The Development and Validation of a Novel 71-Gene Signature for Risk Stratification and Prognosis in Lower Grade Glioma

Jianxiong He

Yichun People's Hospital

Pinggen Li

Yichun People's Hospital

Xianggan Wang

Yichun People's Hospital

Feijun Chen

Yichun People's Hospital

Weijun Wu

Yichun People's Hospital

Nanfei Wang

Yichun People's Hospital

Zhengyang Li

Yichun People's Hospital

Yuan Hu

Yichun People's Hospital

Xiaojun Zhang (✉ zxj873140676@sohu.com)

Yichun People's Hospital

Research Article

Keywords: Lower grade glioma, 71-gene score, nomogram, overall survival, relapse-free survival

Posted Date: August 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-789624/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Lower-grade gliomas (LGG) are a diverse group of primary brain tumors with relatively poor overall survival in young adults. In this study, we aimed to establish novel method that are effectively predictive of prognosis of LGG patients.

Methods: We detected and validated prognosis-associated genes using gene expression and c`Clinical data of LGG patients from The Cancer Genome Atlas (TCGA) and Chinese Glioma Genome Atlas (CGGA) databases. We then established a novel prognostic 71-gene score and 17-gene nomograms and analyzed their relationship with overall survival (OS) and relapse-free survival (RFS) in LGG patients. We also performed Gene Set Enrichment Analysis to investigate the altered signalling pathways associated with the 71-gene score phenotype and hierarchical clustering analysis of 71 genes to detect subgroups of LGG patients with distinct clinical characteristics.

Results: We identified 1489 genes significantly correlated with patients' prognosis in LGG. The 71-gene score was predictive of favourable OS and RFS in LGG patients independently of clinicopathological characteristics. The *wnt* signalling pathway, glutathione metabolism, primary immunodeficiency, galactose metabolism were the potential pathways involved in the prognostication of the 71-gene score. Hierarchical clustering analysis of the 71 genes revealed three subgroups of LGG patients in the TCGA dataset. The cluster2 LGG tumours were associated with higher grade, more frequent radiation therapy, poorer OS and RFS than cluster1 and cluster3 tumours. The 71-gene nomogram incorporating the survival-related clinical factors showed good prediction accuracies for overall survival, 3-year and 5-year survival (area under curve [AUC] = 0.79, 0.67 and 0.75 respectively).

Conclusions: The 71-gene nomogram may turn out to be a useful and robust method to remarkably ameliorate the prognostic prediction in LGG.

Introduction

Gliomas are malignant tumours that start in glial cells and account for 30% of all brain tumors and central nervous system. The incidence of gliomas is approximately 6.03 per 100,000 in USA[1], 1–4/100,000 in China[2]. Gliomas comprise astrocytoma, brain stem glioma, ependymoma, mixed glioma, oligodendroglioma subtypes, with astrocytoma the most common histological subtype [1]. Gliomas are classified into four grades according to histopathologic appearance, clinical criteria and molecular parameters [3]. Lower-grade gliomas [LGG] refer to World Health Organization grades II and III gliomas [1, 3]. Despite a favourable survival for LGG, a large fraction of LGG patients undergo high-grade transformation within 10 years, indicating the importance of prognosis prediction for LGG patients [4].

The treatments for glioma patients include surgical resection, radiotherapy or temozolamide treatment [5]. The average overall survival is around 7 years, only 20% of LGG patients could survive for at least two decades[5]. Several studies have established prognostic genetic markers for LGG patients. For example, patients with 1p-19q co-deletion are related to prolonged chemotherapy response and improved survival in oligodendrogliomas[6]. The *isocitrate dehydrogenase 1* (*IDH1*) mutations are indicative of more favorable prognosis and higher response rate to temozolamide in LGG[7]. The patients with O(6)-methylguanine-DNA methyltransferase promoter (MGMTp) methylation are more likely to benefit from concomitant chemoradiotherapy in glioblastomas[8]. Apart from genetic biomarkers, recent studies uncovered gene expression profiles which successfully classify patients into subgroups of glioma patients with distinct survival probabilities [9, 10]. Therefore, the identification of prognostic molecular biomarkers is critical to the improvement of survival of LGG patients.

In the present study, we aimed to establish a robust and accurate risk score model to effectively predict the overall survival (OS) and relapse-free survival (RFS) in LGG patients. To this aim, we first determined survival-related genes using

the expression and clinical data of LGG patients from The Cancer Genome Atlas (TCGA) database[11]. Then, we established the 71-gene score with the least absolute shrinkage and selection operator (LASSO) model for prognosis prediction and investigated the signalling pathways related to the 71-gene score phenotype. Lastly, the relationship between survival-related genes, 71-gene score and prognosis was validated in an external LGG patient dataset from the Chinese Glioma Genome Atlas (CGGA) database[12].

Methods And Materials

Acquisition of gene expression and clinical data

The RNA-seq data and clinical information of LGG samples were collected from the TCGA database (n = 506). The external validation dataset was obtained from the CGGA database (n = 420). Given all the data were obtained from public databases, the study didn't need to be approved by the ethics committee of Yichun People's Hospital.

Identification of the survival-related genes

We utilized different statistical methods to examine the associations between OS, RFS and clinical factors in LGG patients, with student's t test for quantitative variables and fisher exact test for categorical variables. The prognosis-associated genes were determined with various statistical methods. In brief, the difference in the survival rates was compared by the Kaplan-Meier survival analysis between the high-expression and low-expression groups which were stratified by the median expression value using the survival package [13, 14]. Multivariate Cox regression model was performed to confirm whether gene expression was independently prognostic following the adjustment of clinical factors. Survival-related genes were further grouped into risk genes with hazard ratio (HR) greater than 1 and protective genes with HR ranging from 0 and 1.

The Establishment And Validation Of The 71-gene Score

As the gene expression units are different between the TCGA and CGGA cohorts, gene expression was first normalized following the z score formula $z = (x - \bar{x})/s$, where x is the gene expression value, $\bar{x}$ is the mean, and s is the standard deviation. We performed 10-fold cross-validation of the LASSO model to determine the optimal combination of genes for the prediction of OS in the TCGA dataset using the R package glmnet [15]. The 71-gene score = expression of gene 1 $\times \beta_1$ + expression of gene 2 $\times \beta_2$ + ... + expression of gene n $\times \beta_n$. β values are the regression coefficients derived from the LASSO model. We utilized the same method as described in the identification of the survival-related genes to study the correlations between OS, RFS and the 71-gene score. The correlations between clinical characteristics and the 71-gene score were analyzed by linear regression model in the TCGA cohort (n = 506) and the CGGA cohort (n = 420). P value < 0.05 was considered statistically significant.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) [16] was implemented to analyze the altered signalling pathways with the default parameters between the high 71-gene score and low 71-gene score groups split by the median 71-gene score. P value < 0.05 was considered statistically significant.

Unsupervised Hierarchical Clustering Analysis

Unsupervised hierarchical clustering of 71 genes was conducted in the TCGA and CGGA cohorts using the R package pheatmap[17]. We utilized different statistical methods to compare the differences in clinical factors between subgroups of patients, with student t test for quantitative clinical factors and fisher exact test for categorical variables. Kaplan-Meier curves were plotted using the R package survival[13], and survival rates were compared among the three subgroups of LGG patients using the log-rank test. P value < 0.05 was predefined as statistically significant.

Construction And Validation Of The 71-gene Nomogram

Nomogram was constructed using the rms package in R, and included patient's age, histologic grade, *IDH1* mutation status, and 71-gene signature as they are significantly correlated with OS in LGG. The performance of the nomogram was evaluated in the TCGA cohort and validated in the CGGA cohort using the R package pROC. Area under curve (AUC) values were computed accordingly for the nomogram in the prediction of overall survival, relapse-free survival, three-year and five-year survival.

Results

Analysis of the prognostic clinical factors

Patient's age, histological type, histologic grade, tumor weight, *IDH1* mutation and targeted therapy were significantly related to OS in the TCGA cohort ($P < 0.05$ for all cases, student's t test or fisher exact test, Table 1). Additionally, histologic grade, *IDH1* mutation, MGMTp methylation and codeletion of 1p-19q exhibited significant correlations with OS in the CGGA cohort ($P < 0.05$, fisher exact test, supplementary Table 1). Regarding the associations of clinicopathologic factors with RFS, *IDH1* mutation was predictive of improved RFS in the TCGA cohort ($P < 0.05$, Fisher exact test, Table 1). Meanwhile, histologic grade and chemotherapy were related to RFS in the CGGA dataset ($P > 0.05$ for all cases, fisher exact test or student's t test, supplementary Table 1).

Table 1

Association between the clinicopathologic characteristics and patients' OS as well as RFS in the TCGA dataset

Variables	Group	Alive	Dead	P value	Non-relapse	Relapse	P value	Statistical method
Age		41.68	49.06	≤ 0.001	42.09	43.62	0.29	Student t test
Tumour weight		338.38	285.01	0.02	331.86	328.05	0.89	Student t test
Gender	Female	179	45	0.24	123	60	0.83	Fisher's exact test
	Male	237	45		150	69		
History of cancer	No	183	27	0.34	122	49	0.3	Fisher's exact test
	Yes	107	22		76	40		
Histological type	Astrocytoma	151	41	0.17	94	54	0.26	Fisher's exact test
	Oligoastrocytoma	112	17		77	28		
	Oligodendroglioma	153	32		102	47		
Histologic grade	G2	217	26	≤ 0.001	140	57	0.20	Fisher's exact test
	G3	198	64		132	72		
<i>IDH1</i> mutation	Wild-type	79	34	≤ 0.001	44	41	≤ 0.001	Fisher's exact test
	Mutant	337	56		229	88		
<i>TP53</i> mutation	Wild-type	211	47	0.82	139	57	0.24	Fisher's exact test
	Mutant	205	43		134	72		
Radiation therapy	No	104	14	0.29	65	28	0.88	Fisher's exact test
	Yes	116	24		81	33		
Targeted therapy	No	167	22	0.03	124	46	0.08	Fisher's exact test
	Yes	199	48		145	80		

Analysis Of The Prognosis-related Genes

We firstly utilized Kaplan-Meier survival analysis to screen for prognosis-related genes and identified 3494 genes and 4181 genes that were significantly correlated with favourable or poor OS respectively ($P < 0.05$ for all cases, log rank test, Fig. 1). Then the prognosis-related genes were analyzed to evaluate whether their association with OS remained statistically significant after the adjustment of survival-related clinical features using multivariate Cox regression analysis. The results confirmed 1561 protective genes and 1812 risk genes ($P < 0.05$ for all cases, supplementary Table 2, Fig. 1). The associations between 3373 gene expression levels and OS were further validated in 420 LGG samples of the CGGA dataset. The results confirmed that high expression levels of 421 genes were related to prolonged OS, while high expression levels of 1068 genes were related to shorter OS ($P < 0.05$ for all cases, Fig. 1, supplementary Table 3).

The 71-gene Score Is Indicative Of Favourable Overall Survival

We performed 10-fold cross-validation test of the LASSO model and uncovered the LASSO model comprising 71 genes showed the best performance for prognosis prediction, as evidenced by the maximum area under curve and the log (lambda) equal to -4.2 (supplementary Table 2, 3, 4). The 71-gene scores for deceased LGG patients significantly differed from living LGG patients in the TCGA and CGGA cohorts ($P < 0.05$ for all cases, student t test, Fig. 2B). The Kaplan-Meier survival analysis showed LGG patients with higher 71-gene scores were associated with higher survival probabilities than those with lower 71-gene scores in the TCGA dataset ($P < 0.05$ for all cases, Fig. 2C). The multivariate Cox regression analysis further verified that the 71-gene score was a prognosticator for predicting favourable OS independently of prognosis-related clinical features in LGG. Furthermore, the positive association between the 71-gene score and OS was also confirmed in the CGGA dataset ($P < 0.05$ for all cases, Table 2 and Fig. 2D).

Table 2
Multivariate analyses between OS and the 71-gene score in the TCGA and CGGA datasets

TCGA dataset				CGGA dataset			
Clinical feature	HR	95%CI	P value	Clinical feature	HR	95%CI	P value
Age	1.06	1.04–1.09	< 0.001	Histologic grade	2.19	1.46–3.28	< 0.001
Histologic grade	3.79	1.98–7.24	< 0.001	IDH1	0.7	0.47–1.04	0.08
Tumor weight	1	0.99-1	0.81	1p-19q codeletion	0.65	0.4–1.05	0.08
IDH1	0.57	0.33-1	0.04	The 71-gene score	0.36	0.31–0.41	< 0.001
Targeted therapy	0.43	0.23–0.8	0.01				
The 71-gene score	0.98	0.98–0.99	< 0.001				
Notably, HR and CI refer to hazard ratio and confidence interval respectively.							

The 71-gene score is a positive predictor for relapse-free survival

Long term RFS is an important surrogate for OS in the evaluation of prognosis, we followed the OS analysis methods to study the associations between the 71-gene score and RFS. As compared to the non-recurrent tumors, recurrent tumors exhibited significantly lower 71-gene scores in the TCGA and CGGA cohorts ($P < 0.05$ for all cases, student t test, supplementary Fig. 1A). Kaplan-Meier RFS analysis suggested that high 71-gene score was related to a favourable RFS ($P < 0.05$, log rank test, supplementary Fig. 1B). Multivariate Cox regression analysis exhibited that increased 71-gene score was related to decreased RFS after the adjustment of *IDH1* mutation ($P < 0.05$, supplementary Table 5). We analysed the associations between RFS and the 71-gene score in the CGGA dataset to validate the findings above. The positive correlation between the 71-gene score and RFS was confirmed by the Kaplan-Meier analysis and multivariate analysis ($P < 0.05$, supplementary Fig. 1C, supplementary Table 5).

The 71-gene score performs well for the risk stratification independently of clinical factors

We next would like to characterize the relationship between the 71-gene score and each clinical factor and performed linear regression model analysis. We identified *IDH1* mutation, tumor weight significantly positively correlated with the 71-gene score and age and histologic grade negatively correlated with the 71-gene score in the TCGA cohort ($p < 0.05$ for all cases, Fig. 3A). Similar correlations were also observed in the CGGA cohort ($p < 0.05$ for all cases, Fig. 3B). Then, we divided LGG patients into two subgroups following each clinical feature and analyzed whether the positive association

between the 71-gene score and OS was independent of clinical characteristics. The high 71-gene score was significantly correlated with favourable OS independently of gender, age, tumour weight, histologic grade, *TP53* mutation, *IDH1* mutation, history of cancer, radiotherapy and targeted therapy in the TCGA cohort ($P < 0.05$ for all cases, log rank test, supplementary Fig. 2–4). The CGGA cohort exhibited a similar pattern ($P < 0.05$ for all cases, log rank test, supplementary Fig. 5–7). Moreover, we also demonstrated the high 71-gene score was indicative of favourable RFS regardless of clinical characteristics ($P < 0.05$ for all cases, log rank test, supplementary Table 6).

Identification of the signalling pathways associated with the 71-gene score

GSEA was used to analyze the altered signalling pathways between the high and low 71-gene score groups stratified by the median 71-gene score. The wnt signalling pathway was significantly up-regulated in the high 71-gene score group of the TCGA cohort ($p < 0.05$, Fig. 4A). On the contrary, genes in the pathways of glutathione metabolism, primary immunodeficiency, galactose metabolism, glycosaminoglycan degradation, leukocyte transendothelial migration, peroxisome were significantly up-regulated in the low 71-gene score group of the TCGA cohort ($p < 0.05$ for all cases, Fig. 4B–D). What's more, in the low 71 gene score group of CGGA cohort, 20 KEGG pathways were significantly up-regulated, the pathogenic *Escherichia coli* infection, cell cycle, focal adhesion, small cell lung cancer and pancreatic cancer were the top five pathways (all cases, $P < 0.05$, supplementary table 7). These results suggest that the 71 gene score can accurately predict the survival rate of glioma patients, which may be related to the above signalling pathways.

Classification Of Lgg Patients

We implemented the hierarchical clustering analysis of the 71 genes and identified three subgroups of LGG patients in the TCGA dataset (Fig. 5A). The cluster2 LGG tumours were characterized by higher grade, shorter OS and RFS, more frequent radiation therapy in comparison with cluster1 and cluster3 tumours (P values < 0.05 for all cases, fisher exact test or log-rank test, Fig. 5B, supplementary Fig. 8A and supplementary Table 8). Similarly, three classes of LGG patients were uncovered in the CGGA cohort (Fig. 5C). Cluster2 tumours presented significantly lower 71-gene score, lower frequency of 1p-19q codeletion, poorer OS and RFS than cluster1 and cluster3 tumours as well as higher grade than cluster3 tumours (P values < 0.05 for all cases, student t test, fisher exact test or log-rank test, Fig. 5D, supplementary Fig. 8B and supplementary Table 9). The above results indicate that hierarchical clustering analysis of the 71 gene panel successfully divides LGG patients into different survival groups, which were independently of clinical characteristics.

Nomogram combined 71-gene signature and clinical-related variables predicts patients' prognosis

In the TCGA and CGGA cohorts, patient's age, histologic grade, *IDH1* mutation status and 71-gene signature were significantly associated with OS. Then based on the above analysis results, we established a 71-gene nomogram that incorporated the survival-related clinical factors and 71-gene signature (Fig. 5A). The nomogram predicted well the 3-year and 5-year survival for LGG patients in the TCGA cohort, ROC plot revealed the 71-gene nomogram showed good prediction accuracies for overall survival, 3-year and 5-year survival (AUC: 0.66, 0.8 and 0.75 respectively, Fig. 6B and C). The results were also validated in the CGGA cohort (AUC: 0.79, 0.67 and 0.75 respectively, Fig. 6B and D), demonstrating the clinical value and validity of the 71-gene nomogram for OS evaluation of LGG patients. Furthermore, we also established a 71-gene nomogram for predicting RFS and demonstrated the nomogram performed well for predicting RFS, 3-year and 5-year RFS (AUC: 0.72, 0.6 and 0.62 respectively, supplementary Fig. 9).

Discussion

In the past decade, although significant progress has been made in the treatment of LGG, the global mortality rate of LGG is still relatively high. Several studies have confirmed the associations of molecular tumor markers and gene-signatures

with OS in LGG patients [6, 18–21]. In the present study, we established a 71-gene score model using a linear combination of 71 gene expression levels weighted by β -values of the LASSO model. After adjusting for known clinical factors, the 71 gene score was positively correlated with OS and RFS. As compared to established prognosticators, such as histologic grade, 1p-19q co-deletion and *IDH1* mutation, the 71-gene expression signature is advantageous to existing prognosticators. For example, our LASSO model could accurately predict OS and RFS regardless of clinicopathological characteristics. Furthermore, we established a 16-gene nomogram that incorporated the survival-related clinical factors and 16-gene signature. Therefore, the 71-gene nomogram might provide a better risk classification for those patients who are wildtype for 1p-19q co-deletion and *IDH1* mutation or within the same histological grade.

In addition, the clinical significance of 71-gene score was evaluated and KEGG pathways were identified. The wnt signalling pathway was significantly up-regulated in the high 71-gene group of the TCGA cohort. Genes in the pathways of glutathione metabolism, primary immunodeficiency, galactose metabolism, glycosaminoglycan degradation, leukocyte transendothelial migration, peroxisome were significantly up-regulated in the low 71-gene group of the TCGA cohort. The wnt signalling pathway plays critical roles in the central nervous system development. Aberrantly activated wnt signalling pathway may lead to malignant transformation and development of brain tumors in nervous system [22, 23]. We believe that the wnt signaling pathway and the above signaling pathways are involved in the molecular mechanism of 71 gene score and survival of LGG patients.

The 71 genes may have different functions in the tumorigenesis of cancers. For example, cadherin-like protein 22 (*CDH22*) is a transmembrane glycoprotein involved in cell-cell adhesion and metastasis and related to the morphogenesis and tissue formation of brain and neuroendocrine organs [24–26]. *CDH22* protein level is down-regulated in breast cancer tissues and melanoma samples. Consistent with our findings, dysregulated *CDH22* expression has been reported as an independent predictor of poor survival in breast cancer [27] and melanoma [24]. The eukaryotic initiation factor 3f (eIF3f) is the p47 subunit of the multi-subunit eIF3 complex, which plays a crucial role in translation initiation [28]. eIF3f expression is significantly decreased in many human cancers [29, 30]. Increased expression of eIF3f is a significant factor for a favourable prognosis in gastric cancer [29]. The enhanced expression of eIF3f significantly repressed cancer cell growth and tumorigenic ability in nude mice [31]. Aquaporin 7 (*AQP7*) is a water/glycerol transporting protein that regulates adipocyte glycerol efflux and influences lipid and glucose homeostasis. Low expression of *AQP7* is correlated with tumor grade and the aggressive features in hepatocellular carcinoma [32]. *AQP7* is a prognostic indicator of OS in breast cancer patients. Reduced expression of *AQP7* caused reduced primary tumor burden and lung metastasis in mouse breast cancer models [33]. Finally, these genes may also pave the way for developing targeted therapies for LGG patients. For example, silencing the expression of eIF3f or *AQP7* led to marked decrease in cell proliferation, metastasis in various cancer types [31, 33].

Conclusion

In conclusion, the 71 genes are identified as new prognostic biomarkers. Regardless of the clinicopathological features, the high 71 gene score indicates a good prognosis. The 71-gene nomogram may turn out to be a useful and robust method to remarkably ameliorate the prognostic prediction in LGG.

Declarations

Acknowledgement

None

Authors' contributions

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Xiaojun Zhang designed the study, Jianxiong He and Pinggen Li downloaded expression and clinical data from the TCGA and CGGA database. Jianxiong He, Pinggen Li, Xianggan Wang, Feijun Chen performed the survival analysis and hierarchical clustering analysis. Weijun Wu, Nanfei Wang, Zhengyang Li, Yuan Hu developed the 71-gene risk score and validated the results. Jianxiong He wrote the manuscript, all authors have read and approved the final manuscript.

Funding

None

Ethics approval and consent to participate

None

Consent for publication

None

Competing interests

The authors declare no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are available upon reasonable request.

References

1. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, et al. N E U R O - O N C O L O G Y CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2006–2010. *Neuro Oncol* [Internet]. 2013;12:28–36. Available from: <http://neuro-oncology.oxfordjournals.org/content/12/5/500.abstract>.
2. Chang L, Su J, Jia X, Ren H. Treating malignant glioma in Chinese patients: Update on temozolomide. *Onco Targets Ther*. 2014;7:235–44.
3. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol*. 2007;114:97–109.
4. Kiran M, Chatrath A, Tang X, Keenan DM, Dutta A. A Prognostic Signature for Lower Grade Gliomas Based on Expression of Long Non-Coding RNAs. *Mol Neurobiol*. 2019;56:4786–98.
5. Claus EB, Walsh KM, Wiencke JK, Molinaro AM, Wiemels JL, Schildkraut JM, et al. Survival and low-grade glioma: the emergence of genetic information. *Neurosurg Focus*. 2015;38:1–10.
6. Gadji M, Fortin D, Tsanaclis AM, Drouin R. Is the 1p/19q deletion a diagnostic marker of oligodendrogliomas? *Cancer Genet Cytogenet* [Internet]. Elsevier Inc.; 2009;194:12–22. Available from: <http://dx.doi.org/10.1016/j.cancergencyto.2009.05.004>.
7. Houillier C, Wang X, Kaloshi G, Mokhtari K, Guillemin R, Laffaire J, et al. IDH1 or IDH2 mutations predict longer survival and response to temozolomide in low-grade gliomas. *Neurology*. 2010;75:1560–6.
8. Ducray F, Idhah A, Wang X-W, Cheneau C, Labussiere M, Sanson M. Predictive and prognostic factors for gliomas. *Expert Rev Anticancer Ther England*. 2011;11:781–9.
9. Hsu JBK, Chang TH, Lee GA, Lee TY, Chen CY. Identification of potential biomarkers related to glioma survival by gene expression profile analysis. *BMC Med Genomics*. 2019;11:1–18.

10. Wang Y, Liu X, Guan G, Xiao Z, Zhao W, Zhuang M. Identification of a Five-Pseudogene Signature for Predicting Survival and Its ceRNA Network in Glioma. *Front Oncol.* 2019;9.
11. Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* [Internet]. 2013;45:1113–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24071849>.
12. Wang Y, Qian T, You G, Peng X, Chen C, You Y, et al. Localizing seizure-susceptible brain regions associated with low-grade gliomas using voxel-based lesion-symptom mapping. *Neuro Oncol.* 2015;17:282–8.
13. Therneau T. Survival Analysis. *Cran* [Internet]. 2016; Available from: <https://cran.r-project.org/web/packages/survival/survival.pdf>.
14. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.2264&rep=rep1&type=pdf>.
15. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw.* 2011;39:1–13.
16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* [Internet]. 2005;102:15545–50. Available from: <http://www.pnas.org/content/102/43/15545.abstract>.
17. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R programming tools for plotting data. *R Packag. version.* 2005.
18. Cohen AL, Holmen SL, Colman H. IDH1 and IDH2 mutations in gliomas. *Curr Neurol Neurosci Rep.* 2013;13:345.
19. Chen WL, Wang YY, Zhao A, Xia L, Xie G, Su M, et al. Enhanced Fructose Utilization Mediated by SLC2A5 Is a Unique Metabolic Feature of Acute Myeloid Leukemia with Therapeutic Potential. *Cancer Cell* [Internet]. Elsevier Inc.; 2016;30:779–91. Available from: <http://dx.doi.org/10.1016/j.ccell.2016.09.006>.
20. Zhang YA, Zhou Y, Luo X, Song K, Ma X, Sathe A, et al. SHOX2 is a Potent Independent Biomarker to Predict Survival of WHO Grade II–III Diffuse Gliomas. *EBioMedicine* [Internet]. The Authors; 2016;13:80–9. Available from: <http://dx.doi.org/10.1016/j.ebiom.2016.10.040>.
21. Chen B, Liang T, Yang P, Wang H, Liu Y, Yang F, et al. Classifying lower grade glioma cases according to whole genome gene expression. *Oncotarget.* 2016;7:74031–42.
22. Kalani MYS, Cheshier SH, Cord BJ, Bababeygy SR, Vogel H, Weissman IL, et al. Wnt-mediated self-renewal of neural stem/progenitor cells. *Proc Natl Acad Sci U S A.* 2008;105:16970–5.
23. Pulvirenti T, Van Der Heijden M, Droms LA, Huse JT, Tabar V, Hall A. Dishevelled 2 signaling promotes self-renewal and tumorigenicity in human gliomas. *Cancer Res.* 2011;71:7280–90.
24. Piche B, Khosravi S, Martinka M, Ho V, Li G. CDH22 expression is reduced in metastatic melanoma. *Am J Cancer Res.* 2011;1:233–9.
25. Saarimäki-Vire J, Alitalo A, Partanen J. Analysis of Cdh22 expression and function in the developing mouse brain. *Dev Dyn an Off Publ Am Assoc Anat United States.* 2011;240:1989–2001.
26. Zhou J, Li J, Chen J, Liu Y, Gao W, Ding Y. Over-expression of CDH22 is associated with tumor progression in colorectal cancer. *Tumour Biol J Int Soc Oncodevelopmental Biol Med Netherlands.* 2009;30:130–40.
27. Martín-Sánchez E, Mendaza S, Ulazia-Garmendia A, Monreal-Santesteban I, Córdoba A, Vicente-García F, et al. CDH22 hypermethylation is an independent prognostic biomarker in breast cancer. *Clin Epigenetics* [Internet]. Clinical Epigenetics; 2017;9:1–8. Available from: <http://dx.doi.org/10.1186/s13148-016-0309-z>.
28. Shi J, Kahle A, Hershey JWB, Honchak BM, Warneke JA, Leong SPL, et al. Decreased expression of eukaryotic initiation factor 3f deregulates translation and apoptosis in tumor cells. *Oncogene England.* 2006;25:4923–36.
29. Li G, Wang N, Sun C, Li B. Decreased expression of eukaryotic initiation factor 3f is an adverse prognostic factor for

30. Doldan A, Chandramouli A, Shanas R, Bhattacharyya A, Leong SPL, Nelson MA, et al. Loss of the eukaryotic initiation factor 3f in melanoma. *Mol Carcinog*. 2008;47:806–13.

31. Lee J-Y, Kim H-J, Rho SB, Lee S-H. eIF3f reduces tumor growth by directly interrupting clusterin with anti-apoptotic property in cancer cells. *Oncotarget*. 2016;7:18541–57.

32. Chen X-F, Li C-F, Lü L, Mei Z-C. Expression and clinical significance of aquaglyceroporins in human hepatocellular carcinoma. *Mol Med Rep Greece*. 2016;13:5283–9.

33. Dai C, Charlestin V, Wang M, Walker ZT, Miranda-Vergara MC, Facchine BA, et al. Aquaporin-7 Regulates the Response to Cellular Stress in Breast Cancer. *Cancer Res United States*. 2020;80:4071–86.

Figures

Figure 1
 Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

The intersection of survival-related genes between the TCGA and CGGA datasets. A. The protective genes common to the TCGA and CGGA cohorts, B. The risk genes common to the TCGA and CGGA cohorts. Red and blue refer to the genes which were identified significantly related to OS by the Kaplan-Meier analysis in the CGGA and TCGA cohorts respectively. Green and purple represent the genes which were confirmed to be significantly related to OS by the Multivariate analysis in the CGGA and TCGA cohorts respectively.

Figure 2

The 71-gene score is a positive prognosticator for OS in LGG. A. The relationship between log scaled lambda values and AUC and number of genes with coefficients above zero in the LASSO model. The numbers on the top indicate number of genes with coefficients above zero in the LASSO model. The left and right vertical dotted lines indicate the λ_{min} and $\lambda_{1\text{se}}$ for λ respectively. The λ_{min} is the one which minimizes out-of-sample loss in cross validation. The $\lambda_{1\text{se}}$ is the one which is the largest lambda value within 1 standard error of the minimum. B. The living and the 71-gene scores in the two cohorts. C. The high and low 71-

gene score groups present significant difference in overall survival as demonstrated by the Kaplan-Meier survival analysis in the TCGA cohort. D. The high and low 71-gene score groups present significant difference in overall survival as demonstrated by the Kaplan-Meier survival analysis in the CGGA cohort.

Figure 3

Correlations between the 71-gene score and clinical features demonstrated by the linear regression model analysis in the TCGA cohort (A) and the CGGA cohort (B). Of note, ** and *** represent P value < 0.01 and 0.001 respectively.

Figure 4

KEGG pathways that contribute to the prognostic value of the 71-gene score. GSEA based on the expression of TCGA dataset revealed significant pathways related to the 71-gene score, including wnt signalling pathway (A) enriched in the high 71-gene score group (n=253), glutathione metabolism (B), primary immunodeficiency (C), galactose metabolism (D) enriched in the low 71-gene score group (n=253).

Figure 5

Unsupervised hierarchical clustering of the 71 gene panel uncovered three classes of LGG patients. A. Unsupervised hierarchical clustering of the 23 gene panel uncovered three classes of LGG patients in the TCGA dataset. B. The difference in overall survival between the three subsets of LGG patients in the TCGA cohort. C. Unsupervised hierarchical clustering of the 71 gene panel uncovered three classes of LGG patients in the CGGA dataset. D. The difference in overall survival between the three classes of LGG patients in the CGGA cohort.

Figure 6

The 71-gene nomogram to predict the risk of disease in patients with LGG. (A) mRNA nomogram to predict OS. 2,3 for histologic grade denote grade II and III respectively. High and low for the 71-gene score represent high and low 71-gene scores respectively stratified by the median 17-gene score. B. The ROC plot for the nomogram in predicting overall survival in the TCGA and CGGA cohorts. C. The ROC plot for the nomogram in predicting 3-year and 5-year overall survival in the TCGA cohort. D. The ROC plot for the nomogram in predicting 3-year and 5-year overall survival in the CGGA cohort.

Supplementary Files

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

- [Supplementaryfigures.docx](#)
- [Supplementarytables.docx](#)