

# Nomogram with a novel microenvironment signature is systematically constructed and validated to predict the survival rate of glioma patients.

Tianhua Li<sup>1</sup>, Yiguang Chen<sup>1</sup>, Yongjian Chen<sup>2</sup>, Guangjie Liu<sup>1</sup>, Shisheng Zou<sup>1</sup>, Anming Yang<sup>1</sup>, Yi Liu<sup>1\*</sup>, Jun Fan<sup>1\*</sup>.

<sup>1</sup>Department of Neurosurgery, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

<sup>2</sup>Department of Medical Oncology and Guangdong Key Laboratory of Liver Disease, the Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China.

All authors above are co-first authors.

\*Correspondence: Yi Liu, E-mail:278679433@qq.com; Jun Fan,E-mail:kinvan2002@163.com.

## Abstract

Glioma accounts for the highest proportion of primary intracranial malignant tumors. Microenvironment enormously influences the process of glioma progression. Our study is to establish an individualized prognostic nomogram for glioma patients with microenvironment signature. Glioma samples of Chinese Glioma Genome Atlas (CGGA) were grouped by the immune and stromal score based on ESTIMATE algorithm. Microenvironment-related genes (MRGs) in glioma were analyzed by R. To determine the best prognostic correlation genes, univariate and multivariate Cox regression analysis were used to analyze MRGs. Use the selected genes (CHI3L1, SOCS3, SLC47A2, COL3A1, SRPX2 and SERPINA3), we established the prognostic risk score model (microenvironment signature) and validated it. Gene Set Enrichment Analysis (GSEA) showed that the high-risk group was mainly enriched in immune and stromal function KEGG pathways. Finally, the nomogram was constructed and evaluated. The receiver operating characteristic (ROC) curve, Calibration plots and decision curve analysis (DCA) of training and validation set indicated the excellent predictive performance of nomogram. In conclusion, the 6-gene microenvironment signature can not only provide directions for the basic research of glioma, but also can be included as an independent prognostic index in nomogram for individual prediction to guide clinical treatment.

Keywords: microenvironment signature, prognostic model, glioma, CGGA, ESTIMATE algorithm

## 1. Introduction

In primary intracranial malignant tumors, the proportion of gliomas can be as high as 81%<sup>1</sup>. Although a lot of achievements have been made in the clinical and molecular research of glioma, there are significant deficiencies in the study on the prognostic biomarkers and a more accurate and reliable prognostic index of glioma patients is also needed.

Tumor cell internal genes play essential roles in the evolution of glioma<sup>2,3</sup>. At the same time, tumor microenvironment had vital effects on gene expression in tumor tissues<sup>4-7</sup>. Tumor microenvironment contains two main non-tumor components: immune cells and stromal cells, which are crucial for diagnosis and prognosis of tumors<sup>8,9</sup>. Many studies showed that some microenvironment-related genes (MRGs) play essential roles in glioma in many signal pathways<sup>10,11</sup>. Therefore, MRGs are expected to be clinical prognostic indicators and therapeutic targets for glioma.

Thanks to the continuous development of genome sequencing technologies, several glioma molecular biomarkers have been discovered. There have been many studies on 1p/19q codeletion, tumor protein 53 (TP53) mutations, isocitrate dehydrogenase (IDH) mutation and so on<sup>12,13</sup>. Emerging research suggests that certain single genes do not fully represent tumor characteristics, but global gene expression pattern of multigene could be used as a special molecular biological marker for subgroup classification, early diagnosis, treatment targeting, prognosis prediction and so on in glioma<sup>14,15</sup>. However, there is little research on the global expression pattern based on MRGs in glioma.

Recently, a newly proposed computational algorithm, known as “Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE)” was developed<sup>9</sup> and successfully brought to calculate the degree of infiltration of non-tumor cells in several malignant tumors like prostate cancer<sup>16</sup>, breast cancer<sup>17</sup>, and colon cancer<sup>18</sup>. Therefore, in this study, we use the ESTIMATE algorithm to evaluate the RNA sequencing data of glioma samples, construct and validate a microenvironment signature that can predict prognosis and provide research directions for therapeutic targets in glioma. Moreover, combining clinical parameters and the microenvironment signature, we established an innovative and promising predictive nomogram model, which has more accurate predictive ability for glioma.

## 2. Results

### 2.1 Identification of MRGs and enrichment analysis

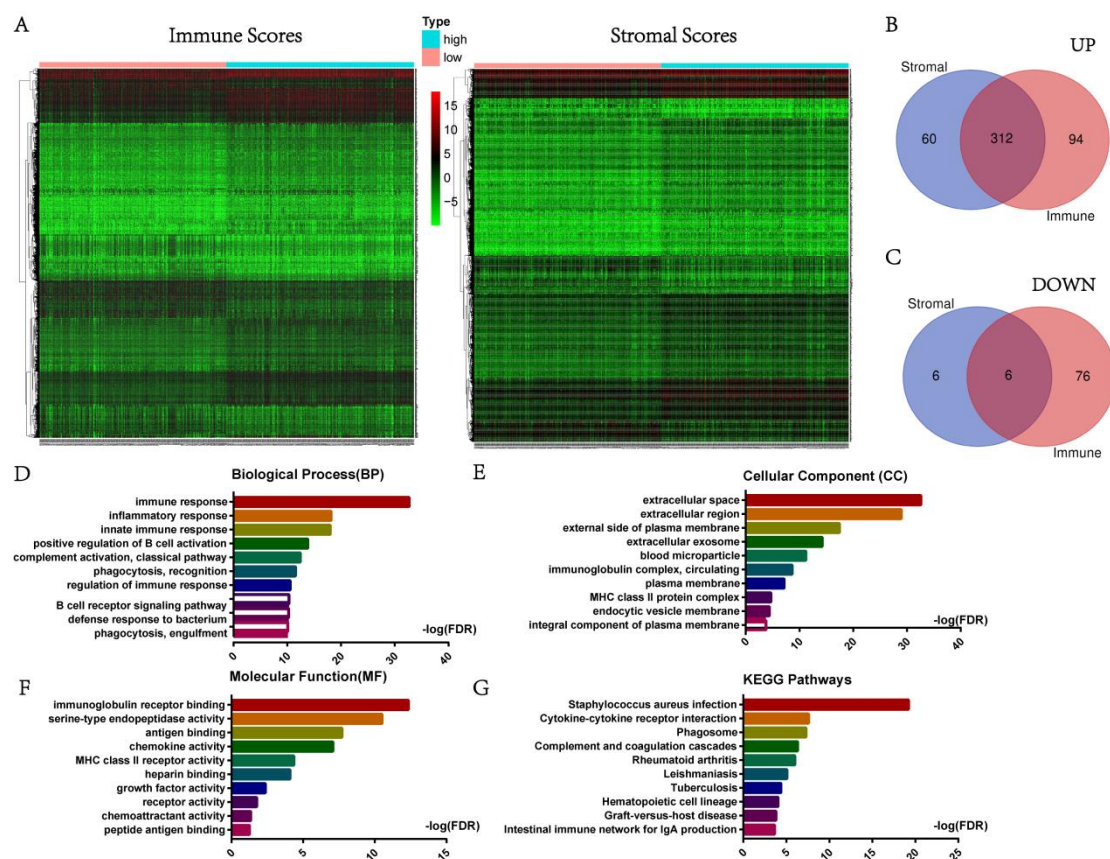
Using the immune or stromal median score as the cut-off, we divided the 693 glioma cases into high/low immune or stromal score groups. The K-M survival curve [Figure S1] showed that, whether in immune ( $p = 0.281$ ) or stromal ( $p = 0.114$ ) groups, the median overall survival of patients with high scores was lower than that of patients with low scores, although they were not statistically significant.

We compared their RNA-seq data based on the high/low immune or stromal score group. The heatmaps [Figure 1A] showed that the gene expression profiles of the cases were different. In the comparison based on immune score, high score group had 406 genes were up-regulated and 82 genes were down-regulated ( $|\log(FC)| \geq 2$ ,  $p < 0.05$ ). Likewise, based on stromal score, high score group had 372 genes were up-regulated and 12 genes were down-regulated ( $|\log(FC)| \geq 2$ ,  $p < 0.05$ ). The intersection of genes was considered to be an important set of

MRGs, for further analysis. Through the Venn diagram [Figure 1B, C], we found that there were 312 common up-regulated genes and 6 common down-regulated genes.

We used DAVID to do enrichment analysis of MRGs. In the biological process (BP), the MRGs were significantly enriched in the terms immune response, inflammatory response and innate immune response [Figure 1D]. In the cellular component (CC), the MRGs were significantly enriched in the terms extracellular space, extracellular region, and external side of plasma membrane [Figure 1E]. In the molecular function (MF), the MRGs were significantly enriched in the terms immunoglobulin receptor binding, serine-type endopeptidase activity, and antigen binding [Figure 1F]. Moreover, KEGG pathway analysis showed that the MRGs were mainly enriched in Phagosome, Cytokine-cytokine receptor interaction and Phagosome [Figure 1G].

**Figure1. Screening and enrichment analysis of microenvironment related genes (MRGs).** (A) Heatmap of the DEGs of immune/ stromal high score vs. low score ( $p < 0.05$ , fold change  $> 2$ ). Venn diagrams show the number of common upregulated (B) and downregulated (C) The intersection of genes (MRGs) in immune and stromal score groups. (D,E,F,G) Top 10 GO terms and KEGG pathways ( $p < 0.05$ ) of MRGs enriched by DAVID.

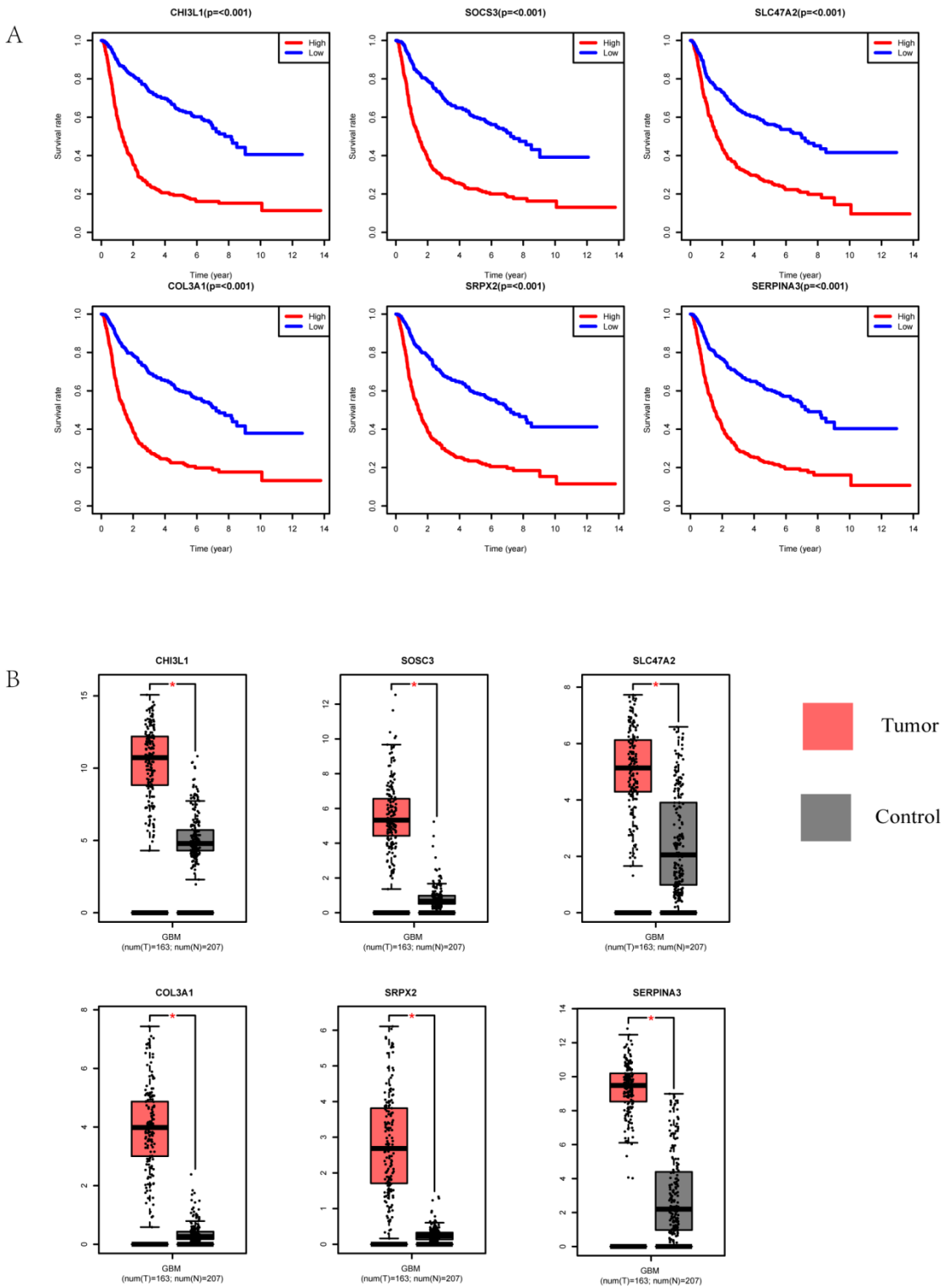


## 2.2 Identification of prognosis-related MRGs

We excluded patients with loss of age and survival time in cohort 1 and performed univariate Cox regression analysis on the 318 MRGs. The significant prognostic genes ( $P < 0.05$ ) were arranged in ascending order, and top 10 genes related to prognosis were identified and

analyzed by multivariate Cox regression analysis. After the analysis, optimal prognosis genes were used to construct the proportional risk model, which includes CHI3L1(coef=0.08507), SOCS3(coef=0.06909), SLC47A2(coef=0.13418), COL3A1(coef=0.09616), SRPX2(coef=-0.12213) and SERPINA3(coef=0.05121) (Table S1). Then, using the median expression quantity of 6 genes as the cut-off, we separated glioma patients into high/low expression groups and constructed the K-M curve. The results showed that the patients with low expression of CHI3L1, SOCS3, SLC47A2, COL3A1, SRPX2 and SERPINA3 had a better prognosis [Figure 2A]. At the same time, we further confirm that these 6 genes were significantly up-regulated in glioma tissues (163) compared with normal tissues (207) on Gene Expression Profiling Interactive Analysis (GEPIA) database<sup>19</sup> [Figure 2B].

**Figure2. Expression and survival analysis for CHI3L1, SOCS3, SLC47A2, COL3A1, SRPX2 and SERPINA3 in glioma.** (A) K-M survival curves based on the expression levels of the 6 genes in glioma patients in the training set. (B) The expression levels of the six genes in tumor and normal tissues were validated in the GEPIA database.

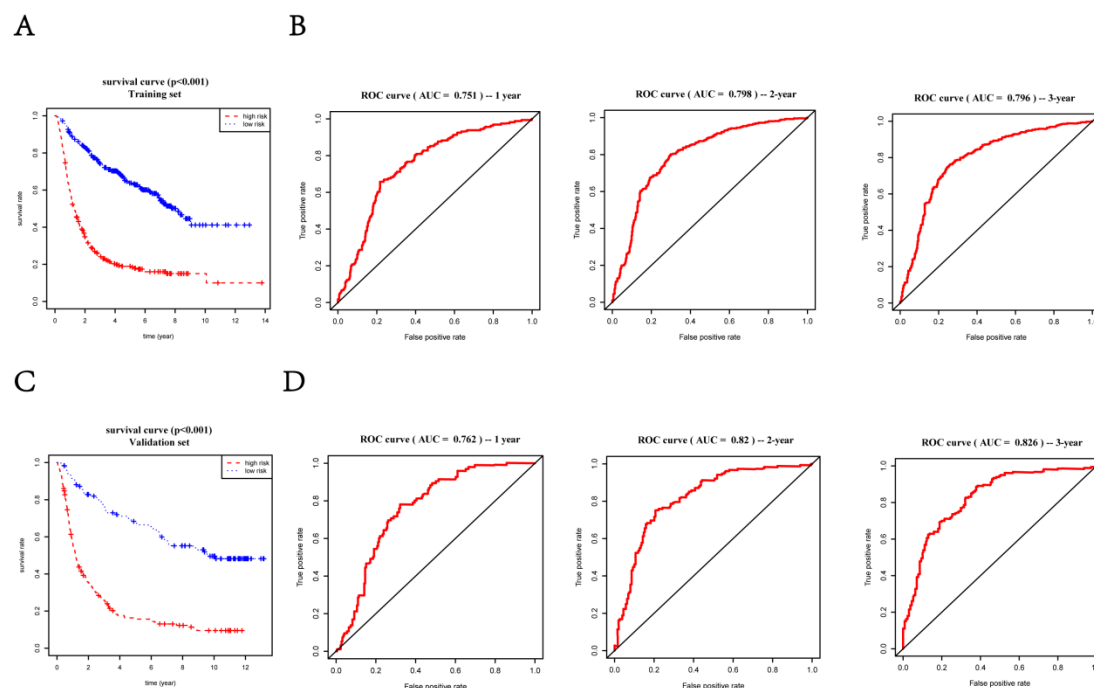


**2.3 Construct and validate prognostic risk scoring model (microenvironment signature) based on MRGs**

We graded each patient according to the risk score formula and divided them into high/low-risk group, using the median risk score as the cut-off point. K-M curve indicated that the OS of patients in high-risk group was notably worse than low-risk group ( $P < 0.001$ ) [Figure 3A]. Moreover, the microenvironment signature showed a favorable predictive ability to predict the

OS rates in 1, 2 and 3 years with the AUC values 0.751, 0.798 and 0.796, respectively [Figure 3B]. We treat cohort 2 as an external validation cohort and verify it in the same way. K-M curve also demonstrates that OS of patients in high-risk group was markedly worse ( $P < 0.001$ ) [Figure 3C]. And microenvironment signature showed a favorable predictive ability to predict the OS rates in 1, 2 and 3 years with the AUC values in the 0.762, 0.82 and 0.826, respectively [Figure 3D].

**Figure3. Survival analysis and prognostic evaluation of the microenvironment signature in glioma.** K-M survival curve of the risk score for patient OS in the training (A) and validation set (C). The OS of patients in high risk group was significantly worse than low risk group. The prognostic evaluation of the microenvironment signature displayed by the ROC curve for predicting the 1,2 and 3-year OS rates in the training (B) and validation set (D).

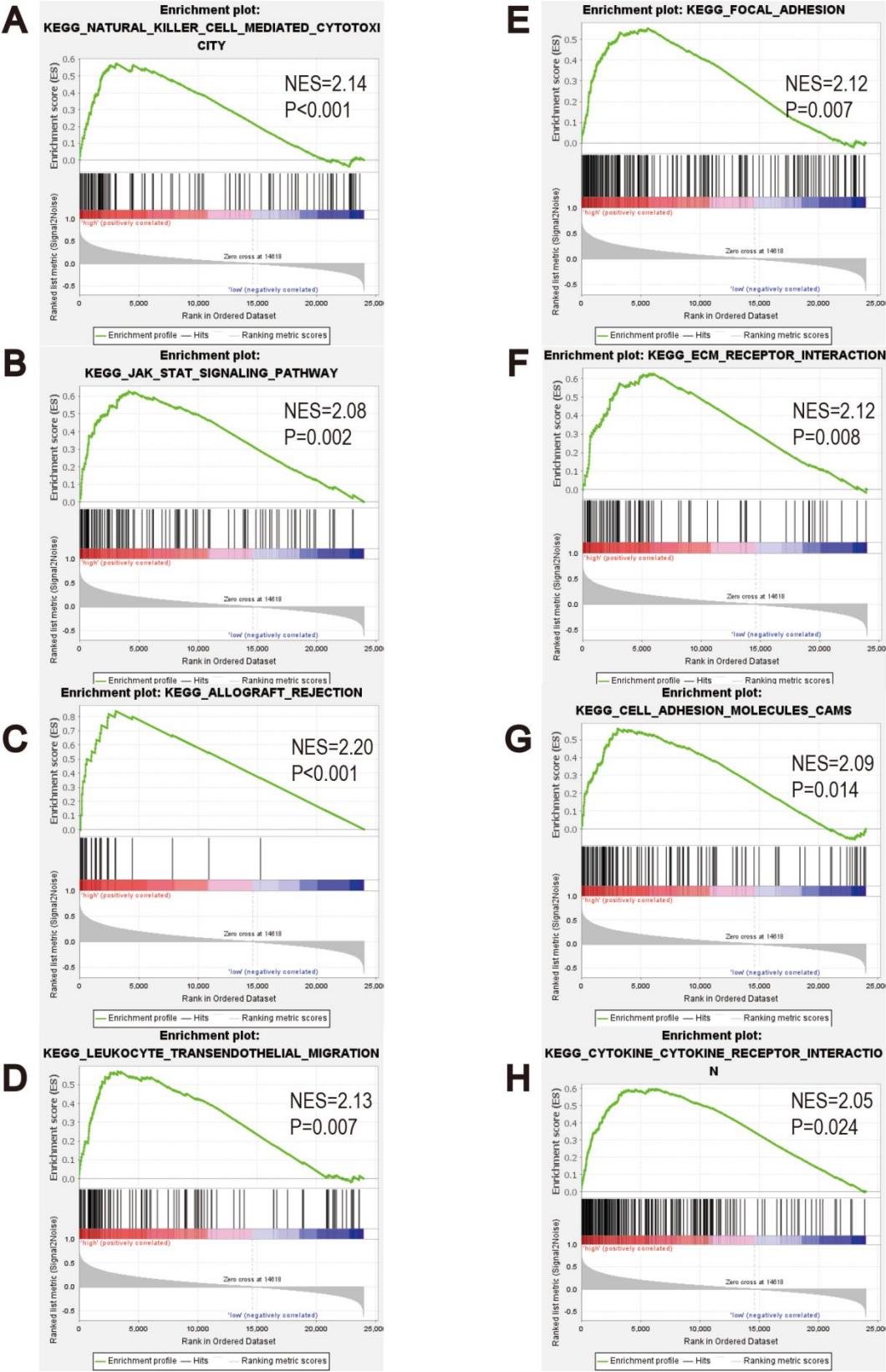


## 2.4 GSEA

GSEA indicated that the gene sets of high-risk group in the cohort 1 were significantly enriched immunity-related pathways included natural killer cell mediate cytotoxicity, JAK/STAT signaling pathway, allograft rejection, leukocyte trans endothelial migration [Fig4 A-D]. The significantly enriched stromal-related pathways included focal adhesion, ECM receptor interaction, cell adhesion molecules cams, cytokine-cytokine receptor interaction [Fig4 E-H]. In summary, these 6-high expression of MRGs in glioma tissue, led to a relatively worse prognosis in high-risk groups through the immune and stromal regulation. And we also inferred that these 6 MRGs could be the potential therapeutic targets for glioma.

**Figure4. GSEA analysis base on high risk group vs. low risk group.** the genesets of high-risk group in the cohort 1 were significantly enriched immunity-related pathways (A-D) and stromal-related pathways (E-H)





**2.5 Identify the microenvironment signature as an independent prognostic factor**  
Base on the microenvironment signature, we counted the clinical information of glioma

patients in the cohort 1 and cohort 2 (Table1). Then, we performed univariate and multivariate Cox regression analysis on them (Table 2). After analyses we found that the microenvironment signature ( $P < 0.001$ ), age ( $P = 0.007$ ) and IDH mutation status ( $P < 0.001$ ) were significantly correlated with OS. Therefore, the microenvironment signature constructed by the training cohort could be an independent prognostic factor for glioma patients. Then, we conducted the same analyses to the validation cohort, and also got the same result (Table 2).

Table 1. Demographics and clinicopathologic characteristics of patients in the training and validation set based on the risk score group.

Variables	Training set			Validation set		
	Total (n=656)	Low risk (n=327)	High risk (n=329)	Total (n=313)	Low risk (n=119)	High risk (n=194)
Age						
<50	463	266	197	216	100	116
>=50	193	61	132	97	19	78
Gender						
Female	283	150	133	116	49	67
Male	373	177	196	197	70	127
Radio status						
Treated	500	246	254	241	99	142
Untreated	131	72	59	62	19	43
NA	25	9	16	10	1	9
Chemo status (TMZ)						
Treated	479	221	258	190	66	124
Untreated	156	97	59	110	49	61
NA	21	9	12	13	4	9
IDH mutation status						
Mutant	333	241	92	167	103	64
Wildtype	275	51	224	145	15	130
NA	48	35	13	1	1	0
MGMTp methylation status						
Methylated	304	156	148	152	65	87
Unmethylated	217	102	115	143	47	96
NA	135	69	66	18	7	11

NA, not available; TMZ, Temozolomide.

Table 2. Univariate and multivariate cox proportional hazards analysis of clinical parameters and microenvironment signature of patients in the training and validation set.



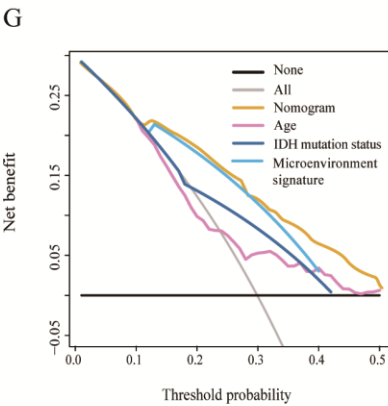
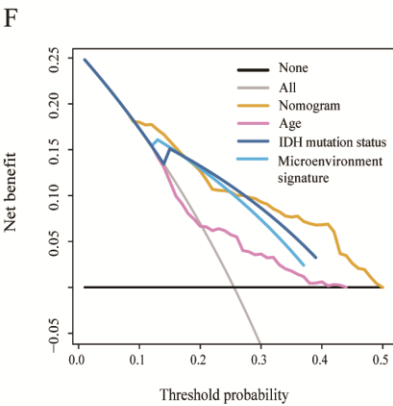
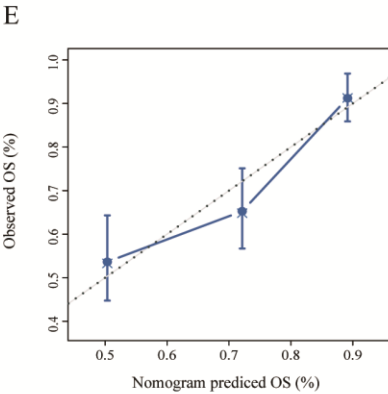
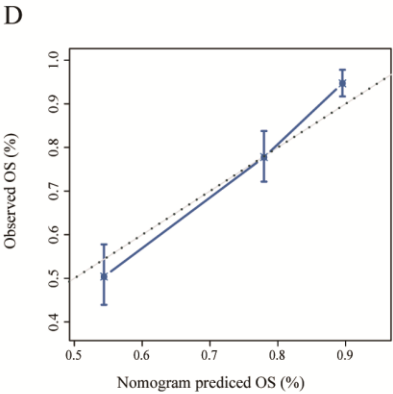
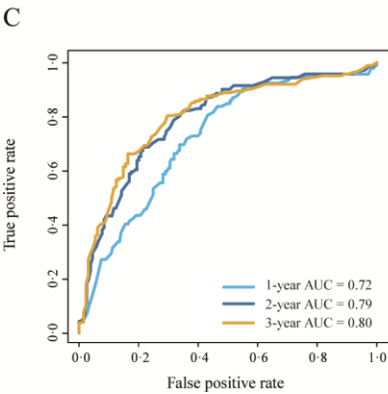
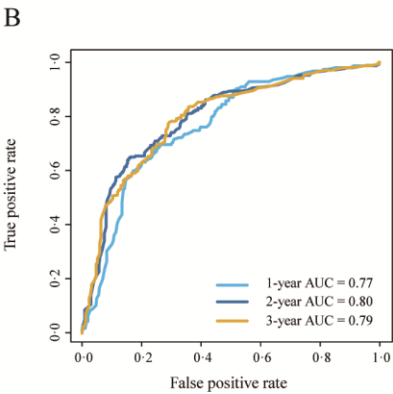
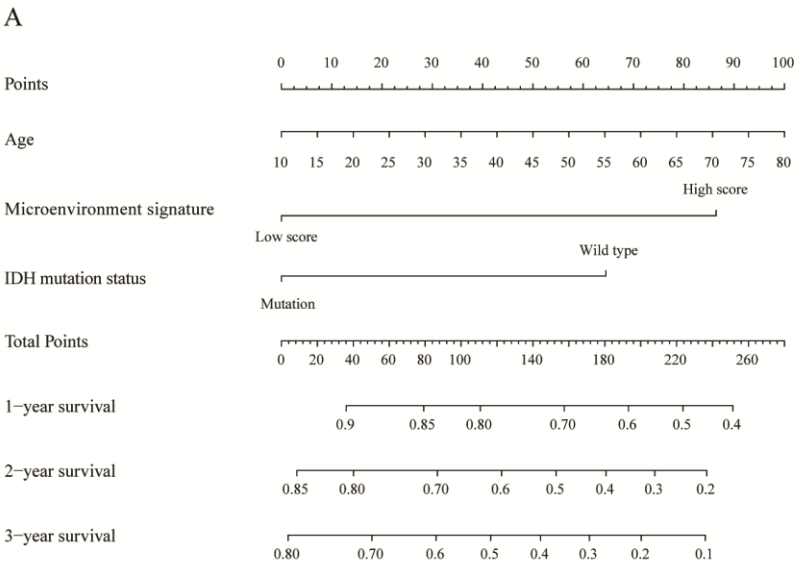
Value <sup>⓪</sup>	Univariate analysis <sup>⓪</sup>		Multivariate analysis <sup>⓪</sup>		Univariate analysis <sup>⓪</sup>		Multivariate analysis <sup>⓪</sup>	
	HR (95%CI) <sup>⓪</sup>	P value <sup>⓪</sup>	HR (95%CI) <sup>⓪</sup>	P value <sup>⓪</sup>	HR (95%CI) <sup>⓪</sup>	P value <sup>⓪</sup>	HR (95%CI) <sup>⓪</sup>	P value <sup>⓪</sup>
	Training cohort <sup>⓪</sup>				Validation cohort <sup>⓪</sup>			
Gender <sup>⓪</sup>	1.06 (0.87-1.30) <sup>⓪</sup>	0.563 <sup>⓪</sup>	/ <sup>⓪</sup>	/ <sup>⓪</sup>	0.94 (0.72-1.24) <sup>⓪</sup>	0.660 <sup>⓪</sup>	/ <sup>⓪</sup>	/ <sup>⓪</sup>
Age <sup>⓪</sup>	1.03 (1.02-1.04) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	1.01 (1.00-1.02) <sup>⓪</sup>	0.007 <sup>⓪</sup>	1.03(1.02-1.05) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	1.02 (1.00-1.03) <sup>⓪</sup>	0.011 <sup>⓪</sup>
Microenvironment signature <sup>⓪</sup>	3.48 (2.82-4.30) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	2.31 (1.74-3.07) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	4.02 (2.94-5.51) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	3.04 (2.14-4.32) <sup>⓪</sup>	<0.001 <sup>⓪</sup>
Radiotherapy status <sup>⓪</sup>	1.24 (0.95-1.62) <sup>⓪</sup>	0.109 <sup>⓪</sup>	/ <sup>⓪</sup>	/ <sup>⓪</sup>	0.63 (0.46-0.87) <sup>⓪</sup>	0.005 <sup>⓪</sup>	0.63 (0.45-0.87) <sup>⓪</sup>	0.005 <sup>⓪</sup>
Chemotherapy status(TMZ) <sup>⓪</sup>	1.27 (0.97-1.59) <sup>⓪</sup>	0.081 <sup>⓪</sup>	/ <sup>⓪</sup>	/ <sup>⓪</sup>	1.45 (1.08-1.94) <sup>⓪</sup>	0.014 <sup>⓪</sup>	1.10 (0.81-1.49) <sup>⓪</sup>	0.548 <sup>⓪</sup>
IDH mutation status <sup>⓪</sup>	0.32 (0.26-0.40) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	0.49 (0.37-0.65) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	0.35 (0.27-0.47) <sup>⓪</sup>	<0.001 <sup>⓪</sup>	0.67(0.48-0.95) <sup>⓪</sup>	0.022 <sup>⓪</sup>
MGMTp methylation status <sup>⓪</sup>	0.80 (0.64-0.99) <sup>⓪</sup>	0.041 <sup>⓪</sup>	0.89 (0.71-1.12) <sup>⓪</sup>	0.325 <sup>⓪</sup>	0.83 (0.63-1.09) <sup>⓪</sup>	0.177 <sup>⓪</sup>	/ <sup>⓪</sup>	/ <sup>⓪</sup>

TMZ, Temozolomide.

2.6 Construct and verify the nomogram

Based on the training cohort, we established a prognostic nomogram, which can predict the 1, 2 and 3-year survival probability of glioma patients. We enrolled three independent prognostic parameters in the prognostic nomogram: age, microenvironment signature and IDH status [Figure 5A]. The C-index of the nomogram for OS prediction was 0.707 (se = 0.013) and the nomogram had a favorable ability to predict the OS rates of 1, 2 and 3-year, and their AUC values were 0.77, 0.80 and 0.79 respectively. Similarly, in the validation cohort, the AUC values were 0.72, 0.79, and 0.80 respectively [Figure 5B, C]. At the same time, there was an excellent agreement between the nomogram prediction and the actual observation in terms of 1, 2 and 3-year survival rates, which could be found in the calibration curves of the training and validation cohorts [Figure 5D, E]. Ultimately, the DCA curve showed this nomogram model had better clinical benefits [Figure 5F, G].

**Figure5. Nomogram to predict the survival probability of glioma patients and the evaluations of it.** (A) Prognostic nomogram to predict 1, 2 and 3-year of the OS of glioma patients based on the raining set. ROC curve of nomogram for predicting survival in the raining (B) and validation set (C). Calibration curve of nomogram to predict 1, 2 and 3-year of the OS of glioma patients in the training (D) and validation set (E). The observed OS (%) is plotted on the y-axis; the nomogram predicted OS (%) is plotted on the x-axis. The DCA curve of the nomogram predicting total OS in the training (F) and validation set (G). Among all the areas formed by the curves and "None" and "All", the nomogram curve is the largest, which showed that the prediction ability of nomogram model is better than that of single parameter model.



### 3. Discussion

The tumor microenvironment of glioma plays an essential role in the development of glioma. The change of microenvironment related genes can affect the expression of tumor tissue, and then affect the clinical outcome<sup>4,5</sup>. Therefore, MRGs are promising prognostic indicators and treatments target for glioma.

With the progress of high-throughput sequencing technology, more and more biomarkers related to the survival of glioma patients have been identified<sup>12,13</sup>. Many global gene expression patterns can be used in prognosis prediction, risk stratification and treatment guidance of glioma<sup>20,21</sup>. However, there is still a lot of room for us to study the global expression pattern based on MRGs in glioma.

ESTIMATE is a bioinformatics tool for predicting non-tumor cell infiltration. It can score each sample by evaluating the particular gene expression feature of stromal and immune cells<sup>9</sup>. In this study, we first use ESTIMATE to grade the cohort 1 samples. Taking the median score as the dividing value, the samples were partitioned for high/low immune or stromal score group. Then, we regard the 318 intersection genes of DEGs between the immune and stromal group as MRGs. Finally, univariate and multivariate Cox regression analysis confirmed that 6 up-regulated genes (CHI3L1, SOCS3, SLC47A2, COL3A1, SRPX2 and SERPINA3) were significantly related to prognosis. TCGA-based GEPIA also proved that they are up-regulated in glioma compared to normal tissues. GSEA showed that gene sets of the high-risk group in the cohort 1 were chiefly enriched in immune and stromal related KEGG pathways, which suggested that glioma with high expression of 6 kinds of MRGs can affect tumor progression through related pathways. CHI3L1, also known as YKL-40, is a pro-inflammatory factor that can be used as a biomarker of glioma and brain injury. High levels expression of YKL-40 in human gliomas can activate AKT<sup>22</sup>. Angiogenesis and malignancy of glioblastoma can be synergistically inhibited by Anti-YKL-40 antibody and ionizing irradiation<sup>23</sup>. SOCS3 is related to tumor progression and therapeutic response in glioma, and it crucial for glioma to acquire anti-radioactivity. Hypermethylation of SOCS3 promoter is an important marker of poor prognosis in glioma<sup>24,25</sup>. The expression of SLC47A2 can be cis-regulated in renal cell carcinoma<sup>26</sup>. Type III collagen is an important signal molecule to promote wound healing, and COL3A1 encodes its alpha 1 chain<sup>27</sup>. SRPX2 enhances EMT process and promotes glioma metastasis through MAPK signaling pathway<sup>28</sup>. The upregulation of SERPINA3 might reshape the extracellular tissue matrix and promote the invasion of glioma, and it was significantly related to the poor survival of patients<sup>29</sup>. In summary, CHI3L1, SOCS3, SRPX2 and SERPINA3 were significantly associated with the evolution of glioma. However, SLC47A2 and COL3A1 had not been studied in glioma. These MRGs can be used not only as independent prognostic biomarkers but also as potential targets to guide the treatment of glioma.

Then, based on the expression of 6 MRGs, we developed and validated a novel risk score model (microenvironment signature) and separated glioma patients into low/high-risk group based on their risk score. Subsequently, the K-M curve showed that the high-risk group had an appreciably poorer prognosis. Therefore, glioma patients with high-risk scores should receive more attention and adopt more aggressive individualized medical strategies. At the same time,

they need to be closely followed up to detect recurrence.

Nomogram can intuitively show the prognosis, which makes it widely used in clinical practice<sup>30</sup>. In this study, we constructed and verified a nomogram based on microenvironment signature, IDH mutation status and age. As far as we know, the nomogram is an innovative combination of microenvironment signature and clinical parameters, which can individually and more precisely predict the survival rate of glioma patients. The ROC curve showed that the nomogram has an excellent ability to predict the OS rate of 1-, 2- and 3- years. The calibration curve showed that the prediction of the nomogram is in outstanding agreement with the actual observation, and the DCA curve showed that the nomogram model was better than the single parameter model. Combining the results of these three indicators, the innovative and promising nomogram demonstrates excellent prediction ability.

There is no denying that there are still some deficiencies in our research. First, the data we download from CGGA is incomplete and limited. Some clinical information of some patients is missing and some clinical parameters, such as operation method, tumor location, tumor size, etc. were not included in the study. Second, a limitation of this prediction model lies in its retrospective property, so it needs to be further verified in future clinical trials.

## 4. MATERIALS AND METHODS

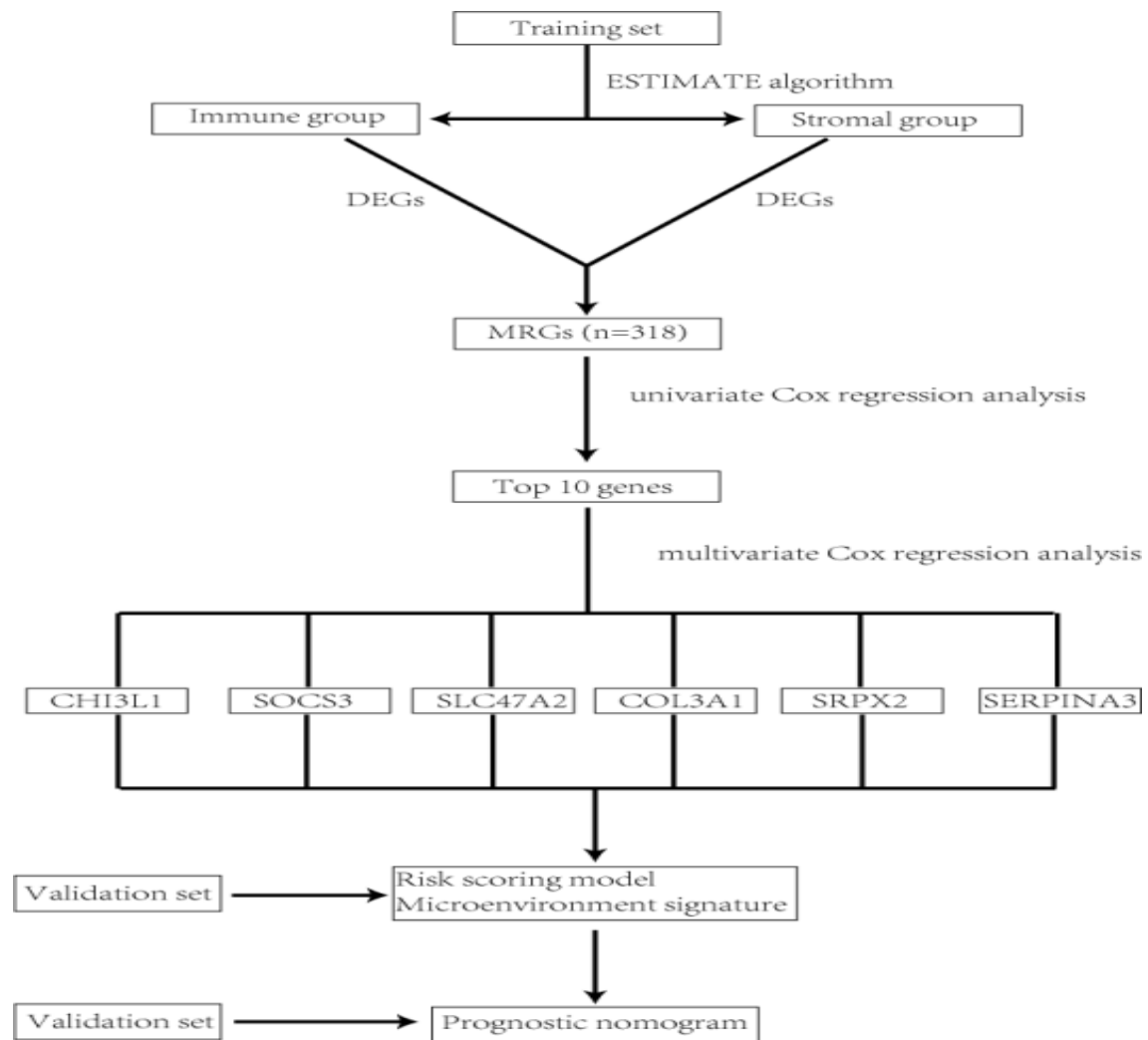
### 4.1 Database

From the Chinese Glioma Genome Atlas (CGGA, <http://www.cgga.org.cn/>) databases, we download clinical information and RNA sequencing data of glioma patients. Cohort 1 (mRNAseq\_693)<sup>31</sup> and cohort 2 (mRNAseq\_325)<sup>32</sup> were selected as training set and validation set respectively.

### 4.2 Screening and enrichment analysis of microenvironment related genes (MRGs)

**Figure 6** showed the schematic diagram for constructing the nomogram. Through the evaluation of ESTIMATE algorithm, each sample was calculated to get its own immune and stromal score. Taking the median score of immune or stromal score as the dividing point, cohort 1 was partitioned for high/low immune or stromal score group. We screened the differentially expressed genes (DEGs) in immune or stromal score group by edgeR<sup>33</sup> (R 3.6.1) (<https://bioconductor.org/>). Then, use the default Benjamini-Hochberg false discovery rate (FDR) method to correct the false-positive results. DEGs with | fold change (FC) | > 2 and FDR < 0.05 were considered to be significantly different<sup>34</sup>. We show the significant DEGs of the immune or stromal score group in the heatmaps and the intersection DEGs of the two groups in a Venn diagram and regard them as meaningful MRGs for later analysis. Then, we used DAVID (<http://david.ncifcrf.gov/>) to do Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of MRGs<sup>35</sup>. P-value < 0.05 was regarded as statistically significant.

**Figure6. The schematic diagram for constructing the prognostic nomogram based on microenvironment signature.**



#### 4.3 Construct and evaluate the prognostic risk score model of MRGs

First of all, univariate Cox regression analysis was conducted to the MRGs in training cohort 1 by using survival software package in R 3.6.1. Genes with  $P < 0.05$  were deemed as statistical significance to overall survival (OS) of glioma patients<sup>36</sup>. Then the first 10 genes with the lowest P-value were analyzed by multivariate Cox regression analysis. After the analysis we used the selected genes as the genes related to the optimal prognosis and established a prognostic risk score model to predict OS<sup>37</sup>.

The risk score was obtained according to the following formula:

$$\text{Risk score} = \sum_{i=1}^n \beta_i x_i$$

Where  $\beta_i$  and  $x_i$  are the coefficient and relative expression value of each selected gene, respectively<sup>38</sup>, and each patient could get a prognostic risk score according to this formula. According to their median risk score, glioma patients were divided into low/ high-risk group. Next, we constructed the Kaplan-Meier (K-M) survival curve of low/ high-risk group, and the survival difference between the two groups was evaluated by two-sided log-rank test. Subsequently, the prediction accuracy of the prediction model based on MRGs was evaluated for 1, 2, and 3- year by using R packet

'survivalROC'<sup>39</sup>(<https://cran.r-project.org/web/packages/survivalROC/>). The area under the ROC curve (AUC) ranges from 0.5 to 1. Among them, 0.5 means no discrimination and 1 means perfect discrimination. Finally, we used the similar approach to validate the performance of the risk scoring model built by cohorts 1 in cohorts 2.

#### 4.4 Gene set enrichment analysis (GSEA)

Use the gene sets database (c2. cp. kegg. v7. 0. symbols. gmt.), and we performed GSEA (high-risk group VS. low-risk group) (<http://software.broadinstitute.org/gsea/index.jsp>) with permutation = phenotype, permutation = 1000. Determination of statistically significant enriched gene sets based on  $P < 0.05$  and  $FDR < 0.25$  as criteria<sup>40</sup>.

#### 4.5 Construction and validation of the nomogram

We combined the MRGs-based prognostic model (microenvironment signature) with other clinicopathological parameters of glioma patients for univariate and multivariate Cox proportional hazard regression analysis in the cohort 1 and cohort 2. After the analyses, we screened out all independent prognostic factors and used the rms R package (<https://cran.r-project.org/web/packages/rms/>) to construct a nomogram of these independent prognostic factors to evaluate the probability of 1, 2 and 3-year OS in cohort 1 glioma patients<sup>14</sup>. The discriminant ability of nomogram was graphically evaluated by using C-index, the AUC value, Calibration plots and decision curve analysis (DCA)<sup>39,41,42</sup>. Finally, cohort 2 was used as an external verification of the prognostic nomogram. All analyses were performed with R, and  $P < 0.05$  was deemed to be statistically significant. Hazard ratios (HRs) and 95% confidence intervals (CIs) were also stated.

### 5. Conclusion

In this study, a promising nomogram contains novel microenvironment signature was constructed and validated for glioma individual prognostic assessment. Further bioinformatics analysis of these MRGs and microenvironment will help to clarify its possible survival mechanism. Next, this model will be further verified in clinical trials and is likely to be translated into meaningful practice to guide the individualized treatment of glioma patients.

#### Supplementary Materials:

**Figure S1. The overall survival of glioma patients based on immune or stromal score groups.** Glioma cases were divided into high/low score groups based on their immune or stromal score. As shown in the K-M survival curves, p value of immune score group is 0.281(A), of stromal score group is 0.114(B).



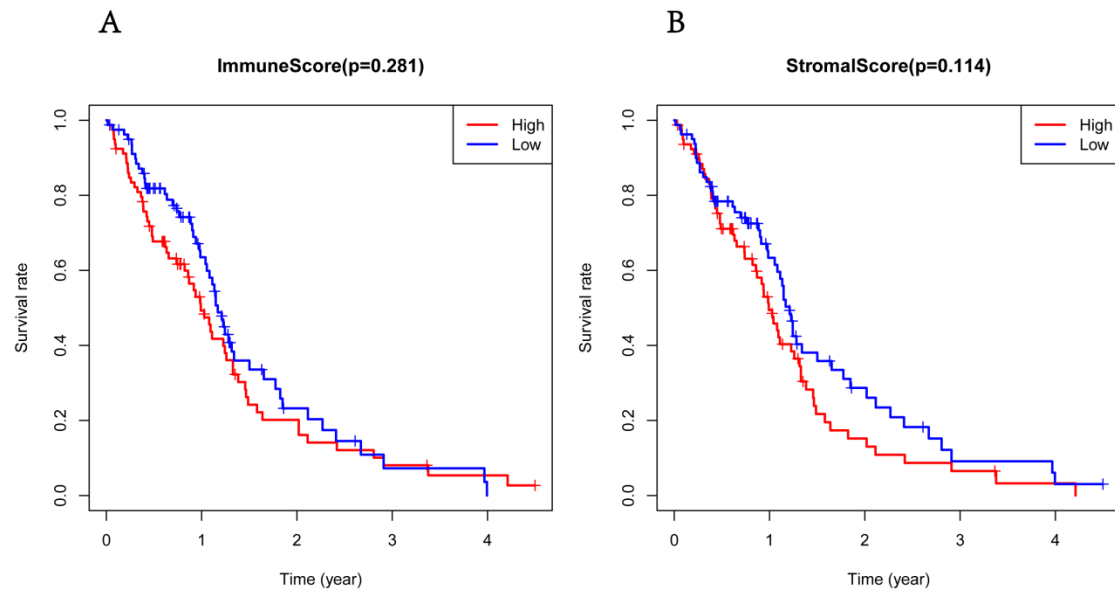


Table S1. Top 10 genes related to prognosis after univariate Cox regression analysis and the Regression coefficient (coef) after multivariate Cox regression analysis.

gene	pvalue	coef
CHI3L1	5.69E-38	0.08507
SOCS3	5.09E-32	0.06909
HSPA6	1.40E-29	/
SERPINE1	1.68E-29	/
SLC47A2	4.82E-29	0.13418
SPOCD1	5.51E-28	/
COL3A1	5.59E-28	0.09616
ABCC3	1.04E-27	/
SRPX2	2.01E-27	-0.12213
SERPINA3	2.83E-27	0.05121

**Author Contributions:** Tianhua Li, Yiguang Chen and Yongjian Chen performed the data curation and analysis. Shisheng Zou and Anming Yang analyzed and interpreted the results. Tianhua Li drafted the manuscript. Guangjie Liu, Yi Liu and Jun Fan reviewed the manuscript. All authors read and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no funding.

**Acknowledgments:** The authors would like to acknowledge the illustrations by Jin Jin.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Abbreviations**

CGGA Chinese Glioma Genome Atlas  
 MRGs Microenvironment-related genes  
 GSEA Gene Set Enrichment Analysis  
 ROC receiver operating characteristic  
 DCA decision curve analysis  
 TP53 tumor protein 53  
 IDH isocitrate dehydrogenase  
 ESTIMATE Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data  
 DEGs differentially expressed genes  
 FDR false discovery rate  
 GO Gene Ontology  
 KEGG Kyoto Encyclopedia of Genes and Genomes  
 OS overall survival  
 K-M Kaplan-Meier  
 AUC area under the ROC curve  
 HRs Hazard ratios  
 Cis confidence intervals  
 BP biological process  
 CC cellular component  
 MF molecular function  
 GEPIA Gene Expression Profiling Interactive Analysis

## Reference

1. Ostrom QT, Bauchet L, Davis FG, et al. The epidemiology of glioma in adults: a “state of the science” review. *Neuro-oncology* 2014;16:896-913.
2. Suvà ML, Rheinbay E, Gillespie SM, et al. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* 2014;157:580-94.
3. Wang Q, Hu B, Hu X, et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer cell* 2017;32:42-56.e6.
4. Cooper LA, Gutman DA, Chisolm C, et al. The tumor microenvironment strongly impacts master transcriptional regulators and gene expression class of glioblastoma. *The American journal of pathology* 2012;180:2108-19.
5. Curry JM, Sprandio J, Cignetti D, et al. Tumor microenvironment in head and neck squamous cell carcinoma. *Seminars in oncology*; 2014: Elsevier. p. 217-34.
6. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer cell* 2012;21:309-22.
7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *cell* 2011;144:646-74.
8. Galon J, Pagès F, Marincola FM, et al. The immune score as a new possible approach for the classification of cancer. *Springer*; 2012.
9. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and

immune cell admixture from expression data. *Nature communications* 2013;4:1-11.

10. Chang AL, Miska J, Wainwright DA, et al. CCL2 produced by the glioma microenvironment is essential for the recruitment of regulatory T cells and myeloid-derived suppressor cells. *Cancer research* 2016;76:5671-82.

11. Qian J, Wang C, Wang B, et al. The IFN- $\gamma$ /PD-L1 axis between T cells and tumor microenvironment: hints for glioma anti-PD-1/PD-L1 therapy. *Journal of neuroinflammation* 2018;15:1-13.

12. Liu X-Y, Gerges N, Korshunov A, et al. Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations. *Acta neuropathologica* 2012;124:615-25.

13. Smith JS, Perry A, Borell TJ, et al. Alterations of chromosome arms 1p and 19q as predictors of survival in oligodendrogliomas, astrocytomas, and mixed oligoastrocytomas. *Journal of Clinical Oncology* 2000;18:636-.

14. Wang Z, Gao L, Guo X, et al. Development and validation of a nomogram with an autophagy-related gene signature for predicting survival in patients with glioblastoma. *Aging (Albany NY)* 2019;11:12246.

15. Gao W-Z, Guo L-M, Xu T-Q, Yin Y-H, Jia F. Identification of a multidimensional transcriptome signature for survival prediction of postoperative glioblastoma multiforme patients. *Journal of translational medicine* 2018;16:368.

16. Shah N, Wang P, Wongvipat J, et al. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. *Elife* 2017;6:e27861.

17. Priedigkeit N, Watters RJ, Lucas PC, et al. Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases. *JCI insight* 2017;2.

18. Alonso MH, Aussó S, Lopez-Doriga A, et al. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *British journal of cancer* 2017;117:421-31.

19. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research* 2017;45:W98-W102.

20. Maher EA, Brennan C, Wen PY, et al. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer research* 2006;66:11502-13.

21. Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *science* 2008;321:1807-12.

22. Wang Y, Wong CW, Yan M, et al. Differential regulation of the pro-inflammatory biomarker, YKL-40/CHI3L1, by PTEN/Phosphoinositide 3-kinase and JAK2/STAT3 pathways in glioblastoma. *Cancer letters* 2018;429:54-65.

23. Shao R, Francescone R, Ngernyung N, et al. Anti-YKL-40 antibody and ionizing irradiation synergistically inhibit tumor vascularization and malignancy in glioblastoma. *Carcinogenesis* 2014;35:373-82.

24. Feng Y, Wang Z, Bao Z, et al. SOCS3 promoter hypermethylation is a favorable prognosticator and a novel indicator for G-CIMP-positive GBM patients. *PloS one* 2014;9:e91829.

25. Ventero MP, Fuentes-Baile M, Quereda C, et al. Radiotherapy resistance acquisition in Glioblastoma. Role of SOCS1 and SOCS3. *PloS one* 2019;14:e0212581.
26. Gao Z, Chen M, Tian X, et al. A novel human lncRNA SANT1 cis-regulates the expression of SLC47A2 by altering SFPQ/E2F1/HDAC1 binding to the promoter region in renal cell carcinoma. *RNA biology* 2019;16:940-9.
27. Kuivaniemi H, Tromp G. Type III collagen (COL3A1): Gene and protein structure, tissue distribution, and associated diseases. *Gene* 2019;707:151-71.
28. Tang H, Zhao J, Zhang L, Zhao J, Zhuang Y, Liang P. SRPX2 enhances the epithelial–mesenchymal transition and temozolomide resistance in glioblastoma cells. *Cellular and molecular neurobiology* 2016;36:1067-76.
29. Li Y, Dong X, Cai J, et al. SERPINA3 induced by astroglia/microglia co-culture facilitates glioblastoma stem-like cell invasion. *Oncology Letters* 2018;15:285-91.
30. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *The lancet oncology* 2015;16:e173-e80.
31. Wang Y, Qian T, You G, et al. Localizing seizure-susceptible brain regions associated with low-grade gliomas using voxel-based lesion-symptom mapping. *Neuro-oncology* 2015;17:282-8.
32. Bao Z-S, Chen H-M, Yang M-Y, et al. RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome research* 2014;24:1765-73.
33. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-40.
34. Feng H, Gu Z-Y, Li Q, Liu Q-H, Yang X-Y, Zhang J-J. Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. *Journal of ovarian research* 2019;12:35.
35. Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 2009;4:44.
36. Linden A, Yarnold PR. Modeling time - to - event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice* 2017;23:1299-308.
37. Nagashima K, Sato Y. Information criteria for Firth's penalized partial likelihood approach in Cox regression models. *Statistics in medicine* 2017;36:3422-36.
38. Itzykson R, Kosmider O, Renneville A, et al. Prognostic score including gene mutations in chronic myelomonocytic leukemia. *Journal of clinical oncology* 2013;31:2428-36.
39. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *Jama* 2017;318:1377-84.
40. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* 2007;23:3251-3.
41. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* 1996;15:361-87.
42. Van Calster B, Wynants L, Verbeek JF, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *European urology* 2018;74:796-804.