

Identification and Validation of a Five-lncRNA Signature Related to Glioma Using Bioinformatics Analysis

Daofeng Tian (✉ 307367133@qq.com)

Wuhan University Renmin Hospital

Haitao Liu

jia xing xue yuan: Jiaxing University

Pengfei Xu

The Seventh Affiliated Hospital Sun Yat-sen University

Liguo Ye

Wuhan University Renmin Hospital

Long Wang

Wuhan University Renmin Hospital

Yinqiu Tan

Wuhan University Renmin Hospital

Baohui Liu

Wuhan University Renmin Hospital

Qianxue Chen

Wuhan University Renmin Hospital

Research article

Keywords: WGCNA, TCGA, CGGA, glioma, prognostic signature

Posted Date: September 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-70707/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on March 9th, 2021. See the published version at <https://doi.org/10.1186/s12885-021-07972-9>.

Abstract

Background

To accurately predict the prognosis of glioma patients.

Methods and Results

A total of 541 samples from the TCGA cohort and 181 observations from the CGGA database were included in our study. By weighted gene co-expression network analysis (WGCNA), 14 long non-coding RNAs (lncRNAs) associated with glioma grade were identified. Using univariate and multivariate Cox analysis Five lncRNAs (CYTOR, MIR155HG, LINC00641, AC120036.4 and PWAR6) were selected to develop the prognostic signature. The Kaplan-Meier curve depicted that the patients in high risk group had poor prognosis in both cohorts. The areas under the receiver operating characteristic curve of the signature in predicting the survival of glioma patients at 1, 3, and 5 years were 0.84, 0.92, and 0.90 in the CGGA cohort and 0.8, 0.85 and 0.77 in the TCGA set. Multivariate Cox analysis demonstrated that the five-lncRNA signature was an independent prognostic indicator in both sets (HR = 2.002, $p < 0.001$; HR = 1.243, $p = 0.007$, respectively). A nomogram including the lncRNAs signature and clinical covariates was constructed and demonstrated high predictive accuracy in predicting 1-, 3- and 5-year survival probability of glioma patients.

Conclusion

We established a five-lncRNA signature as a potentially reliable tool for survival prediction of glioma patients.

Introduction

Glioma is the most prevalent neoplasm in central nervous system (CNS), with the highest malignancy and the worst prognosis compared with other tumors in brain. In spite of the fact that clinical interventions have been improved dramatically, but the outcomes of glioma patients fail to satisfy the expected goals [1]. The main factors of the poor prognosis are due to the atypical symptoms and the aggressiveness of glioma, resulting in a large proportion of glioma patients diagnosed and treated at the advanced grade [2]. Therefore, further exploration on the molecular mechanisms of tumorigenesis and progression is crucial for improving diagnosis and treatment and investigations into the novel biomarkers are needed.

The past few years witness the great accumulation of evidence indicating the subtle and complex regulatory process in tumor biological behaviors, insights into which in all probability facilitate the development of more effective treatment strategies and improvement in clinical benefits [3]. High-throughput sequencing makes it possible to comprehensively analyze the transcriptome and genome, especially for long non-coding RNA (lncRNA), mRNA, and others [4, 5]. lncRNA typically has more than

200 nucleotides in length, disabling to encode proteins [6]. This type of RNA has been confirmed to be involved in several biological functions including transcription [7], RNA splicing [6], N6-methyladenosine (m6A) [8] and others. The expression profiles of lncRNAs are cell-specific and tumor-specific [9]. Meanwhile, there is mounting evidence that the dysregulated lncRNAs act a pivotal part in many biological processes of malignancy, demonstrating great potentiality as novel diagnostic or prognostic molecular biomarkers [10, 11]. Regarding glioma, some lncRNAs are dysregulated and implicated in cell growth, apoptosis, invasion and angiogenesis. lncRNA-MALAT1 has been considered as a tumor suppressor and down expression of MALAT1 to cause remarkable promotion of invasion and proliferation of the glioma cells [12]. lncRNA-NEAT1 has been revealed to be modulated by the epidermal growth factor receptor (EGFR) pathway, leading to glioblastoma multiform progression by the WNT/ β -Catenin Pathway by Scaffolding enhancer of zeste homolog 2 (EZH2) [13]. Nevertheless, the molecular functions and mechanisms of the great majority of lncRNAs remain ill-defined and the research concerning lncRNAs with diagnosis or prognosis potentiality in glioma is still in the initial stage.

In this study, we used the univariate and multivariate Cox regression analysis to develop a prognostic predictor based on multiple lncRNAs for glioma patients. The prognosis prediction accuracy of the signature was evaluated in the CGGA cohort and verified in the TCGA group.

Materials And Methods

Data obtaining and Processing

Two sets of data were included in our study. The both RNA-sequencing data were downloaded from the TANRIC database[14], and corresponding clinical information were separately obtained from two databases, including the Chinese Glioma Genome Atlas (CGGA, <http://www.cgga.org.cn/>) and the The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>). The CGGA dataset consisting of 89 LGG and 92 GBM tissue samples, acting as a training set, was used to establish a multi-lncRNA signature. The TCGA dataset, consisting of 405 LGG and 136 GBM observations, was treated as an independent validation set to verify the statistical significance of the signature. The clinical characteristics of glioma patients in the training and validation sets were shown in Table S1.

Weighted gene co-expression network analysis (WGCNA)

The process of WGCNA included the construction of gene expression similarity matrix, adjacency matrix, and then co-expression network [15]. The power value of the soft threshold β of the adjacency matrix was set as 6 through the R function pickSoftThreshold. And then we used the ScaleFree plot to evaluate whether the network met the criterion of the scale-free topology network. The hierarchical clustering method on the basis of average-linkage, was adopted to define modules. When clustering, we used the smallest module size of 30. The Pearson correlation coefficients between modules and clinical variables were calculated. The module with absolute value of the correlation coefficient > 0.5 and $p < 0.01$ was worth further analysis.

Survival analysis

The association between the hub lncRNAs and overall survival (OS) were first calculated using univariate Cox analysis. The lncRNAs with $p < 0.05$ were chosen to carry out multivariate Cox regression analysis and step-wise regression method was used to select variables to construct prognostic model on the basis of the Akaike information criterion (AIC). Then, risk score was obtained based on coefficients () from multivariate Cox analysis and expression data of lncRNAs (). The detailed formula was as follows: risk score = $\beta_1 \times X_1 + \beta_2 \times X_2 + \dots + \beta_n \times X_n$. In the light of the median values of the risk scores in training and validation cohorts, patients with glioma were separated into high and low risk groups, respectively. To analyze whether the survival time of the patients in two risk groups was significantly different, Kaplan-Meier (KM) curve was used and visualized by R package survival. To evaluate the predicting accuracy of the established model, receiver operating characteristic (ROC) curve was conducted.

Construction of Nomogram Survival Model

Nomogram could illustrate the relationship between different variables in a graphical way [16]. Meanwhile, the advantage of nomogram is that it can personally calculate the survival rate of specific tumor patients, so it has great value in clinical application [16]. To assess the probabilities of overall survival at 1, 3, and 5 years for glioma patients, a nomogram was constructed here, consisting of risk score, age, 1p19q status and glioma WHO grading. In addition, calibration curve was used to assess the degree to which actual results were consistent with the predicted results of nomogram. Areas under the ROC curve were used to making a comparison between the predictive reliability of the nomogram and clinical variables.

Differential expression of the hub lncRNAs

To investigate whether the key lncRNAs were expression-dysregulated between LGG and GBM samples, t-test was designed to perform differential expression analysis using CGGA and TCGA RNA-Seq expression profiles. lncRNAs with $p < 0.05$ were considered differentially expressed.

Construction of ceRNA Regulatory Network and Functional Annotations

Several lncRNAs have recently been observed to participate in the regulation of gene expression, through the absorption of miRNAs, resulting in the occurrence and development of malignant tumors [17]. Through the starBase v2.0 database [18], we first predicted the lncRNA-miRNA interactions. Three online tools, including miRTarBase[19], TargetScan [20] and miRDB [21] were used to explore the target mRNAs of the miRNAs. Through the four online tools mentioned above, we obtained miRNAs targeting candidate lncRNAs and target genes regulated by these miRNAs. Meanwhile, we obtained the differentially expressed mRNAs (DEmRNAs) and miRNAs (DEmiRNAs) between LGG and GBM from the CGGA database. Only the intersections of target miRNAs and DEmiRNAs and of target mRNAs and DEmRNAs were included in our ceRNA network. Visualization of ceRNA regulatory network was accomplished by software cytoscape (<https://cytoscape.org/>). The mRNAs in the ceRNA network were subject to

enrichment analysis by the R package clusterProfiler [22]. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment results with $p < 0.05$ were considered significant.

Results

Co-expression network Construction

Through conjoint analysis of co-expression network and clinical features, modules with biological significance can be confirmed in this step [15]. When the value of k was chosen as 6, the scale-free topology fit index reached 0.98, meeting the standard of approximate scale-free topology (Fig. S1). LncRNAs with similar expression patterns were divided into the same module by cluster dendrogram trees and eight modules were obtained here. The results of the investigation into the relationships between modules and glioma WHO grading were presented in Fig. 1A. By setting the threshold values as Pearson correlation coefficient $> |0.5|$ and $p < 0.01$ to select the significant modules, and by inquiring into mean gene significance across all genes in one module, the brown, green and yellow module were considered to be closely related to the WHO grade of glioma (Fig. 1B). Meanwhile, the module membership (MM) vs. gene significance (GS) analysis of the brown, green and yellow modules showed that the three were endowed with higher correlation between MM and GS (Fig. 1C-E). Within these three modules, we obtained 14 hub lncRNAs in total, by selecting lncRNAs with $|GS| > 0.85$ and $|MM| > 0.65$.

Survival analysis of risk score and clinical features

Survival analysis had been firstly performed on the training set of 181 glioma patients. From the univariate Cox analysis, all the 14 hub lncRNAs detected in the step of WGCNA were considered to be statistically connected with the clinical outcomes of gliomas (Fig. 2). Subsequently, the risk score system based on the multivariate Cox analysis was constructed, and the following formula was adopted here: $\text{risk score} = \text{CYTOR} * 0.3447 + \text{MIR155HG} * -0.8509 + \text{LINC00641} * -0.7135 + \text{AC120036.4} * -0.5505 + \text{PWAR6} * -0.5748$. The median risk score of samples was calculated in the training set, and set as a cut-off in risk stratification for glioma patients (Fig. 3A, 3C). The same analysis was conducted for the validation dataset in order to validate the training set results (Fig. 3B, 3D). And the expression profiles of the five lncRNAs were visualized by heat map in the training (Fig. 3E) and validation (Fig. 3F) datasets. Moreover, we used the K-M curve evaluated whether there were significant survival time differences for glioma patients after risk stratification. The results showed the mortality of patients in the high-risk group was significantly higher than that in the low-risk group (Fig. 4A, 4B). The values of AUC of the signature at 1-, 3-, and 5-year OS were 0.84, 0.92 and 0.9 in the training set and 0.8, 0.85 and 0.77 in the validation cohort, demonstrating the great reliability of the prognosis signature (Fig. 4C, 4D). In the training set, univariate Cox regression suggested that risk score, age, 1p19q status and WHO grading had prognostic values ($p < 0.05$), while gender not. Then, in multivariate Cox regression analysis of risk score and clinicopathological variables, risk score was still an independent and powerful prognosis-predicting factor (HR = 2.002, 95%CI [1.584-2.530], $p < 0.001$). The similar results were obtained in the validation set (HR = 1.243, 95%CI [1.063-1.469], $p = 0.007$) (Table 1).

Nomogram construction and Accuracy Assessment

To facilitate the clinical prognosis assessment for glioma patients, we established a nomogram to perform the prediction of the overall survival probability at 1-, 3-, and 5-year in the training cohort (Fig. 5A). The observed and predicted probabilities for the specific observations decrease along the diagonal line in calibration plot (Fig. 5B-D). To measure the predictive accuracy of the merged nomogram and the clinical risk factors at 1-, 3- and 5-year OS, the values of the area under the ROC curve were calculated. The 1-year AUC was 0.87 for nomogram, and 0.37 for age, 0.80 for glioma WHO grading, 0.38 for 1p19q status (Fig. 5E). In assessing the predicting efficacy at 3-year OS, AUC was 0.94 for nomograph, and 0.38 for age, 0.89 for glioma WHO grading, 0.37 for 1p19q status (Fig. 5F). Furthermore, the 5-year AUC was 0.93 for alignment chart, and 0.38 for age, 0.86 for glioma WHO grading, 0.33 for 1p19q status (Fig. 5G).

Differential expression analysis of hub lncRNAs

We confirmed that the 5 hub lncRNAs were expression-dysregulated between LGG and GBM tissues by t-test in the training and validation sets. In comparison with expression levels of lncRNAs in low grade glioma, lnc-CYTOR and MIR155HG were significantly over-expressed in GBM tissues, while three lncRNAs (LINC00641, AC120036.4, PWAR6) were remarkably down-regulated in GBM samples (Fig. S2).

Construction of ceRNA Regulation Network and Enrichment Analysis

A competing endogenous RNA (ceRNA) regulation network was constructed and visualized by Cytoscape software (Fig. 6A). And in the network, there were 238 nodes and 512 edges. Moreover, we explored the possible biological mechanisms of the hub lncRNAs related to glioma. By applying R package clusterProfiler, mRNAs in the network were subject to function annotation analysis. It indicated that the mRNAs regulated by the hub lncRNAs were mainly involved in regulation of cell cycle process, regulation of cellular senescence and in cell-matrix adhesion. Moreover, pathways analysis of target mRNAs showed a statistically significant association with glioma and p53 signaling pathway, confirming potential roles in cancer development of the hub lncRNAs (Fig. 6B, 6C).

Discussion

Recent research demonstrates multiple biomarkers integrated into a single model can greatly improve the prognostic value of a single model [23]. Glioma patients are rather heterogeneous. Even in the same pathological grade, there are considerable differences in their clinical outcomes [1]. The present study was designed to detect and validate a signature containing multiple markers to improve the prognostic accuracy for gliomas.

In particular, we applied WGCNA to glioma patients with RNA-seq lncRNA expression data, which has been extensively utilized in exploring biomarkers in cancers such as pancreatic carcinoma and breast cancer[24, 25]. In our research, through the construction of a lncRNA co-expression network with data available at the TANRIC database, three modules-green, yellow and brown associated with glioma WHO

grading were identified using the WGCNA algorithm. Total 14 hub lncRNAs with $|GS| > 0.85$ and $|MM| > 0.65$ were selected from the three modules. Then, univariate Cox analysis of the hub lncRNAs and overall survival was firstly conducted and demonstrated all the variables having outstanding statistical significance. In multivariate Cox analysis, stepwise regression was used to select markers and a risk model was constructed to predict glioma prognosis. Finally, five lncRNAs (CYTOR, MIR155HG, LINC00641, AC120036.4 and PWAR6) were chosen to establish a risk score system for prediction the prognosis of gliomas. The signature remained a strong and independent prognostic indicator for OS in the training and validation cohort by inclusion of statistically significant clinical factors determined by univariate Cox regression analysis and the risk signature into multivariate Cox regression analysis. In assessing of model predictive performance, the 1-year, 3-year, and 5-year AUC values in the training and validation cohort were all more than 0.75. The results indicated that our predictor has higher accuracy in predicting the prognosis of glioma patients at three prediction time points. Nomogram has been widely applied in cancer prognosis assessment for providing an opportunity of predicting individual survival probability. Here, we constructed a robust nomogram composed of risk score, age, grade and 1p19q status to predict the prognosis of glioma patients. In addition, the AUC values of the integrated nomogram at 1 year, 3 years and 5 years were greater than the AUC values of age, grade, 1p19q status and the five-lncRNA signature. The calibration curves demonstrated that the predicted and observed probabilities shown great consistency with each other. These results revealed that the nomogram performed well in predicting the prognosis of glioma patients.

In the TCGA and CGGA datasets, differential expression analysis showed that CYTOR and MIR155HG were significantly overexpressed in GBM compared to LGG samples, while AC120036.4, LINC00641 and PWAR6 showed an inverse expression pattern in glioma samples. Insights into the molecular function of lncRNAs in glioma would accelerate their clinical research and practice. Among the hub lncRNAs, some have been proven to be related to glioma. Recent research reveals overexpression of CYTOR in glioma tissues correlated with metastasis and knockdown of CYTOR can attenuate the tumor cell proliferation and invasion [26]. MIR155HG has been considered to be significantly upregulated in GBM samples in comparison with LGG tissues [27]. The expression of MIR155HG has been found to be positively correlated with tumor grade and prognosis of glioma patients. Meanwhile, MIR155HG knockdown decreases glioma cell proliferation both in vitro and in vivo [27]. Recent study demonstrates MIR155HG is highly expressed in mesenchymal GBM tissues and suppression of MIR155HG can decrease expression of mesenchymal transition-associated proteins, such as β -catenin, N-cadherin, suggesting a function as a regulator in mesenchymal transition progression [28]. Identical to our research, recent bioinformatics analysis reveals PWAR6 and LINC00641 are expression-dysregulated in glioma tissues [29-31]. Meanwhile, PWAR6 has also been considered as a modulator of tumor immunoreaction and epithelial-mesenchymal transition (EMT) [29]. However, the in-depth analysis of mechanisms of AC120036.4 and LINC00641 has not been identified in glioma.

A huge number of putative lncRNAs have been predicted or validated in human. However, the functions or molecular mechanisms of the great part of lncRNAs remain unclear. To infer the potential roles of the hub lncRNAs with dysregulated expression levels, we constructed a lncRNA-miRNA-mRNA regulatory network.

We found that several miRNAs in our ceRNA regulatory network are mainly involved in glioma pathway. Namely miR-155 and miR-107 are associated with poor prognosis in glioma patients and involved in cell proliferation by changing the tumor cell cycle [32, 33]. And the target genes such as WEE1 (WEE1 G2 checkpoint kinase) and CDK6 (cyclin-dependent kinase 6) in our ceRNA network have been considered as key regulators in cell cycle and proliferation [34, 35]. Meanwhile, based on the mRNAs in the network, we carried out the enrichment analysis. Pathway enrichment analysis indicated that the genes were mainly enriched in p53 signaling pathway, which play a sophisticated role in the progression of glioma [36], demonstrating the hub lncRNAs were expected to be involved in the glioma biogenesis and development.

Conclusion

In summary, the current study utilized comprehensive bioinformatics analysis to determine a risk signature based on lncRNAs, providing a potential tool for assessment of clinical prognosis of patients with glioma.

Abbreviations

LGG: low grade glioma; GBM: glioblastoma multiforme; lncRNAs: long non-coding RNAs; CGGA : Chinese Glioma Genome Atlas; TCGA: The Cancer Genome Atlas; TANRIC: The Atlas of Noncoding RNAs in Cancer; ROC: calibration; AUC: area under curve; WGCNA: Weighted gene co-expression network analysis; OS: overall survival; HR: hazard ratio; CI: confidence interval; DEmRNA: differential expressed mRNA; DEmiRNA: differential expressed miRNA; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; ceRNA: competing endogenous RNA; EMT: epithelial-mesenchymal transition.

Declarations

Authors' contributions

L W and BH L collected the data. CY Z and YQ T designed the research and conducted statistical analyses. PF X, LG Y and HT L searched the relevant literature. CY Z drafted the manuscript, QX C and DF T revised the manuscript.

Acknowledgment

We sincerely appreciate valuable advice and support from Dr. Jinping Zhang.

Competing interests

The authors declare no competing interests in this research.

Funding

This study was supported by Hubei Province Health and Family Planning Scientific Research Project (WJ2017M019).

Availability of data and materials

The data that support the findings of this study are available in TCGA database at <https://portal.gdc.cancer.gov/>, at TANRIC database <https://bioinformatics.mdanderson.org/public-datasets/>, and at CGGA database <http://www.cgga.org.cn/>.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

References

1. Weller M, Wick W, Aldape K, Brada M, Berger M, Pfister SM, Nishikawa R, Rosenthal M, Wen PY, Stupp R *et al*: **Glioma**. *Nat Rev Dis Primers* 2015, **16**:1-15017.
2. Lapointe S, Perry A, Butowski NA: **Primary brain tumours in adults**. *Lancet (London, England)* 2018, **392**(10145):432-446.
3. Aldape K, Brindle KM, Chesler L, Chopra R, Gajjar A, Gilbert MR, Gottardo N, Gutmann DH, Hargrave D, Holland EC *et al*: **Challenges to curing primary brain tumours**. *Nature reviews Clinical oncology* 2019, **6**:509–520.
4. Sultan M, Marcel H. Schulz², Richard² H, Magen¹ A, Klingenhoff⁴ A, Scherf M: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome**. *Science* 2008, **Vol. 321**(5891):956-960.
5. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. *Nat Methods* 2008, **5**(7):621-628.
6. Peng Z, Liu C, Wu M: **New insights into long noncoding RNAs and their roles in glioma**. *Molecular cancer* 2018, **17**(1):61.
7. Long Y, Wang X, Youmans DT, Cech TR: **How do lncRNAs regulate transcription?***Sci Adv* 2017, **3**(9):eaao2110.
8. Yang D, Qiao J, Wang G, Lan Y, Li G, Guo X, Xi J, Ye D, Zhu S, Chen W *et al*: **N6-Methyladenosine modification of lincRNA 1281 is critically required for mESC differentiation potential**. *Nucleic acids research* 2018, **46**(8):3906-3920.
9. Arun G, Diermeier SD, Spector DL: **Therapeutic Targeting of Long Non-Coding RNAs in Cancer**. *Trends Mol Med* 2018, **24**(3):257-277.

10. Jiao ZY, Tian Q, Li N, Wang HB, Li KZ: **Plasma long non-coding RNAs (lncRNAs) serve as potential biomarkers for predicting breast cancer.** *Eur Rev Med Pharmacol Sci* 2018, **22**(7):1994-1999.
11. Dastmalchi N, Safaralizadeh R, Nargesi MM: **LncRNAs: Potential Novel Prognostic and Diagnostic Biomarkers in Colorectal Cancer.** *Curr Med Chem* 2019.
12. Han Y, Wu Z, Wu T, Huang Y, Cheng Z, Li X, Sun T, Xie X, Zhou Y, Du Z: **Tumor-suppressive function of long noncoding RNA MALAT1 in glioma cells by downregulation of MMP2 and inactivation of ERK/MAPK signaling.** *Cell Death Dis* 2016, **7**:e2123.
13. Chen Q, Cai J, Wang Q, Wang Y, Liu M, Yang J, Zhou J, Kang C, Li M, Jiang C: **Long Noncoding RNA, Regulated by the EGFR Pathway, Contributes to Glioblastoma Progression Through the WNT/-Catenin Pathway by Scaffolding EZH2.** *Clin Cancer Res* 2018, **24**(3):684-695.
14. Li J, Han L, Roebuck P, Diao L, Liu L, Yuan Y, Weinstein JN, Liang H: **TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer.** *Cancer research* 2015, **75**(18):3728-3737.
15. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC bioinformatics* 2008, **9**:559.
16. Park SY: **Nomogram: An analogue tool to deliver digital knowledge.** *J Thorac Cardiovasc Surg* 2018, **155**(4):1793.
17. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP: **A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?** *Cell* 2011, **146**(3):353-358.
18. Li JH, Liu S, Zhou H, Qu LH, Yang JH: **starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data.** *Nucleic acids research* 2014, **42**(Database issue):D92-97.
19. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH *et al*: **miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions.** *Nucleic acids research* 2018, **46**(D1):D296-d302.
20. Agarwal V, Bell GW, Nam JW, Bartel DP: **Predicting effective microRNA target sites in mammalian mRNAs.** *eLife* 2015, **4**.
21. Liu W, Wang X: **Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data.** *Genome biology* 2019, **20**(1):18.
22. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *Omics : a journal of integrative biology* 2012, **16**(5):284-287.
23. Zhang J-X, Song W, Chen Z-H, Wei J-H, Liao Y-J, Lei J, Hu M, Chen G-Z, Liao B, Lu J *et al*: **Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis.** *The Lancet Oncology* 2013, **14**(13):1295-1306.
24. Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, Zhao Q: **Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis.** *International journal of biological sciences* 2018, **14**(2):124-136.

25. Guo X, Xiao H, Guo S, Dong L, Chen J: **Identification of breast cancer mechanism based on weighted gene coexpression network analysis.** *Cancer gene therapy* 2017, **24**(8):333-341.
26. Zou SF, Yang XY, Li JB, Ding H, Bao YY, Xu J: **UPF1 alleviates the progression of glioma via targeting lncRNA CYTOR.** *Eur Rev Med Pharmacol Sci* 2019, **23**(22):10005-10012.
27. Wu X, Wang Y, Yu T, Nie E, Hu Q, Wu W, Zhi T, Jiang K, Wang X, Lu X *et al.*: **Blocking MIR155HG/miR-155 axis inhibits mesenchymal transition in glioma.** *Neuro-oncology* 2017, **19**(9):1195-1205.
28. Balasubramaniyan V, Bhat KP: **Targeting MIR155HG in glioma: a novel approach.** *Neuro-oncology* 2017, **19**(9):1152-1153.
29. Lin X, Jiang T, Bai J, Li J, Wang T, Xiao J, Tian Y, Jin X, Shao T, Xu J *et al.*: **Characterization of Transcriptome Transition Associates Long Noncoding RNAs with Glioma Progression.** *Mol Ther Nucleic Acids* 2018, **13**:620-632.
30. 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.** *Molecular neurobiology* 2019, **56**(7):4786-4798.
31. Liang R, Zhi Y, Zheng G, Zhang B, Zhu H, Wang M: **Analysis of long non-coding RNAs in glioblastoma for prognosis prediction using weighted gene co-expression network analysis, Cox regression, and L1-LASSO penalization.** *Onco Targets Ther* 2019, **12**:157-168.
32. Chen L, Chen X-r, Chen F-f, Liu Y, Li P, Zhang R, Yan K, Yi Y-j, Xu Z-m, Jiang X-D: **MicroRNA-107 inhibits U87 glioma stem cells growth and invasion.** *Cell Mol Neurobiol* 2013, **33**(5):651-657.
33. Yang L, Li C, Liang F, Fan Y, Zhang S: **MiRNA-155 promotes proliferation by targeting caudal-type homeobox 1 (CDX1) in glioma cells.** *Biomed Pharmacother* 2017, **95**:1759-1764.
34. Yang X, Xiao Z, Du X, Huang L, Du G: **Silencing of the long non-coding RNA NEAT1 suppresses glioma stem-like properties through modulation of the miR-107/CDK6 pathway.** *Oncol Rep* 2017, **37**(1):555-562.
35. Kratassiouk G, Pritchard LL, Cuvellier S, Vislovukh A, Meng Q, Groisman R, Degerny C, Deforz E, Harel-Bellan A, Groisman I: **The WEE1 regulators CPEB1 and miR-15b switch from inhibitor to activators at G2/M.** *Cell cycle (Georgetown, Tex)* 2016, **15**(5):667-677.
36. Ohgaki H, Kleihues P: **Genetic alterations and signaling pathways in the evolution of gliomas.** *Cancer science* 2009, **100**(12):2235-2241.

Tables

Table 1. Univariate and multivariate Cox regression analysis of the risk score, clinical variables, and survival in the training and validation cohort						
Variates	Univariate Cox analysis			Multivariate Cox analysis		
	HR	95%CI	P	HR	95%CI	P
CGGA glioma training set (n = 181)						
Grade (ref. WHO IV)	0.101	0.070-0.171	<0.001	0.377	0.208-0.682	0.001
Gender (ref. Female)	0.966	0.664-1.404	0.854			
Age (median years, ref. <42)	1.910	1.310-2.786	<0.001	0.937	0.633-1.387	0.746
1p19q status (ref. codeletion)	5.442	2.901-10.206	<0.001	2.600	1.337-5.048	0.005
Risk score (continuous)	2.718	2.269-3.256	<0.001	2.002	1.584-2.530	<0.001
TCGA glioma validation set (n = 541)						
Grade (ref. WHO IV)	0.106	0.079-0.143	<0.001	0.327	0.204-0.500	<0.001
Gender (ref. Female)	1.129	0.854-1.495	0.394			
Age (median, ref. <42)	5.250	3.654-7.544	<0.001	3.954	2.005-4.163	<0.001
1p19q status (ref. codeletion)	4.793	2.951-7.785	<0.001	3.166	1.873-5.352	<0.001
Risk score (continuous)	2.076	1.853-2.325	<0.001	1.243	1.063-1.469	0.007
Abbreviations: HR, hazard ratio; CI, confidence interval.						

Figures

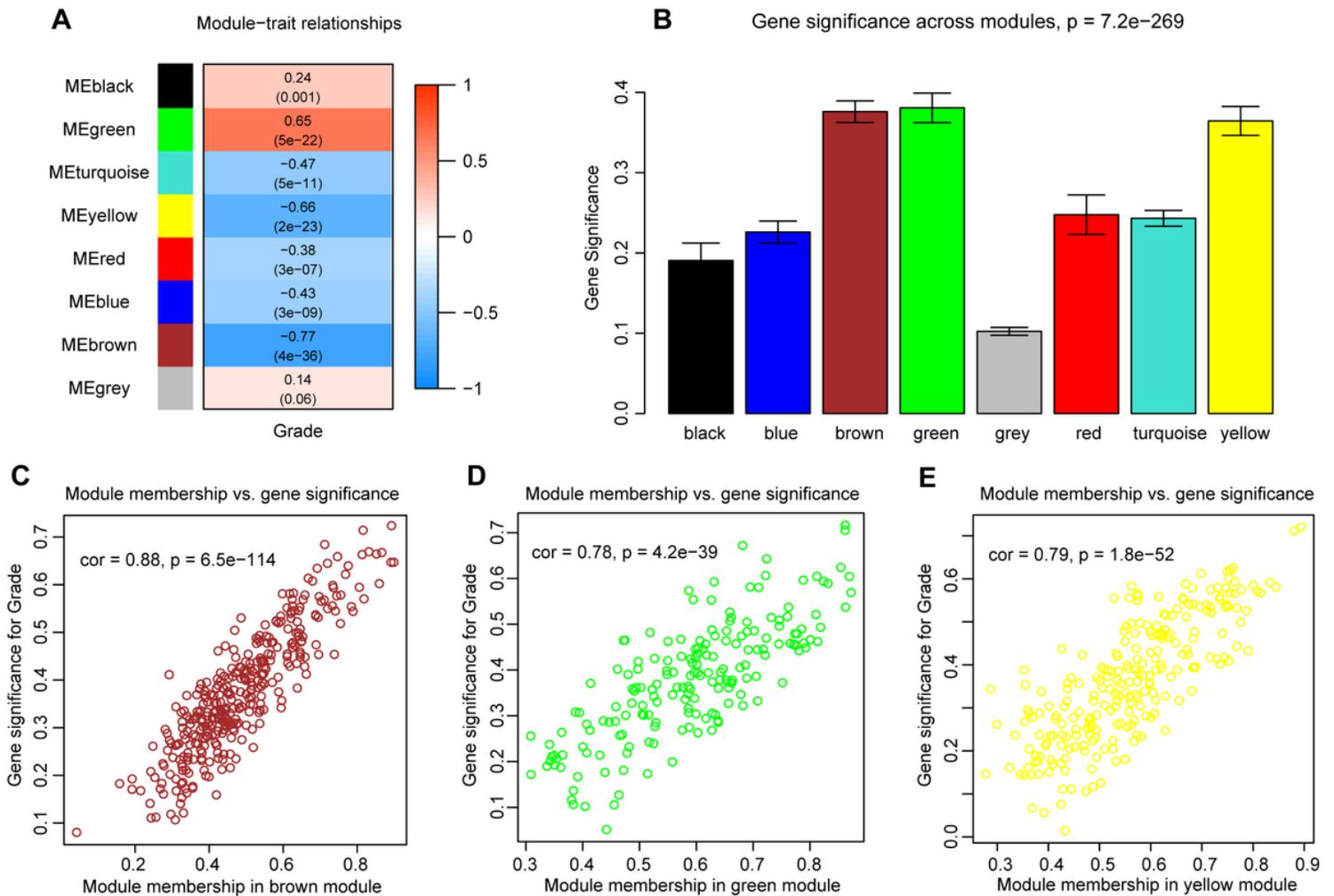


Figure 1

Weighted gene co-expression analysis (WGCNA) of lncRNA expression profiles in the training set. (A) Heatmap of Pearson correlation analysis of modules and the clinical trait of glioma. Rows represent module eigengenes and columns traits. (B) Barplot of average gene significance in eight modules. The green, yellow and brown modules were worth analyzing. (C-E) Scatter plot of module eigengenes in brown, green, and yellow modules.

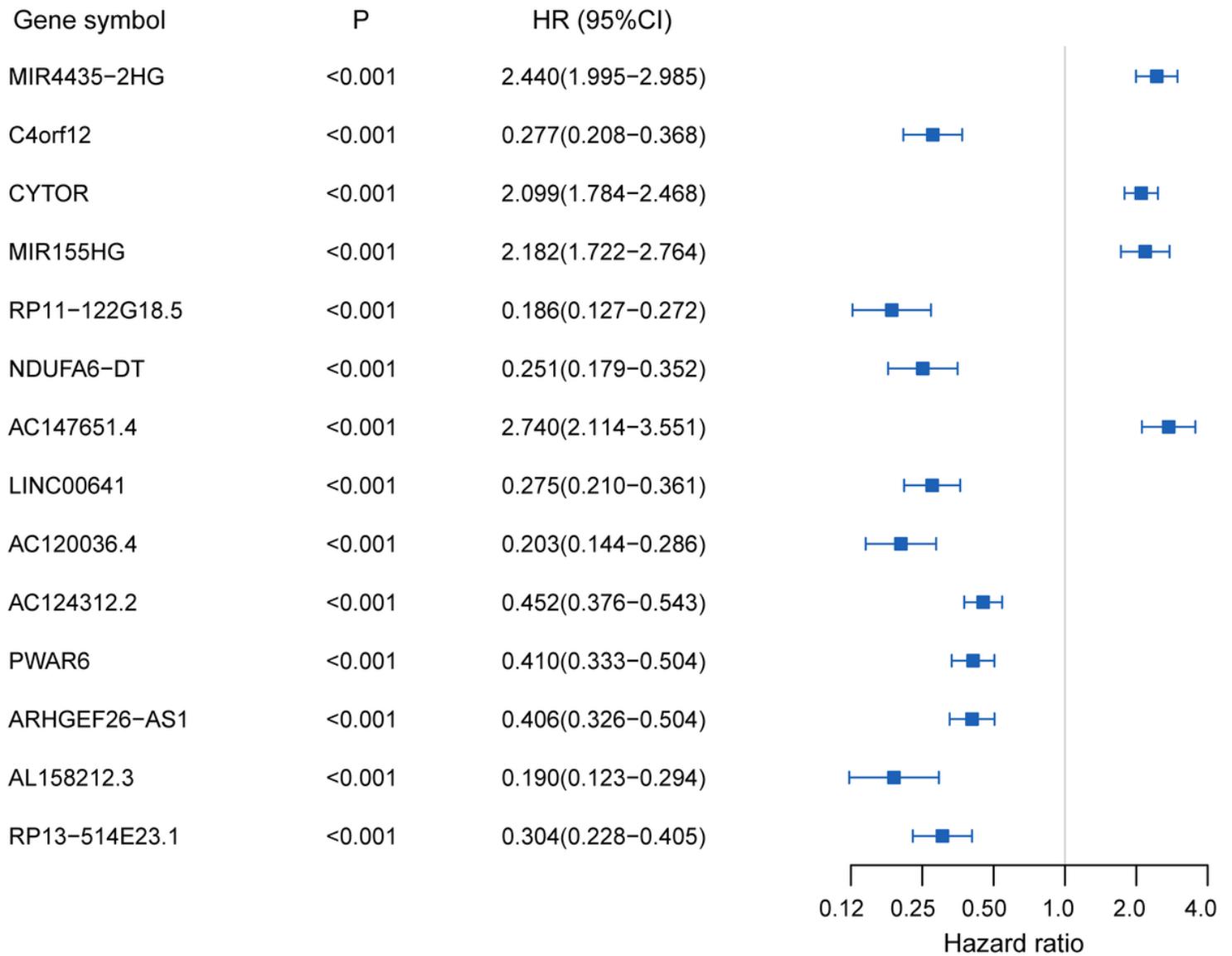


Figure 2

Forest plot of univariate Cox regression analysis of the 14 candidate lncRNAs with overall survival of glioma patients. Abbreviations: HR, hazard ratio; CI, confidence interval.

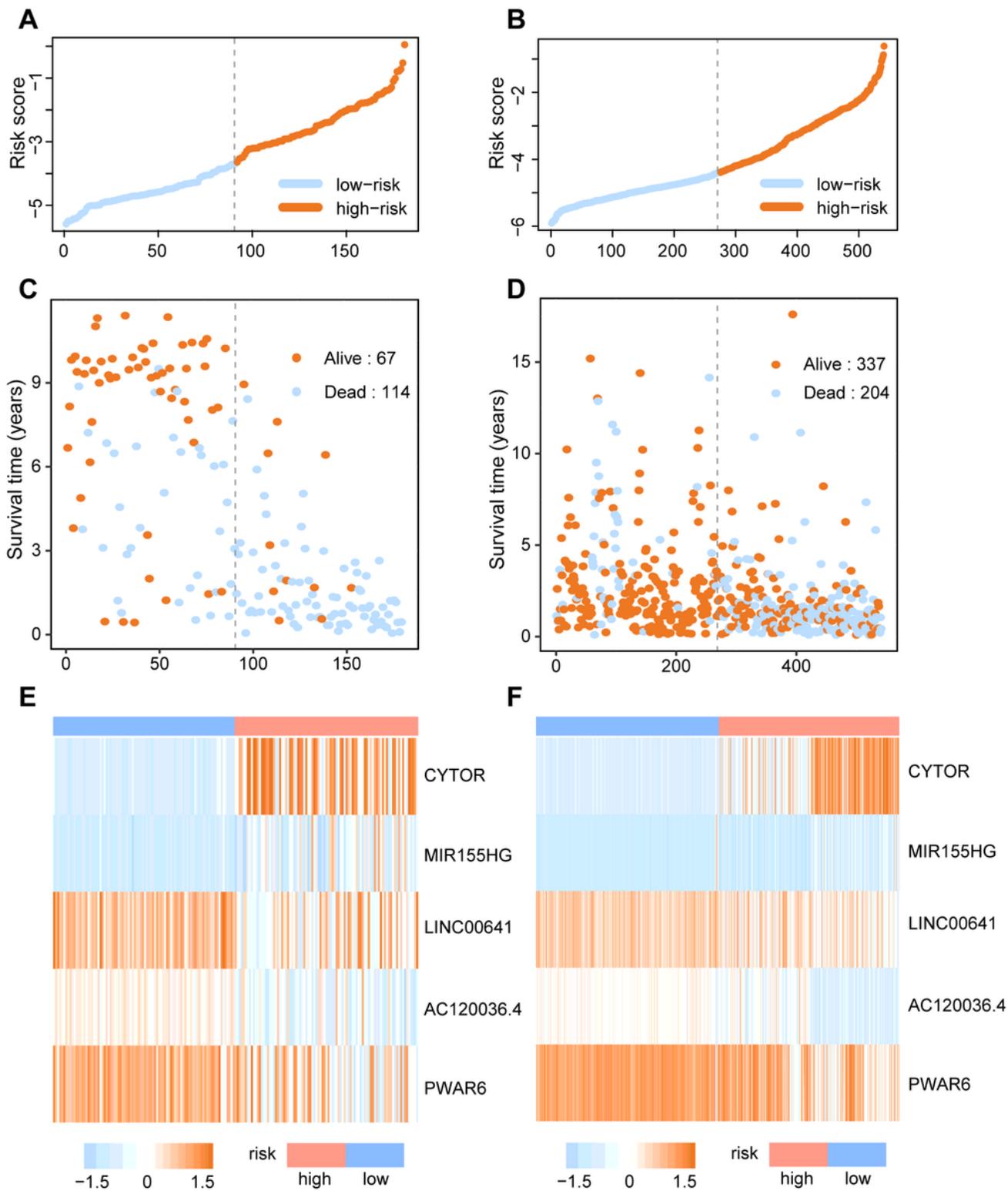


Figure 3

Patients were grouped according to the median risk score in training (left) and validation sets (right). Scatter plot of distribution of risk scores (A-B), survival state of glioma patients (C-D) and heatmap of five hub lncRNAs expression profiles in training (E) and validation (F) cohorts.

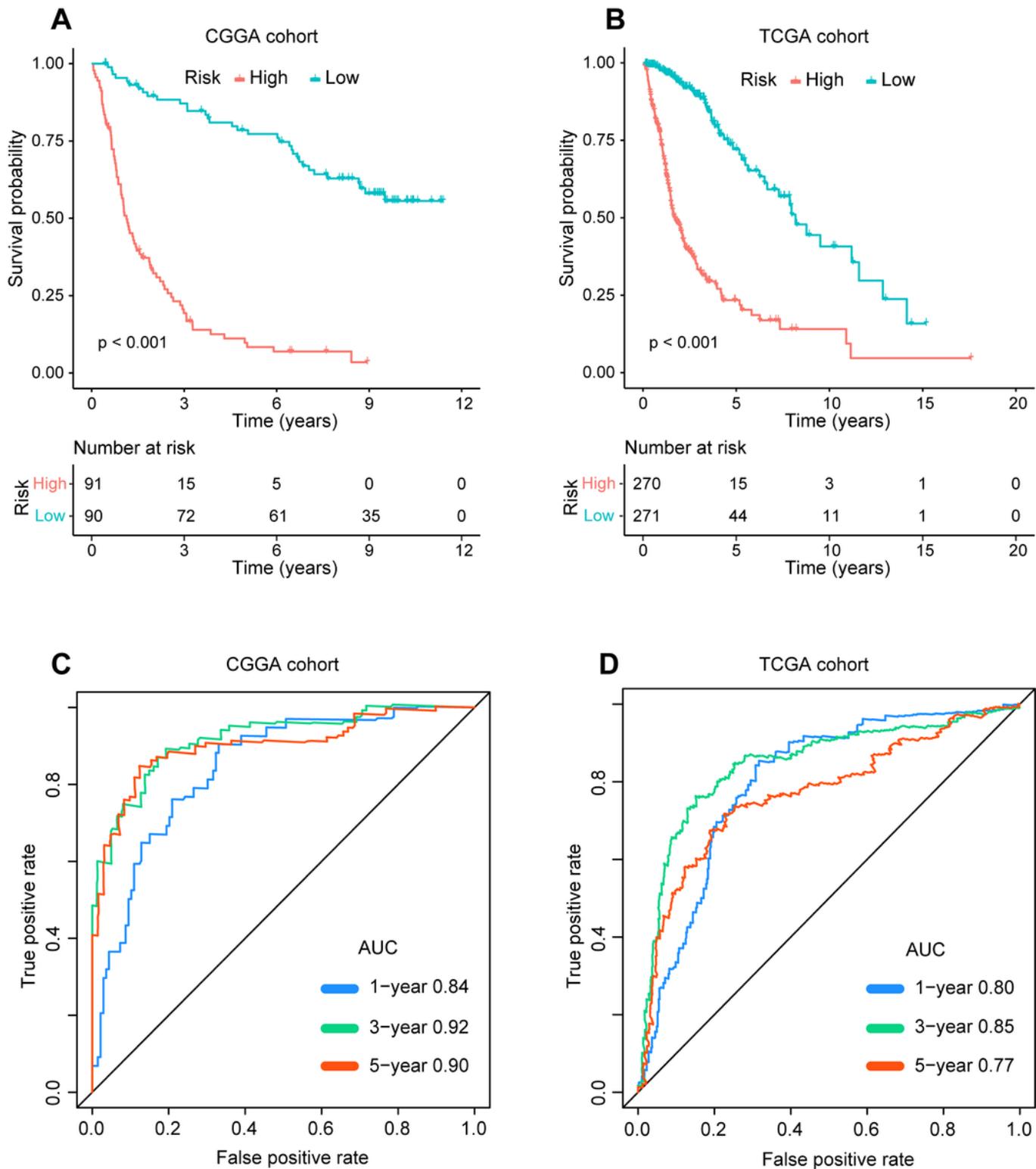


Figure 4

Survival analysis and predictive accuracy assessment of the five-lncRNA signature. Kaplan-Meier curve demonstrated significant survival differences between high and low groups in the training (A) and validation (B) sets. Receiver operating characteristic (ROC) curve showed the great predictive accuracy of the signature in the training (C) and validation (D) sets.

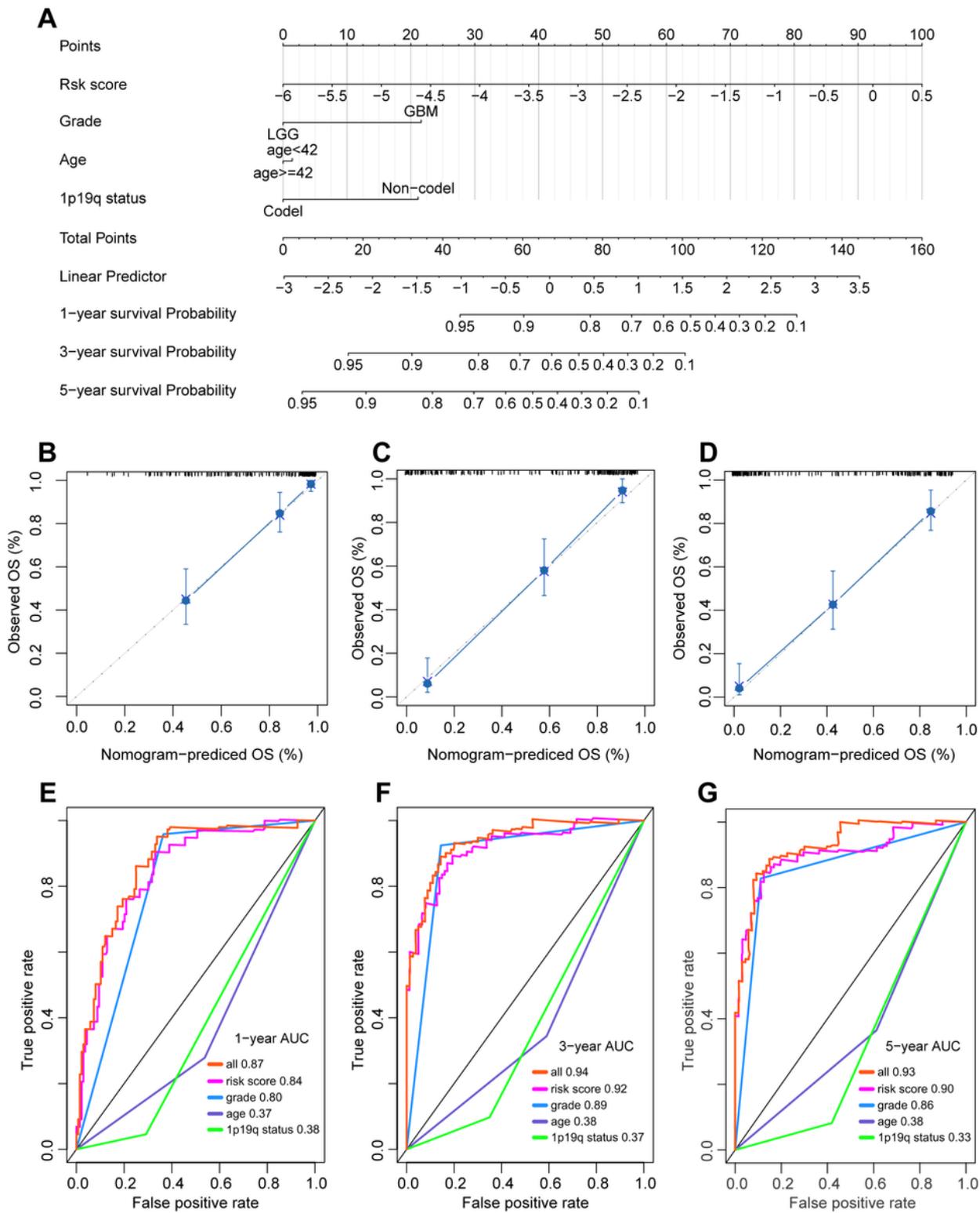


Figure 5

Nomogram construction. (A) A nomogram predicted the survival possibility of gliomas in the training set. (B-D) Calibration plot for the nomogram demonstrated the observed and predicted OS probabilities were highly consistent. (E-G) ROC curve assessed the prognosis prediction accuracy of the nomogram at 1-, 3- and 5-year overall survival. Abbreviations: OS, overall survival; AUC, area under curve.

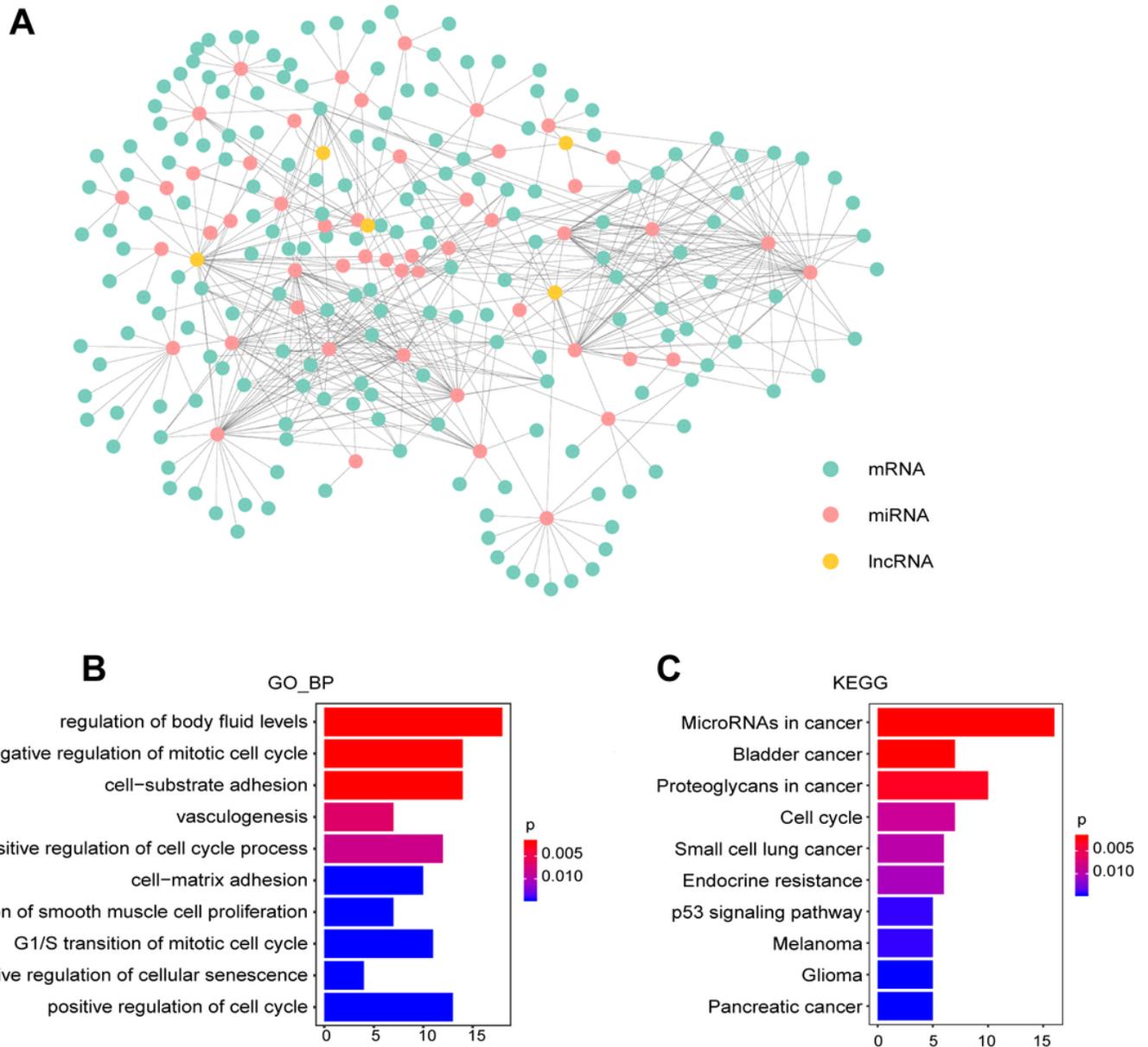


Figure 6

The ceRNA network construction and function annotation. (A) The ceRNA regulatory network contained 70 miRNAs, 182 mRNAs, and 5 lncRNAs. Nodes represented RNAs and lines interactions between RNAs. (B) GO enrichment analysis of mRNAs in ceRNA network. (C) Pathway enrichment analysis of mRNAs in ceRNA network. Abbreviations: ceRNA, competing endogenous RNA; GO, gene ontology; BP, biological process.

Supplementary Files

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

- [Supplementarymaterials.docx](#)