, Zhu Y, Zhang Y. SCGPred: A score-based method for gene structure prediction by combining multiple sources of evidence. Genomics Proteomics Bioinformatics 2008;6(3-4):175-85. https://doi.org/ 10.1016/S1672-0229(09)60005-X.
49. Wang W, Zhao Z, Wu F, Wang H, Wang J, Lan Q, et al. Bioinformatic analysis of gene expression and methylation regulation in glioblastoma. J Neurooncol 2018;136(3):495-503. https://doi.org/ 10.1007/s11060-017-2688-1.
50. Lv W, Ren Y, Hou K, Hu W, Yi Y, Xiong M, et al. Epigenetic modification mechanisms involved in keloid: Current status and prospect. Clin Epigenetics 2020;12(1):183. https://doi.org/ 10.1186/s13148-020-00981-8.
51. Wang W, Zhao Z, Wu F, Wang H, Wang J, Lan Q, et al. Bioinformatic analysis of gene expression and methylation regulation in glioblastoma. J Neurooncol 2018;136(3):495-503. https://doi.org/ 10.1007/s11060-017-2688-1.
52. Liu S, Zheng Y, Zhang Y, Zhang J, Xie F, Guo S, et al. Methylation-mediated LINC00261 suppresses pancreatic cancer progression by epigenetically inhibiting c-Myc transcription. Theranostics 2020;10(23):10634-10651. https://doi.org/ 10.7150/thno.44278.
53. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Bio 2014;15(12):786-801. https://doi.org/ 10.1038/nrm3904.
54. Zhang H, Shi Q, Yang Z, Wang K, Zhang Z, Huang Z, et al. An extracellular Matrix-Based signature associated with immune microenvironment predicts the prognosis and therapeutic responses of patients with oesophageal squamous cell carcinoma. Frontiers in Molecular Biosciences 2021;8. https://doi.org/ 10.3389/fmolb.2021.598427.
55. Pau CT, Mosbruger T, Saxena R, Welt CK. Phenotype and tissue expression as a function of genetic risk in polycystic ovary syndrome. Plos One 2017;12(1):e0168870. https://doi.org/ 10.1371/journal.pone.0168870.
56. Dong R, Bai M, Zhao J, Wang D, Ning X, Sun S. A comparative study of the gut microbiota associated with immunoglobulin a nephropathy and membranous nephropathy. Front Cell Infect Microbiol 2020;10:557368. https://doi.org/ 10.3389/fcimb.2020.557368.
57. Han S, Zhang C, Li Q, Dong J, Liu Y, Huang Y, et al. Tumour-infiltrating CD4+ and CD8+ lymphocytes as predictors of clinical outcome in glioma. Brit J Cancer 2014;110(10):2560-2568. https://doi.org/ 10.1038/bjc.2014.162.
58. Matheson CJ, Backos DS, Reigan P. Targeting WEE1 kinase in cancer. Trends Pharmacol Sci 2016;37(10):872-881. https://doi.org/ 10.1016/j.tips.2016.06.006.
59. Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer. Nat Rev Cancer 2019;19(6):326-338. https://doi.org/ 10.1038/s41568-019-0143-7.

60. Lopes D, Maiato H. The tubulin code in mitosis and cancer. Cells-Basel 2020;9(11):2356. https://doi.org/ 10.3390/cells9112356.
61. Huang K, Yue X, Zheng Y, Zhang Z, Cheng M, Li L, et al. Development and validation of an Mesenchymal-Related long Non-Coding RNA prognostic model in glioma. Front Oncol 2021;11:726745. https://doi.org/ 10.3389/fonc.2021.726745.

Figures

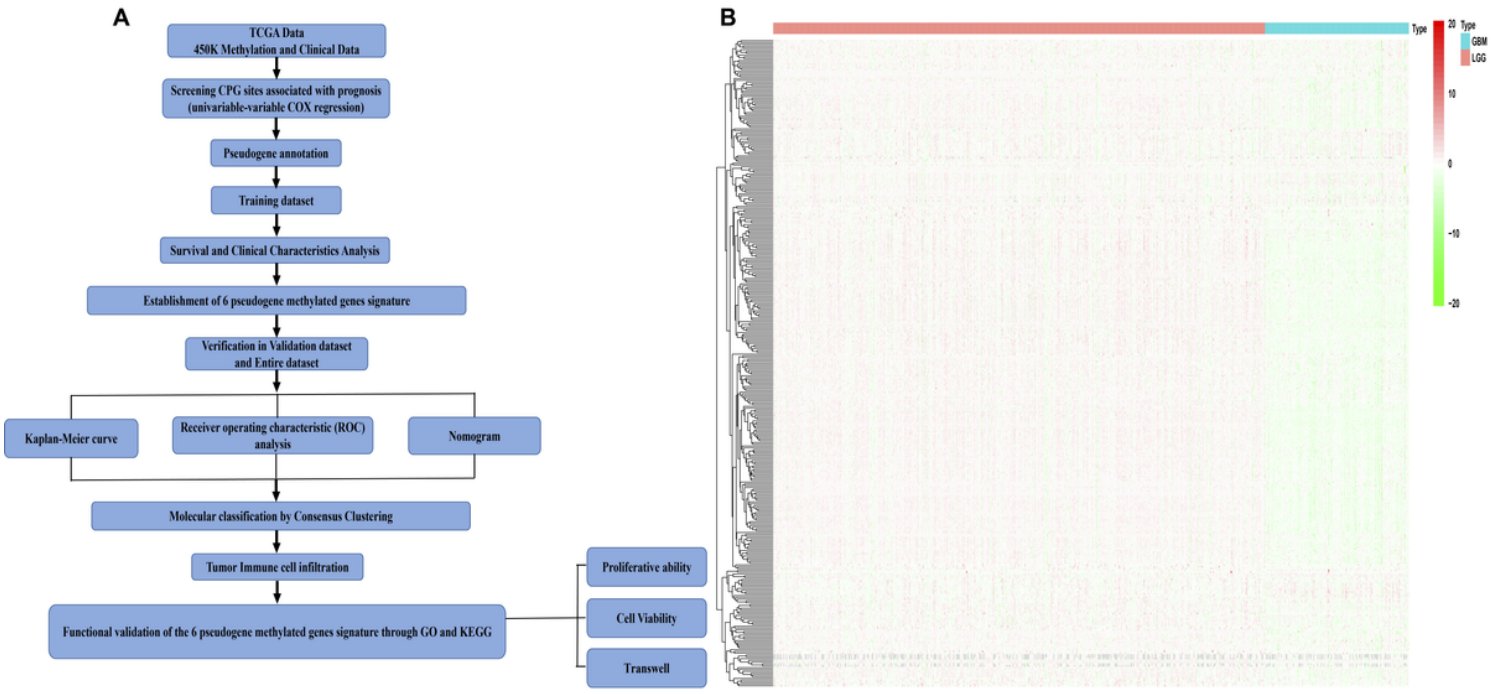

Figure 1

Identification of methylated pseudogenes. (A) Flow chart of the study. **(B)**Heat map showing the difference of pseudogene methylation in GBM and LGG.

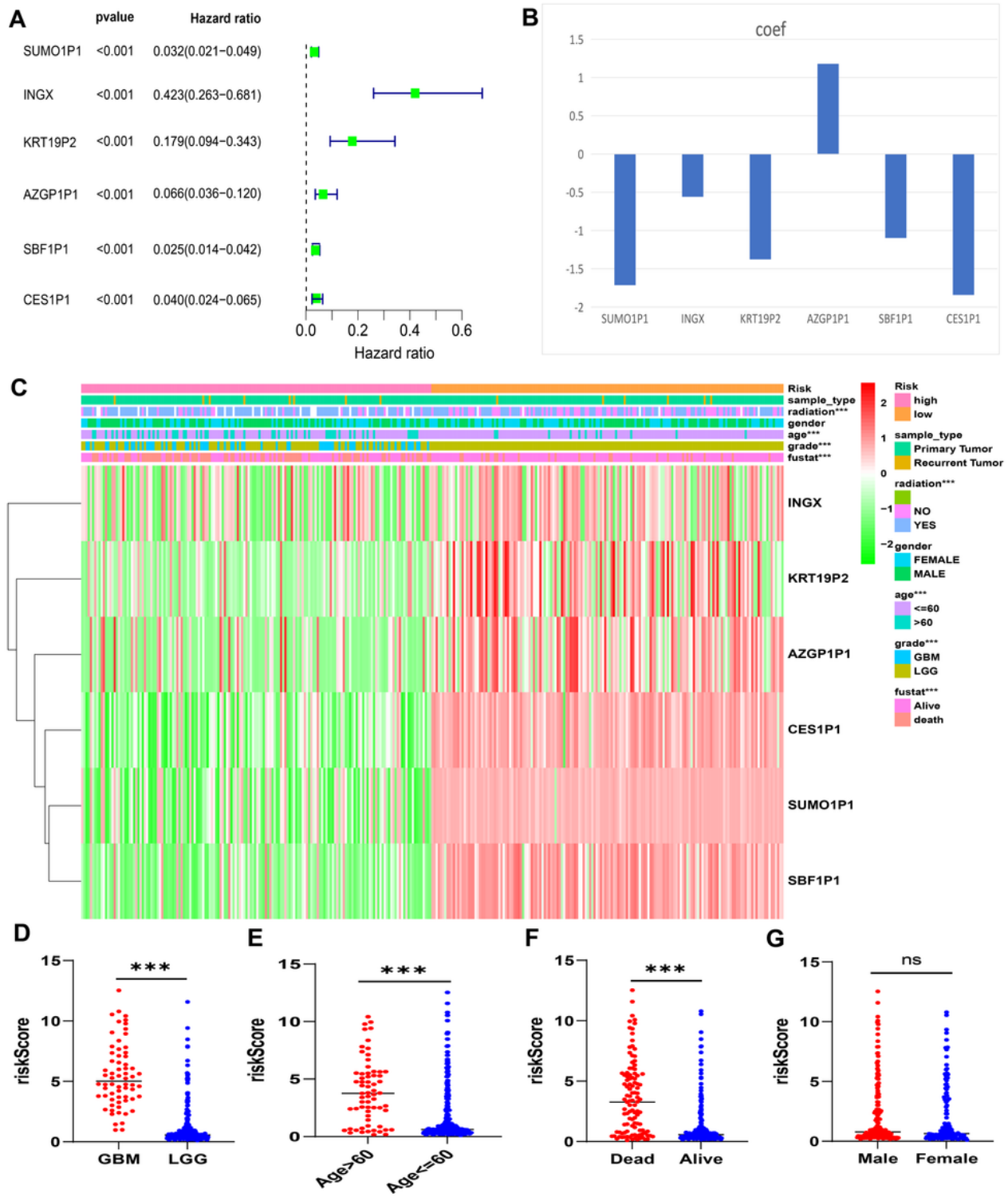

Figure 2

Construction of a risk signature based on 6 methylated pseudogenes. **(A)** P value and hazard ratio of the six-methylated pseudogenes. **(B)** Coefficient values for the six-methylated pseudogenes. **(C)** Heat map showing risk scores in a risk model in relation to clinicopathological features. **(D–G)** The box chart shows the risk assessment scores of patients with glioma at different WHO grades **(D)** , survival statuses **(E)** , ages **(F)** , and genders **(G)** . ns: no significance; ***P<0.001.

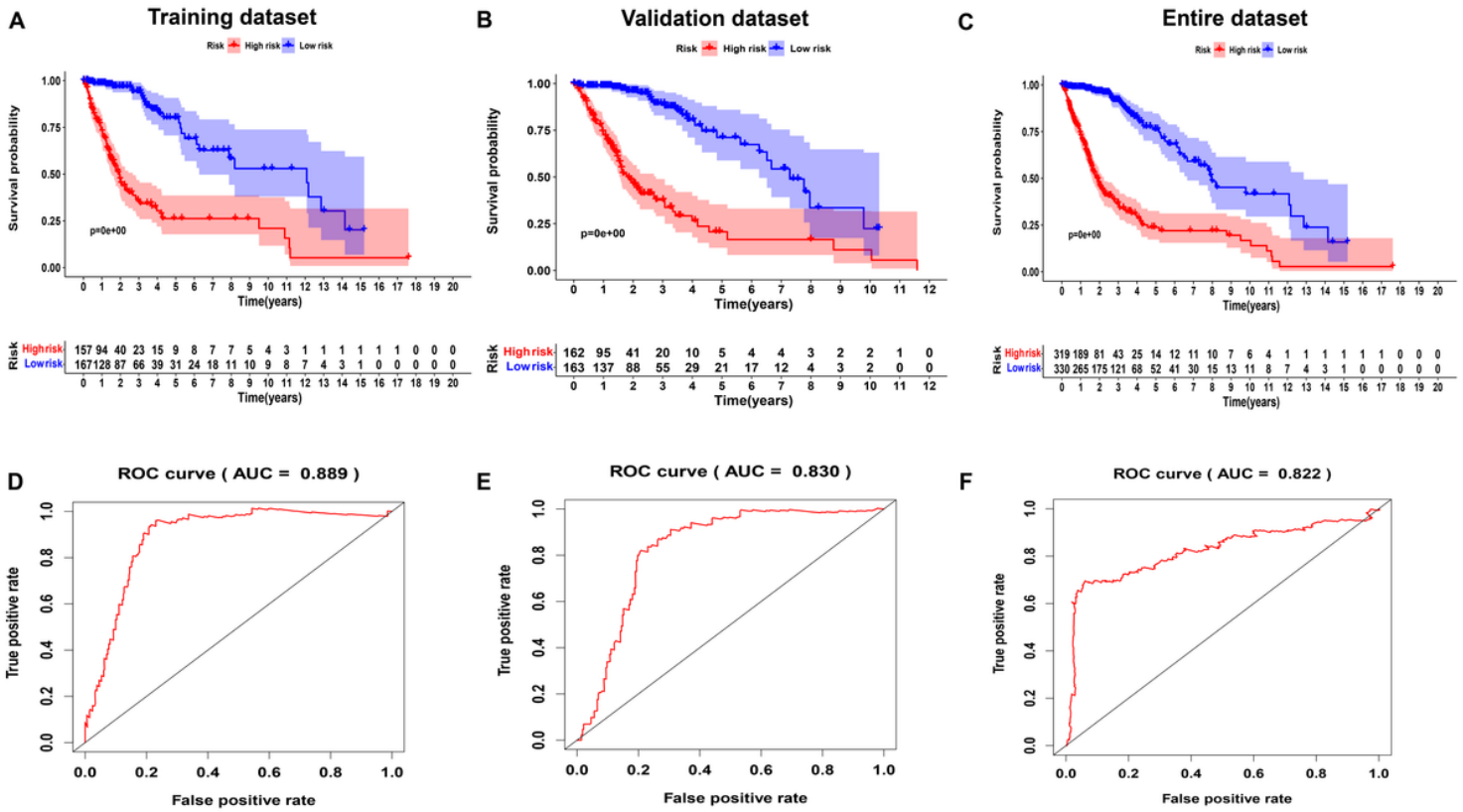

Figure 3

Kaplan-Meier survival and ROC curve analysis in the training, validation and entire dataset. (A-C) KaplanMeier curves of the overall survival of patients with high-risk scores and low-risk scores in training**(A)**, validation**(B)** and entire dataset(**C**). **(D-F)** ROC curve accessing the sensitivity and specificity of the prognostic model in training dataset**(D)**, validation dataset**(E)** and entire dataset**(F)**.

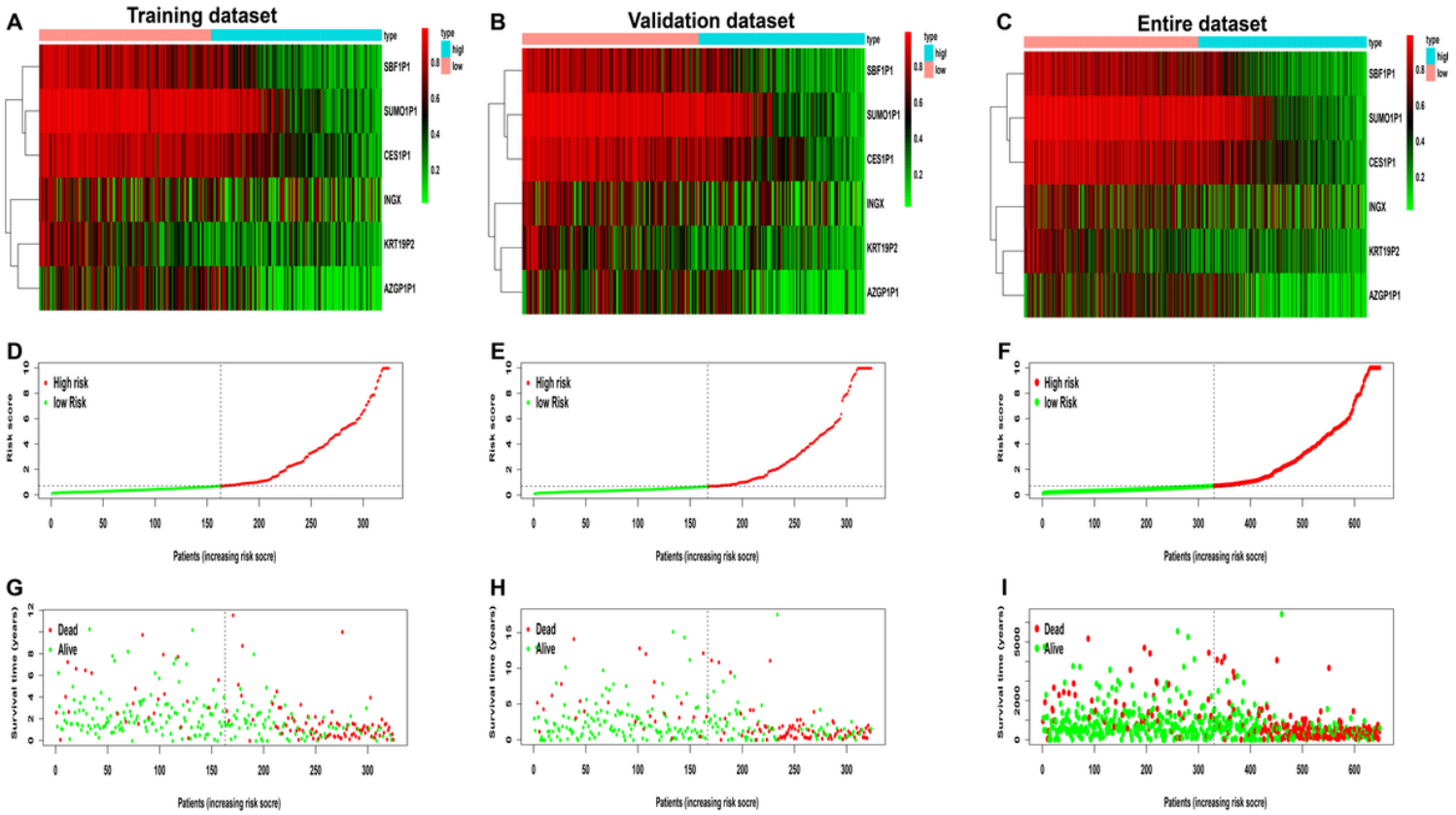

Figure 4

After we constructed the six-gene signature, we first validated their ability to effectively dissect the survival time, survival status and survival degree of patients. (A-C) Heat map showing the degree of methylation of six selected pseudogene promoter regions in training dataset **(A)** , validation dataset **(B)** and entire dataset**(C)** of glioma patients. **(D-F)** Then according to the median of the risk score in validation dataset, patients were categorized into high-risk group and low-risk group in the training dataset **(D)** , the validation dataset **(E)** and entire dataset **(F)** . **(G-I)**The scatter plot shows the survival of glioma patients in the training dataset **(G)** and the validation dataset **(H)** and entire dataset **(I)**. The red circles represent those who are dead, and the green circle represents those who are still alive.

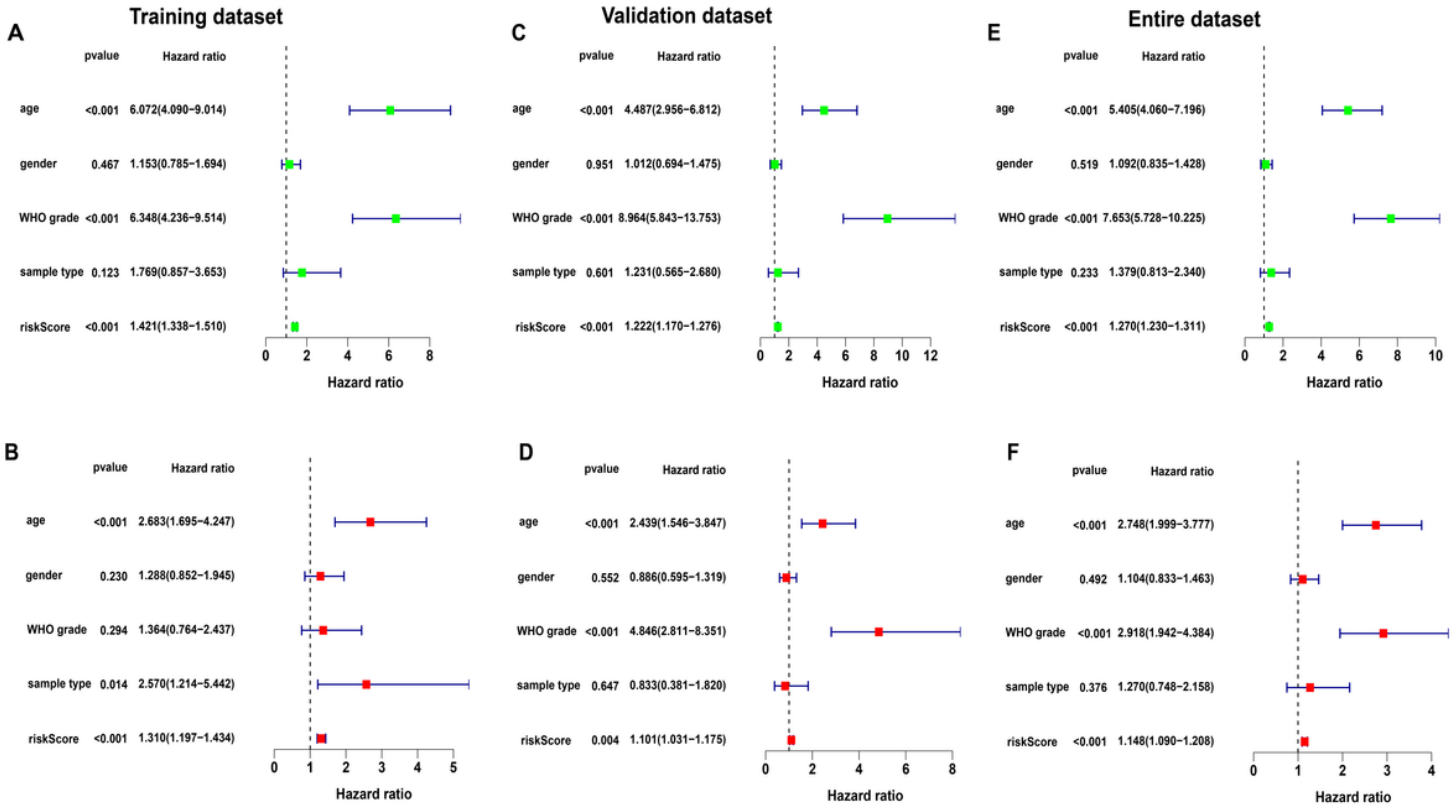

Figure 5

Forest maps were used to analyze the relevance between risk scores and other clinical features. (A-C) Univariate COX analysis of pseudogenes in training dataset **(A)**, the validation dataset **(B)** and the entire dataset **(C)**. **(D-F)**Multivariate COX analysis of pseudogenes in training dataset **(D)**, validation dataset **(E)** and entire dataset **(F)**.

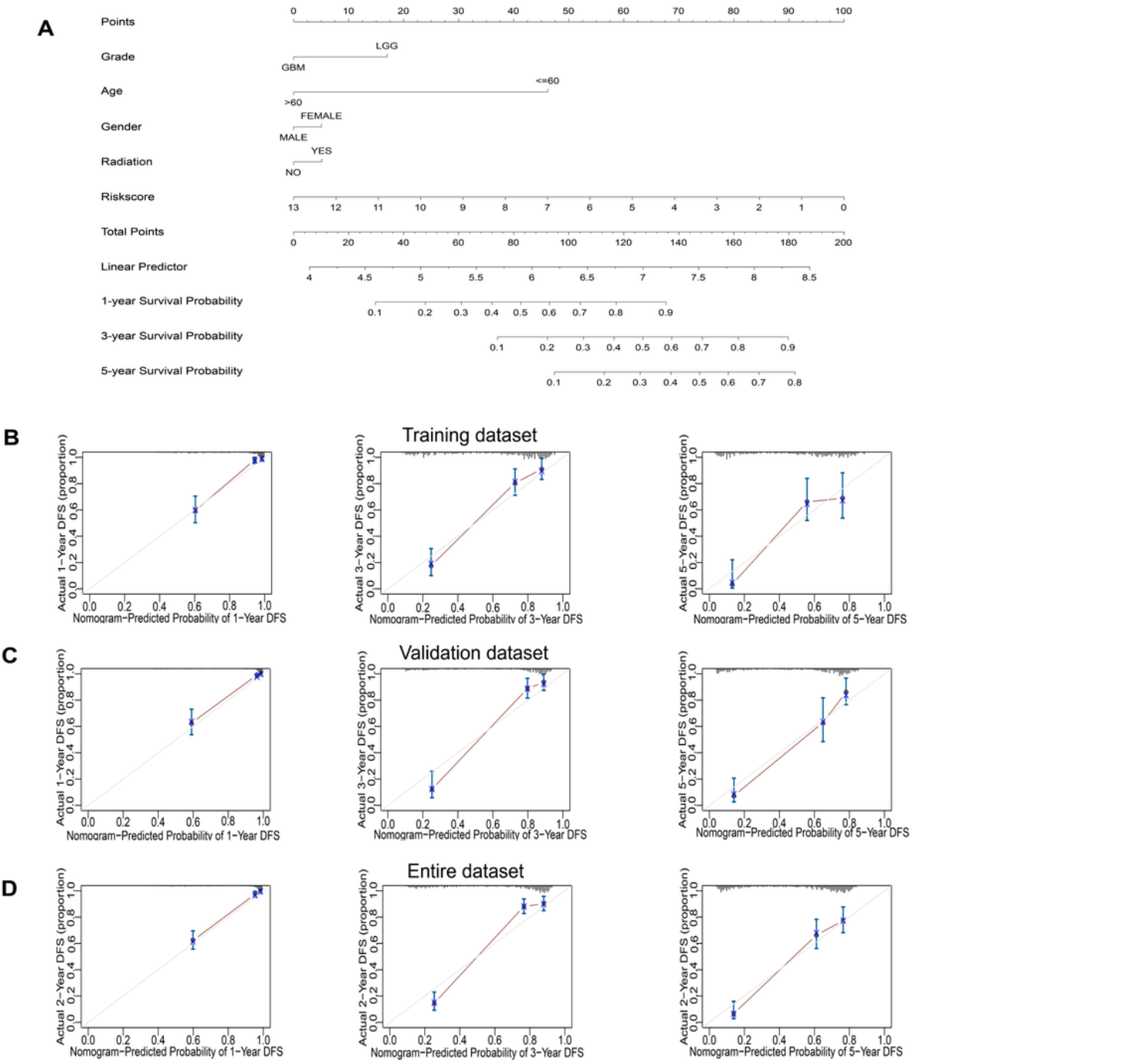

Figure 6

A validation of nomogram in predicting the overall survival rate of glioma. (A) The prognostic nomogram of 1-, 3-, and 5-year overall survival of glioma. **(B-J)** Calibration analysis of the six-pseudogenes containing nomogram for 1- , 3-, and 5-year overall survival in the training dataset (**B**), the validation dataset **(C)** and the entire dataset **(D)**.

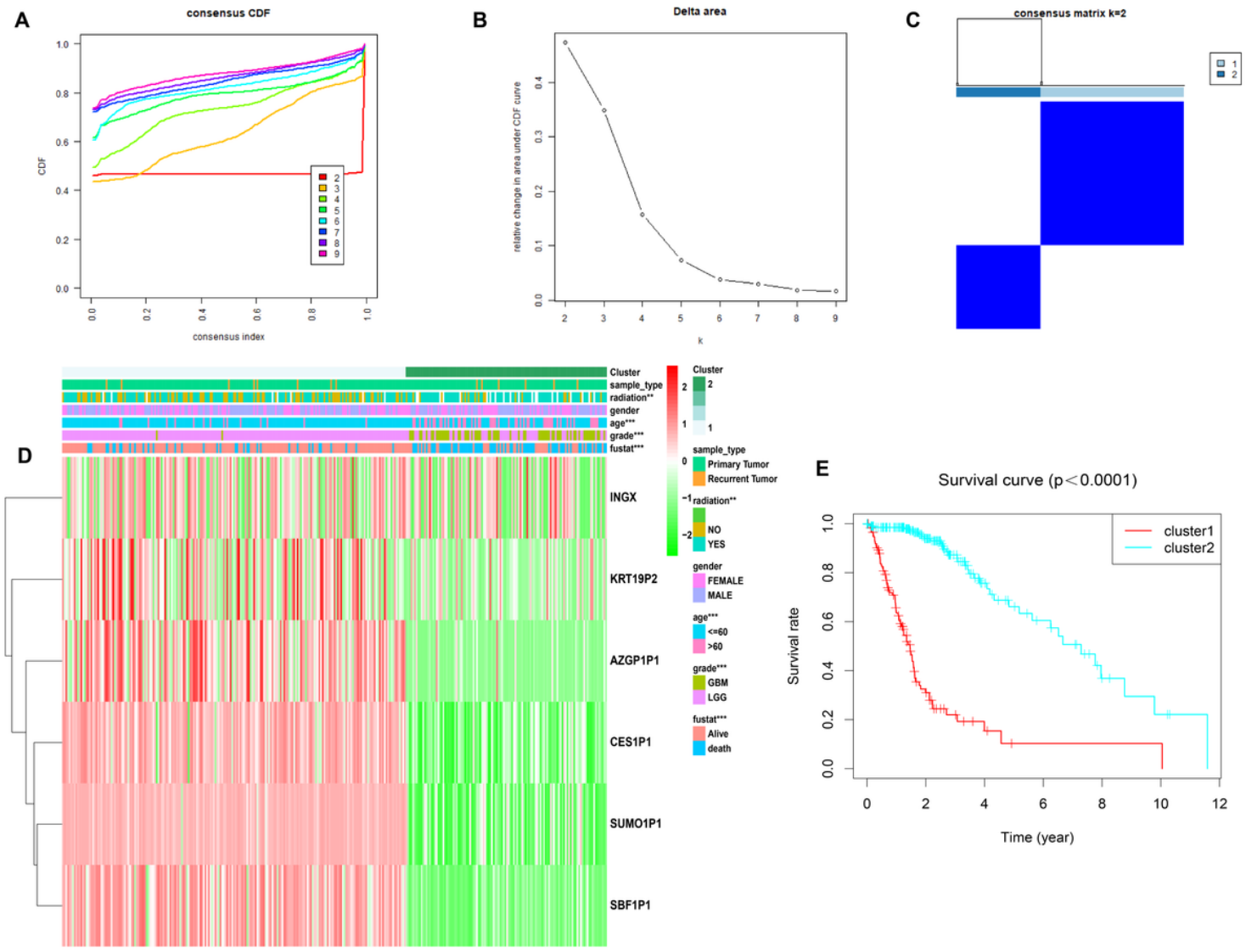

Figure 7

Glioma prognosis-related pseudogenes identified glioma subtypes for patients with different clinical and molecular features. (A) Consensus clustering cumulative distribution function curves for different subtype numbers (k = 2 to k = 9). **(B)** Relative change in area under CDF curve for k = 2 to k = 9. **(C)** The heatmap(consensus matrix) illustrating the consensus matrix at k = 2. **(D)** Heat map showing the methylation difference of the 6 candidate pseudogenes with methylation classification and different clinicopathological characteristics. **(E)** Survival analysis of patients in cluster 1 and cluster 2.

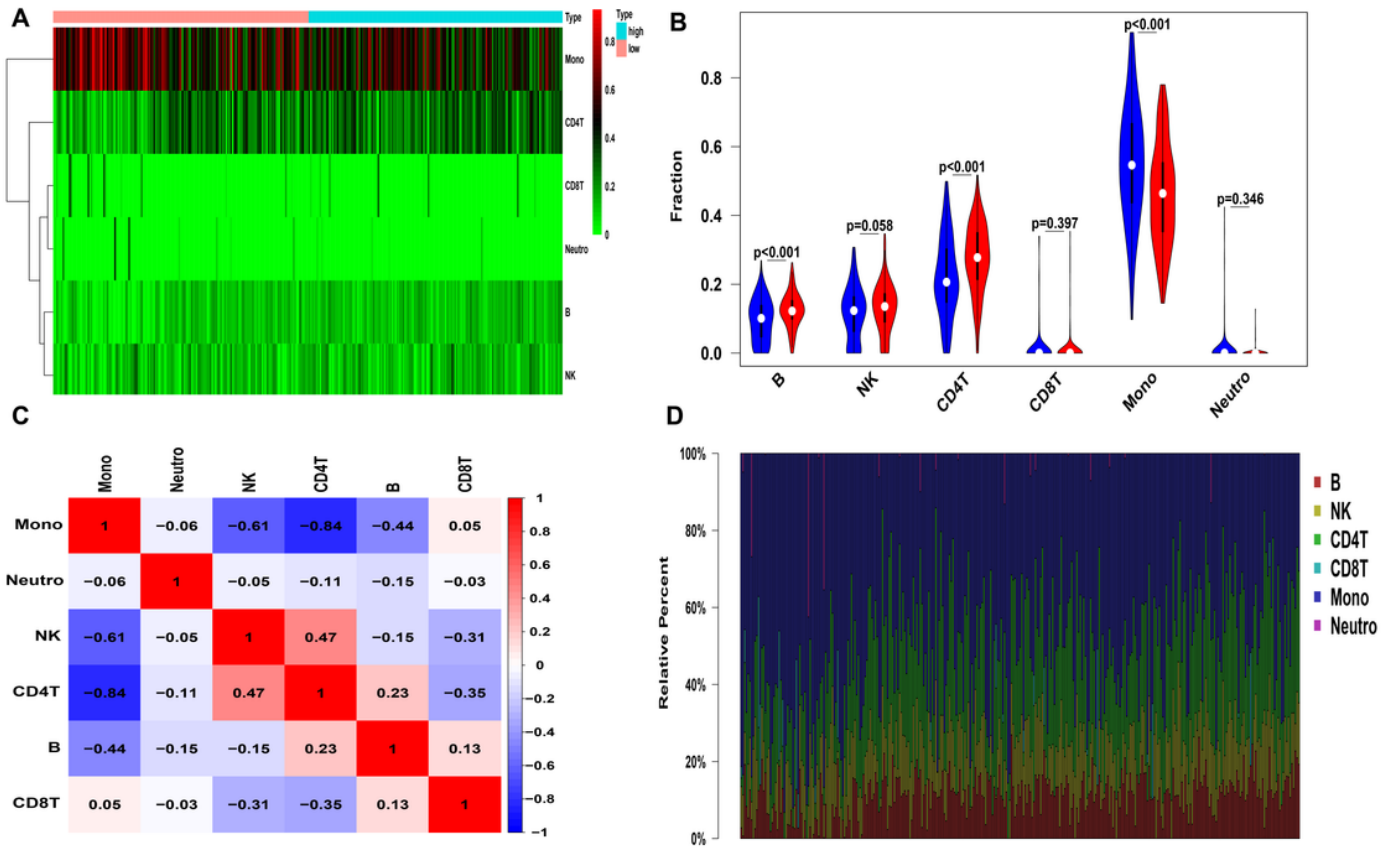

Figure 8

Infiltration of immune cells. (A) Heatmap shows the 6 immune cells' proportions in of the samples in the training dataset. **(B)** Violin plot analysis presents the distribution in the abundance of different immune cell infiltrations in two groups**. (C)** Correlation between different immune cells. **(D)**Histogram showing the individual percent of six immune infiltration cells in each sample.

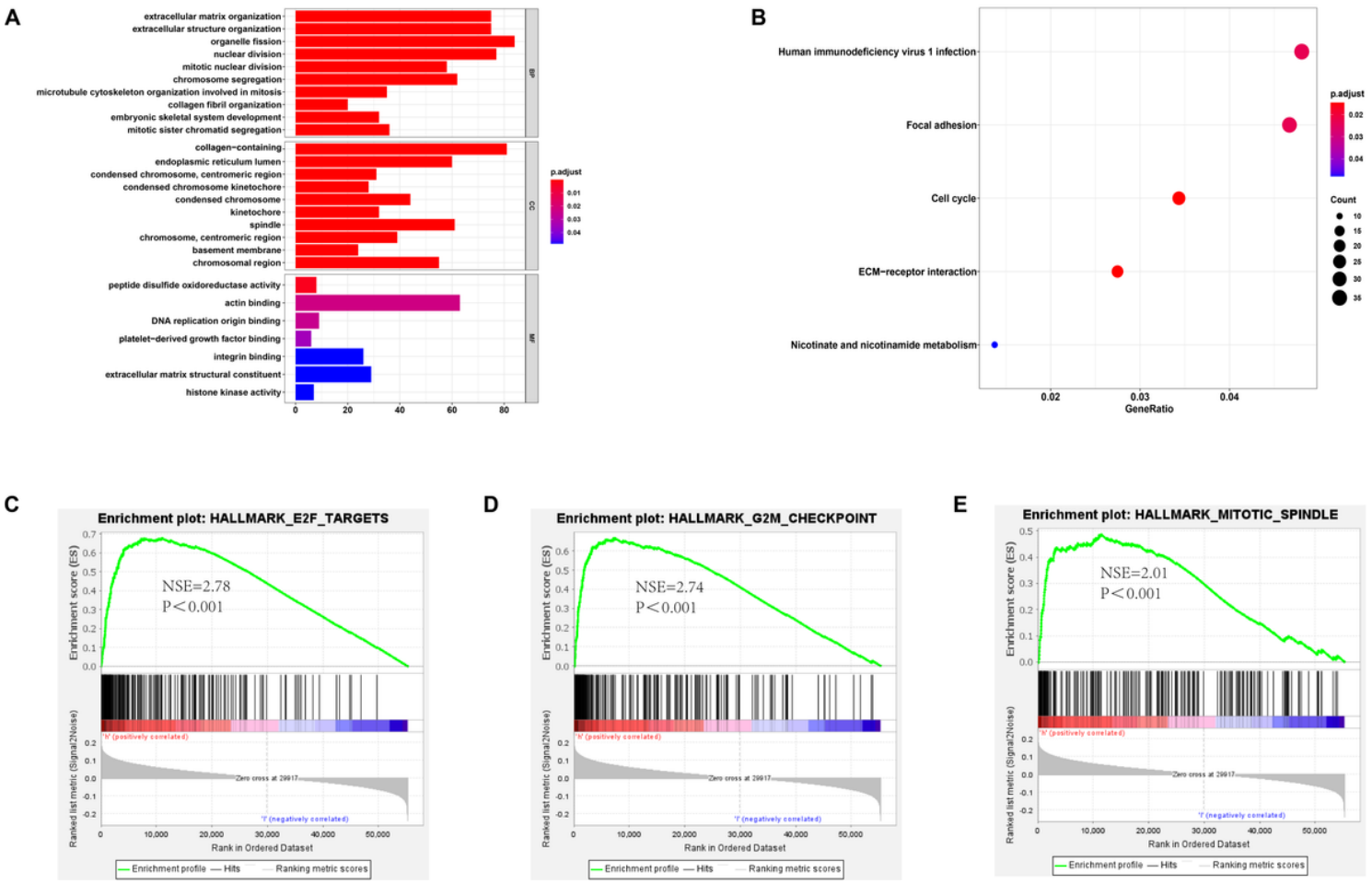

Figure 9

Functional enrichment analysis of the six-pseudogenes signature. (A) GO annotations were carried out and according to the top 1747 genes positively correlated with the six-pseudogenes model. **(B)** KEGG analyses were positively associated with risk score. **(C-E)**GSEA results showing that the high-risk group had strong characteristics of malignancy.

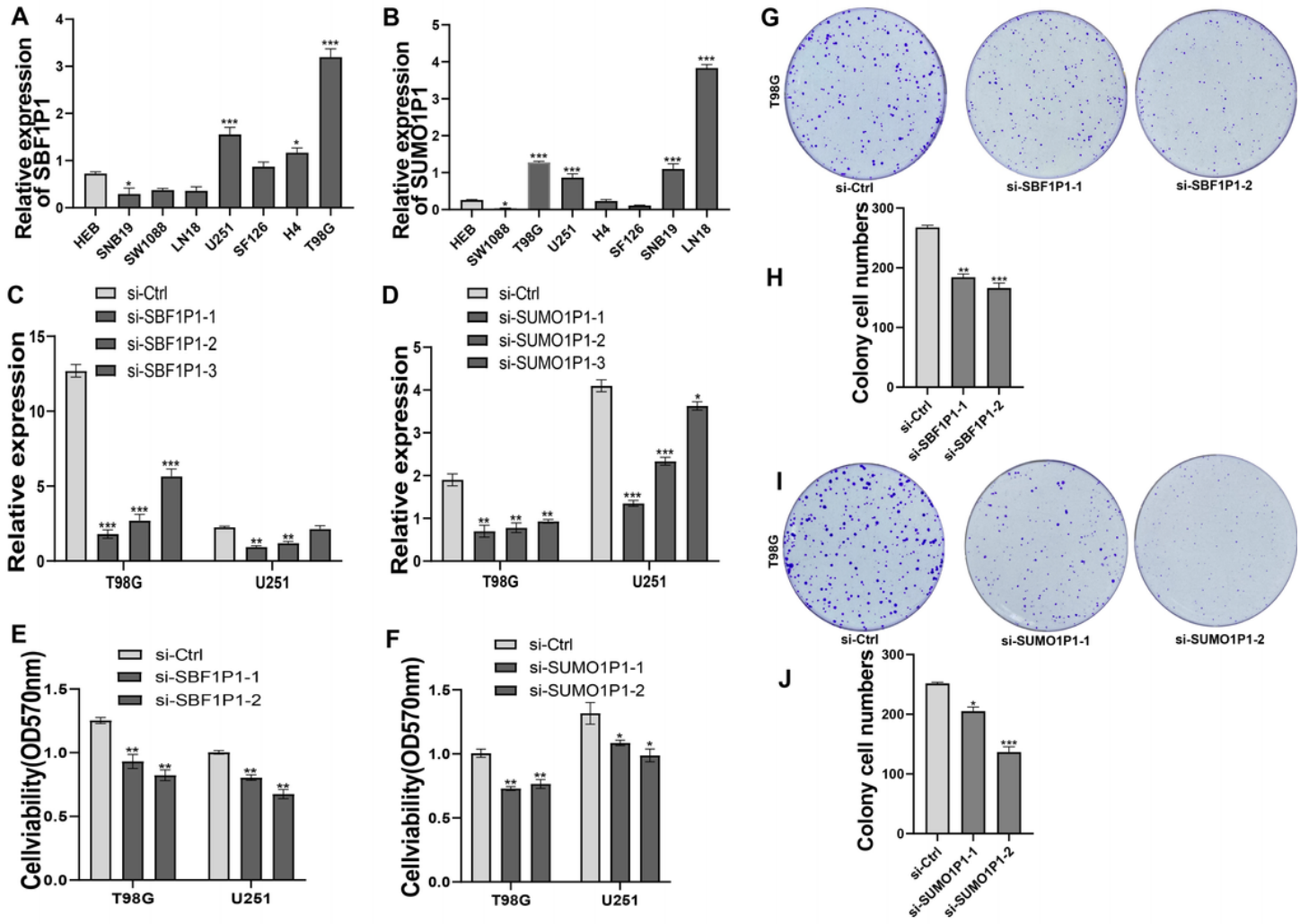

Figure 10

Knockdown of SBF1P1 and SUMO1P1 restrained the activity and proliferation of glioma cells. (A,B) Relative expression of SBF1P1 **(A)** and SUMO1P1(B) in nine cell lines. **(C,D)** The relative silencing levels of SBF1P1 **(C)** and SUMO1P1 **(D)** in T98G and U251 cells was detect by RT-qPCR. **(E,F)** Cell proliferation was performed by the MTT assay after silencing with SBF1P1 **(E)** and SUMO1P1 **(F)** in T98G and U251 cells. **(G-J)**Colony formation assays for T98G cells in which SBF1P1 (**G,H)** and SUMO1P1 **(I,J)** was silenced expressed for 14 days. ns: no significance; *P<0.05; **P<0.01; ***P<0.001.

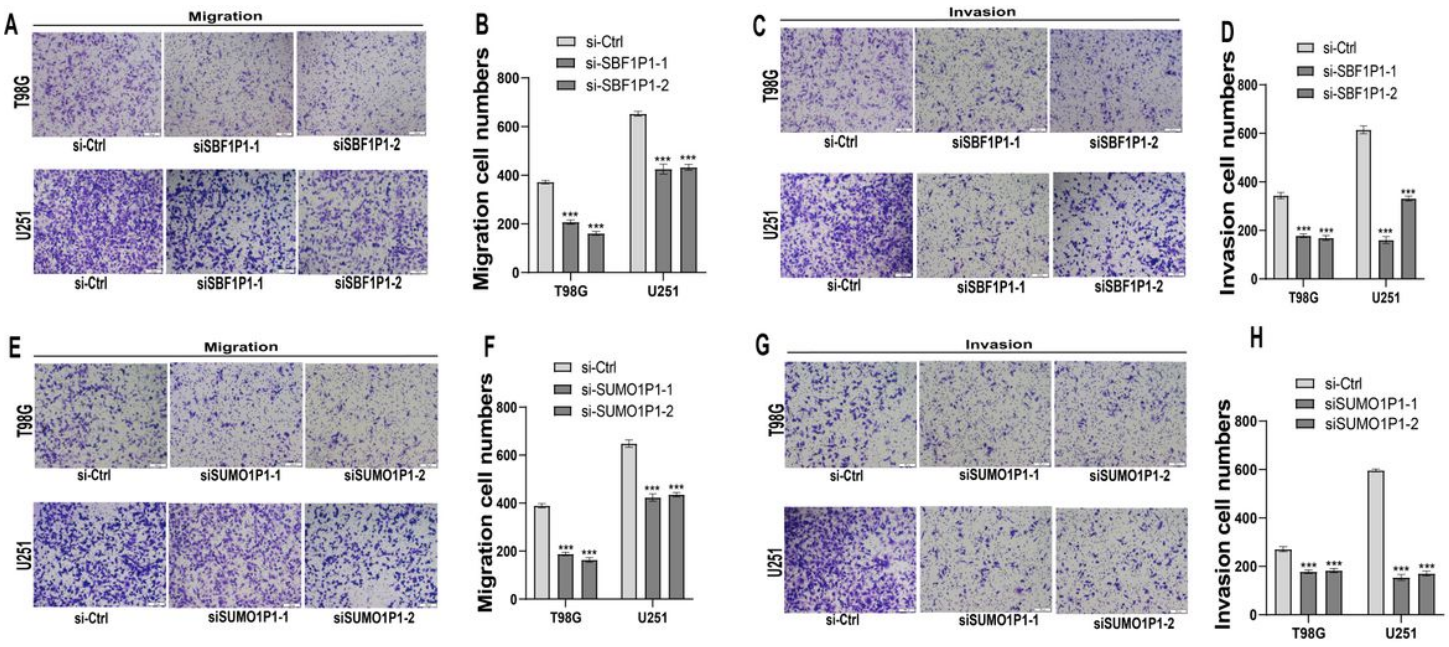

Figure 11

ilencing SBF1P1 and SUMO1P1 can restrain the ability of glioma cells to migrate and invade. (A-H) Typical imaging or counting of migration and invasion assays after silencing SBF1P1**(A-D)** and SUMO1P1**(E-H)**. Scale bar, 100μm. *P<0.05; **P<0.01; ***P<0.001.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFig1.png
- SupplementaryFig2.png
- SupplementaryFigureLegends.docx