Int J Clin Exp Med 2021;14(6):1957-1970
www.ijcem.com /ISSN:1940-5901/IJCEM0123882

Original Article

A 7-lncRNA assessment model predicts glioblastoma multiforme patients' prognoses and survival

Kai Zhao[1], Chaojun Yu[2], Qian Wen[1], Ying Dai[1], Ji Luo[1]

[1]*Neurosurgery Department, The Second Affiliated Hospital of Kunming Medical University, Kunming 650100, Yunnan Province, China;* [2]*Neurosurgery Department, 903 Hospital, Jiangyou 621709, Sichuan Province, China*

Received October 10, 2020; Accepted January 20, 2021; Epub June 15, 2021; Published June 30, 2021

Abstract: Glioblastoma multiforme (GBM) is the most common type of central nervous system tumor and the type with the highest rates of malignancy and mortality. Due to its complex molecular mechanism and pathogenesis, the treatments are often ineffective, and there are significant individual treatment and prognostic differences. Long non-coding RNAs (lncRNAs) may play an important role in the pathogenesis of GBM. Through The Cancer Genome Atlas (TCGA) database, we used a univariate COX regression, a LASSO regression, a Kaplan-Meier (K-M) survival analysis, and multivariate COX to screen and construct an integrated lncRNA model to predict GBM patients' survival times. Using a co-expression analysis, Gene Ontology (GO), and a Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, we can indirectly explain the function of LncRNAs in the model. We constructed predictive models of 7-lncRNAs, including AC010327.4, AC080038.3, AL132822.1, BNC2-AS1, HOXC-AS3, NALT1, and SNHG18. The specificity and sensitivity of the model were verified using the ROC. The AUC (area under the curve) values of 3 years and 5 years were 0.842 and 0.88 respectively, indicating that the model predicted the patients' survival performance well. Through a coexpression analysis, the 7-lncRNA model may act as potential prognostic biomarkers and play an important role in the pathogenesis of GBM. We have constructed a clinical survival prediction model based on the 7-lncRNA expression profile. It can be used to predict the overall three- and five-year survival times, and the model has been shown to be accurate and reliable.

Keywords: TCGA database, glioblastoma multiforme, lncRNA, bioinformatics, survival, prognostic model

Introduction

Gliomas are the most common tumors in the central nervous system, accounting for about 30% [1]. They are characterized by infiltrative growth, a high degree of malignancy and recurrence, and a poor prognosis, among which GBM is the most common type and the type with the worst prognosis [2, 3]. The 5-year survival rate is less than 1%, and the median survival period is only 14-17 months [4]. In 2016, biomolecule markers were added to the glioma classification basis as the WHO central nervous system classification guide, to classify glioma more accurately [5]. Studies on mRNA, miRNA, and DNA methylation, and copy number variation have been widely reported in gliomas. These markers are helpful to distinguish the glioma subtypes [4].

LncRNAs are a non-coding RNA molecules larger than 200 nt that lack a continuous open reading frame, so they cannot encode proteins [6, 7]. Previously, lncRNAs were considered to be the "noise" of genome transcription, a by-product of RNA polymerase II transcription, devoid of any biological function. Recent studies have shown that lncRNAs have a large number and different lengths, and they carry more information, so it is easier to form a variety of functional advanced structures, involving various cellular physiological processes, including chromatin structure, epigenetics, RNA transcription, splicing, editing, and translation [8], overall playing an extensive regulatory role in various levels of gene expression [9]. These lncRNAs may become markers for the early diagnosis of gliomas, they can effectively predict patient prognosis, and we can even develop new treatments using these lncRNAs [10-12]. lncRNAs are of great significance for the efficient diagnosis and treatment of gliomas [13]. To make more effective use of the existing information and data resources, our re-

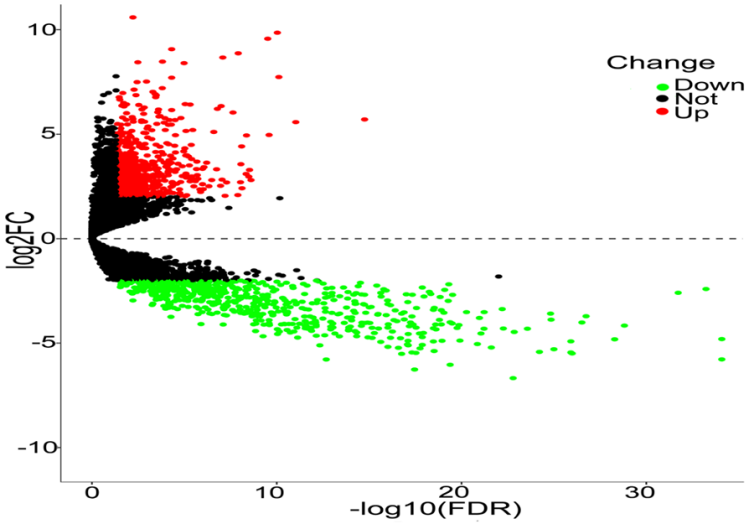

Figure 1. A volcano map of the differential lncRNAs. The edgeR package was used to do a differential analysis of the lncRNA expression in the normal tissues and tumor tissues. In the figure, red indicates that the genes were highly expressed in the tumor samples, while greenindicates a low expression in the samples, and black indicates no difference between the tumor samples and the normal samples.

Table 1. 16 lncRNAs with significant univariate Cox regressions

LncRNA	HR	*P*-Value
HOXC-AS2	1.187910129	0.000249224
NALT1	1.327634215	0.000933215
HOXC-AS3	1.223405126	0.000965228
AC005632.5	1.468289091	0.001697464
HCG21	1.235258873	0.002584322
AC011603.2	1.259766024	0.003142749
HOTAIR	1.112044234	0.00388437
AL356019.2	0.692030194	0.004520976
AP002807.1	1.259107699	0.005158091
AL132822.1	1.276899805	0.005868751
AC080112.2	1.249236677	0.00612997
AC080038.3	1.329389665	0.006190354
BNC2-AS1	1.168771953	0.006804842
HOTAIRM1	1.179011188	0.007957243
AC010327.4	1.227320491	0.008661644
SNHG18	1.188548315	0.00879218

search used the TCGA database to find valuable differential lncRNAs, establish survival-related clinical prognostic models, and predict patients' three- and five-year survival times. Moreover, to understand the potential functions and mechanisms of these lncRNAs, a co-expression analysis can predict the corresponding protein-coding target genes and indirectly explain their mechanism of action in tumorigenesis and tumor development.

Materials and methods

Data

The gene expression and corresponding clinical information of GBM patients came from the Cancer Genome Atlas (TCGA). The website is https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga. The database access date was August 1, 2020, and our searches were limited to human species. Among our search results, five were normal tissues and 169 were glioblastoma. The samples were sequenced on the Illumina HiSeq RNA-Seq platform. The patient expression profile data and the clinical information are public and freely available. Therefore, this study does not require an ethics committee approval.

lncRNA differential expression profile

First, we downloaded the GBM gene expression profile data from the TCGA database, we downloaded the human gene name conversion annotation file in the GENCODE database (https://www.gencodegenes.org/), and then we annotated the relevant genes in the expression profile with gene symbols.

Then we selected all the lncRNAs, using the above steps, and a total of 14,082 lncRNAs expression profiles were retrieved. Next, we grouped normal samples and the tumor samples, and we used the edgeR package to estimate the different lncRNAs. When |log2FC| >2 and P<0.05, this was considered differentially expressed lncRNA in our subsequent analysis.

Model establishment and verification

We sorted out the clinical data downloaded the TCGA, we then extracted the patient sample numbers, the survival status, and the overall survival times. We deleted the normal samples because there is a possibility of duplicate records in the clinical samples, so we deleted

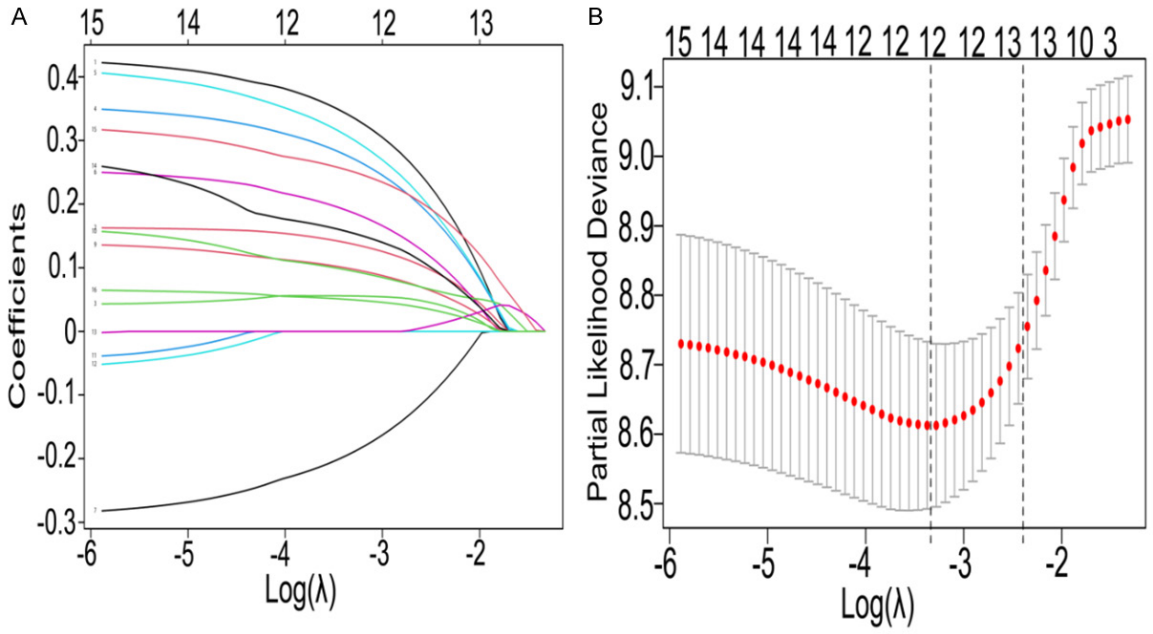

Figure 2. A LASSO regression screening for the lncRNAs. A. LASSO coefficient profiles of the 16 texture features. B. Selection of the tuning parameter (λ) in the LASSO model via a 10-fold cross-validation based on the minimum criteria. Optimal resulted in 12 lncRNAs.

the duplicate sample data, and then integrated the samples with the expression data, and then we performed a univariate cox model analysis to calculate the relationship between the expression level of each lncRNA and the overall patient survival. The lncRNAs with P< 0.01 were selected, and these lncRNAs were statistically significant in our univariate cox analysis. Next, we used lasso regression to select the lncRNAs, and this step served to screen out the lncRNAs based on the minimum lambda method (R package "glmnet").

We performed a KM survival analysis on the above lncRNAs to obtain the lncRNAs that have an impact on survival. Then, a multivariate cox analysis and a stepwise regression were used to select the best model to obtain the lncRNAs for the final clinical prediction model, and the 'survminer package' was used to visualize the multivariate cox results. The prognostic index can be calculated according to the lncRNA expressions obtained using the regression coefficient β of the model and the multivariate cox regression, and the prognostic performance can be evaluated using the 3-year and 5-year time-dependent receiver operating characteristic (ROC) curve. Based on the median prognostic index, the GBM patients were divided into high-risk and low-risk groups, and a Kaplan-Meier survival analysis of the high-risk and low-risk cases was performed.

The co-expression method predicts the lncRNA protein coding genes (PCG) and functions

Through the Pearson correlation coefficient, the target gene prediction of lncRNAs in the prognostic model indirectly clarifies its potential function and mechanism. The protein coding genes (PCG) with a Pearson correlation coefficient >0.40 and P<0.01 are considered to be protein-coding genes related to lncRNAs, so we selected the 'clusterProfiler package' for the target genes predicted by lncRNA to do the GO and KEGG enrichment pathway analyses [14]. The above biometric analysis was performed using R (R-4.0.0).

Results

Differentially expressed lncRNAs in GBM patients

Using 174 GBM samples (5 normal samples, 169 tumor samples) in the TCGA database,

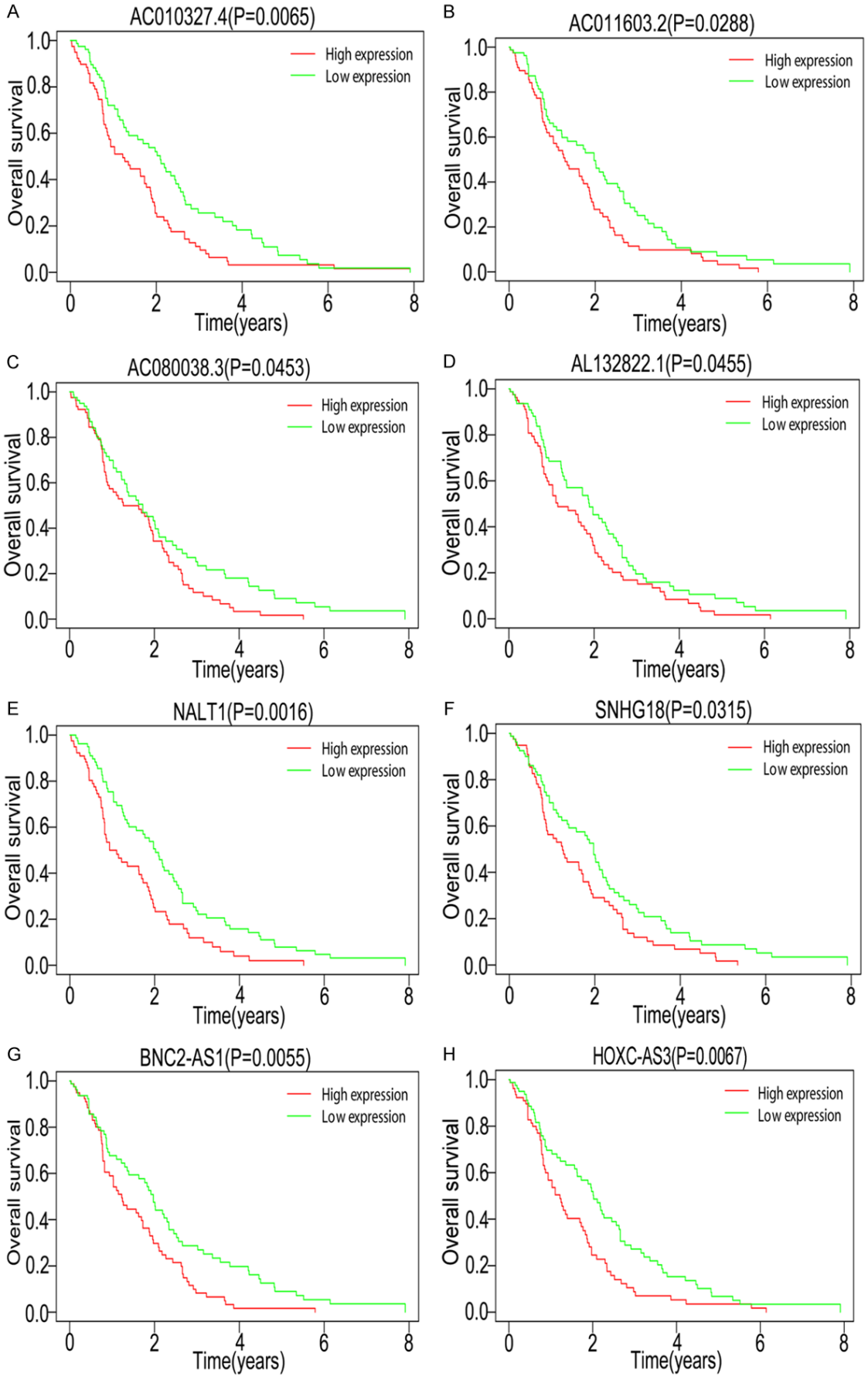
A
AC010327.4(P=0.0065)
B
AC011603.2(P=0.0288)
C
AC080038.3(P=0.0453)
D
AL132822.1(P=0.0455)
E
NALT1(P=0.0016)
F
SNHG18(P=0.0315)
G
BNC2-AS1(P=0.0055)
H
HOXC-AS3(P=0.0067)
High expression
Low expression
Overall survival
Time(years)

Figure 3. The Kaplan-Meier survival curves of 8 lncRNAs. (A) AC010327.4 (B) AC011603.2 (C) AC080038.3 (D) AL132822.1 (E) NALT1 (F) SNHG18 (G) BNC-AS1 and (H) HOXC-AS3. These genes with high expressions were significantly shorter than those with low expressions.

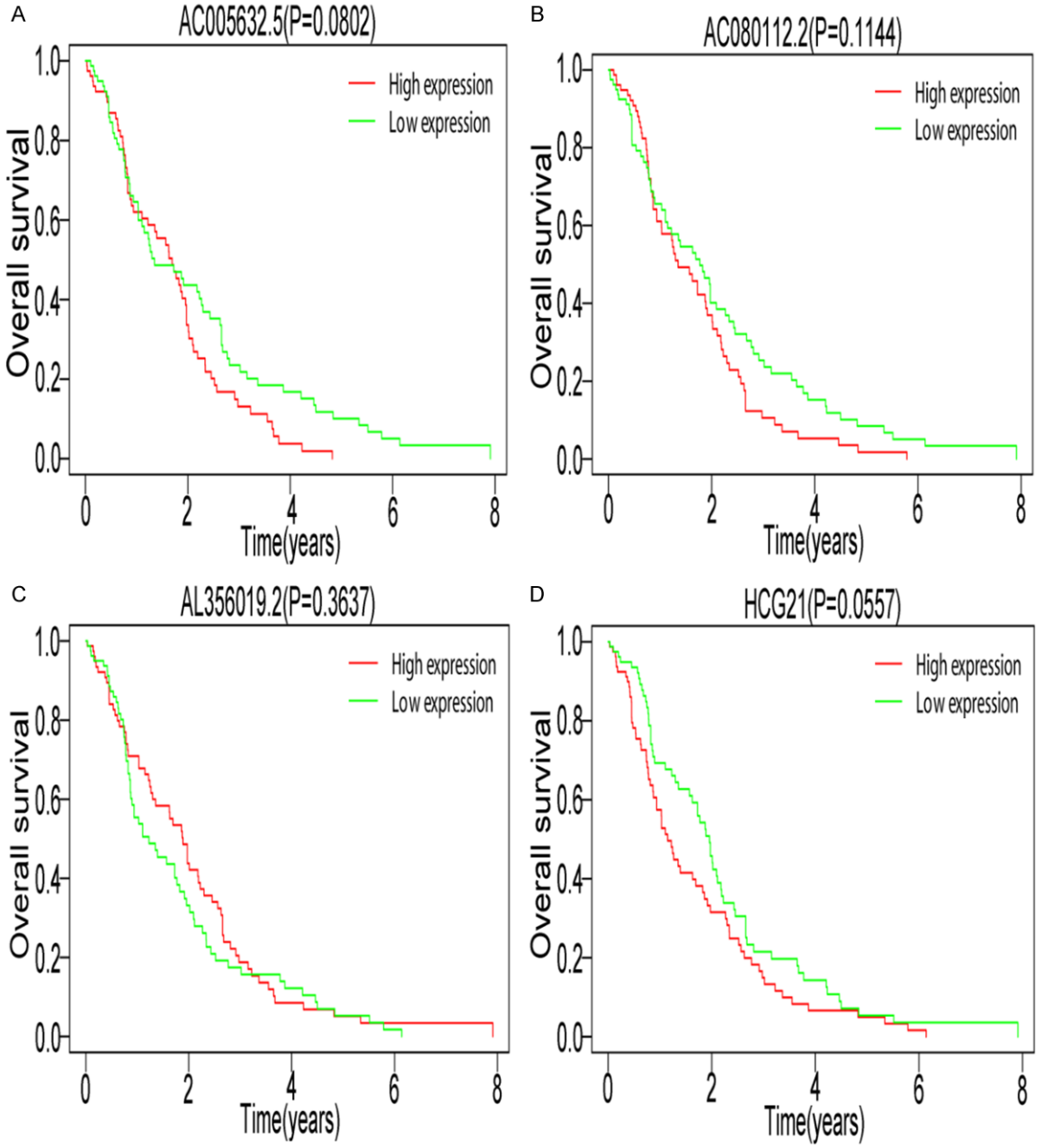

Figure 4. (A) AC005632.5 (B) AC080112.2 (C) AL356019.2 and (D) HCG21 had no significant effect on patient survival.

we used the 'edgeR package' to select the lncRNAs with different expressions for these patients, set |logFC| >2 and P<0.05, and finally screened the 1,226 lncRNAs with differential expressions. 610 were up-regulated, and 616 were down-regulated (**Figure 1**).

Construction of the lncRNA prognostic model

We kept the sample numbers, the survival statuses, and the overall survival times from the clinical information of the GBM patients downloaded using TCGA. The tumor samples were

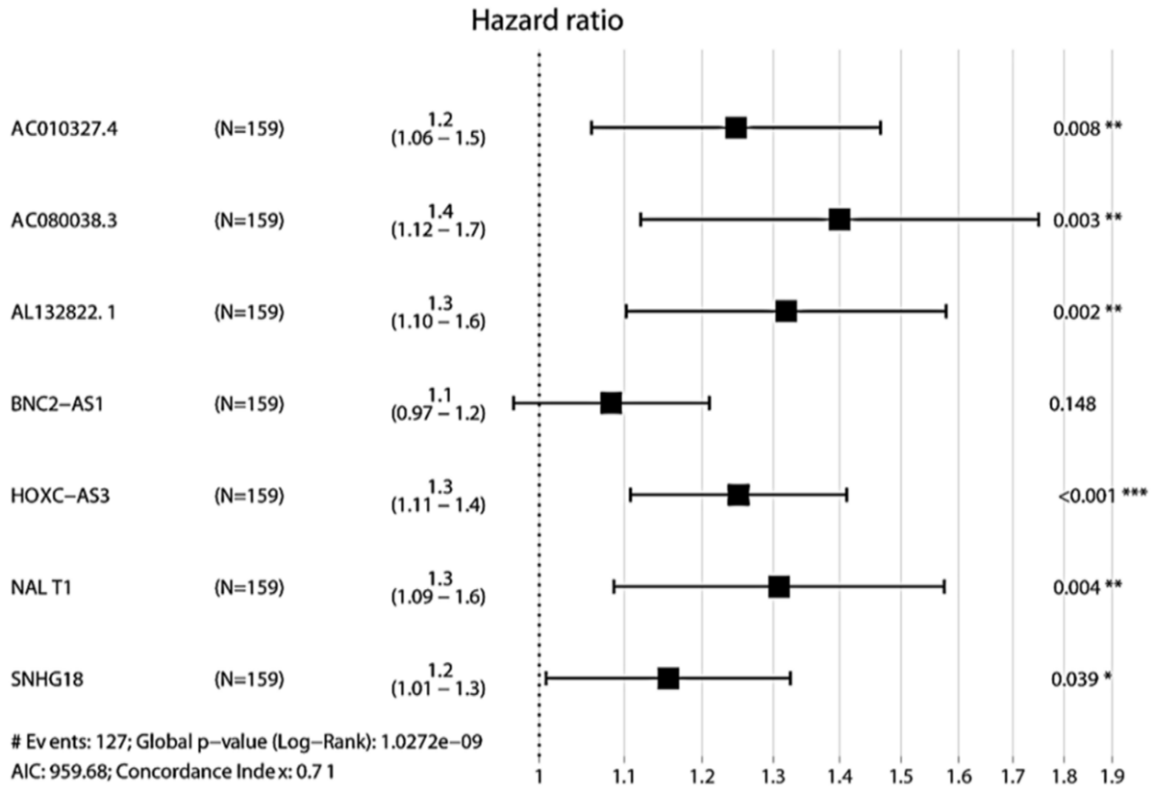

Figure 5. 7 A lncRNA forest map in the multivariate Cox regression model. HR>0 indicates that the expression of the 7 lncRNAs was positively correlated with the survival risk of the patients. The Concordance Index 0.71 indicated that the model has moderate predictive power.

selected to establish a prognostic model, so the lncRNA expression data from the 5 normal samples were deleted. In R, the clinical information and expression profile data were integrated using the sample numbers. We got rid of the missing data in the survival status and the data with duplicate sample numbers. Because of the follow-up lasso regression, the samples with a survival time of 0 were excluded. Finally, 159 GBM patients were left for the follow-up univariate cox regression analysis, and the lncRNAs with P<0.01 were selected for further analysis. These lncRNAs: (HOXC-AS2, NALT1, HOXC-AS3, AC005632.5, HCG21, AC011603.2, HOTAIR, AL356019.2, AP002-807.1, AL132822.1, AC080112.2, AC0800-38.3, BNC2-AS1, HOTAIRM1, AC010327.4, and SNHG18) met the conditions (**Table 1**).

Furthermore, the "glmnet" package was selected, and 12 lncRNAs were screened from 16 genes selected using a univariate Cox regression using lasso regression based on the minimum λ method (**Figure 2**). They were AC005-632.5, AC010327.4, AC011603.2, AC0800-38.3, AC080112.2, AL132822.1, AL3560-19.2, BNC2-AS1, HCG21, HOXC-AS3, NALT1, and SNHG18.

Then the survival analysis of the 12 genes showed that 8 lncRNAs (AC010327.4, AC-011603.2, AC080038.3, AL132822.1, NALT1, SNHG18, BNC2-AS1, and HOXC-AS3) had significant effects on survival, P<0.05 (**Figure 3**). The other four lncRNAs (AC005632.5, AC-080112.2, AL356019.2, and HCG21) (**Figure 4**) had no significant effect on survival. Through a multivariate cox stepwise analysis, the results showed that 7-lncRNAs (AC010327.4, AC080038.3, AL132822.1, BNC2-AS1, HOXC-AS3, NALT1, and SNHG18) make up the optimal prediction model (**Figure 5**). Based on the expression level multiplicative regression model (β) and the linear combination of the following expression, a prognostic index (PI) based on lncRNAs was established. In our model, PI = (0.22056 * AC010327.4 expression) + (0.33664 * AC080038.3 expression)

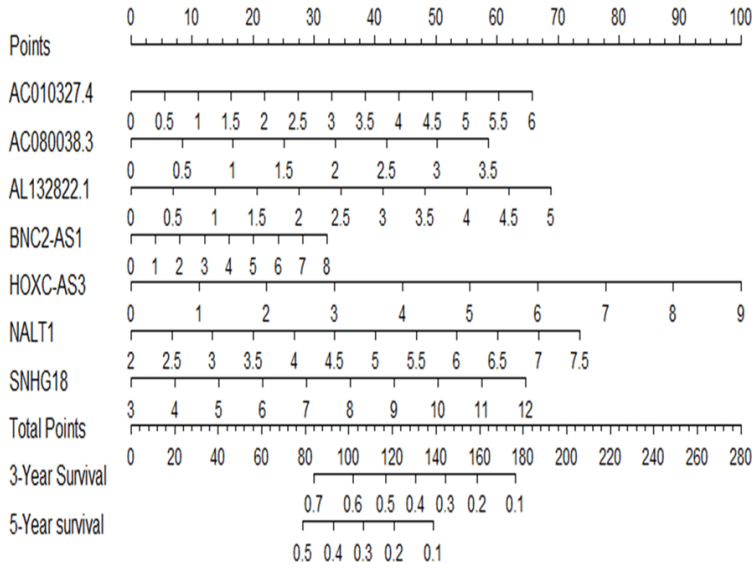

Figure 6. Prediction of the patient survival times by drawing a nomogram of the 7 lncRNAs in the model.

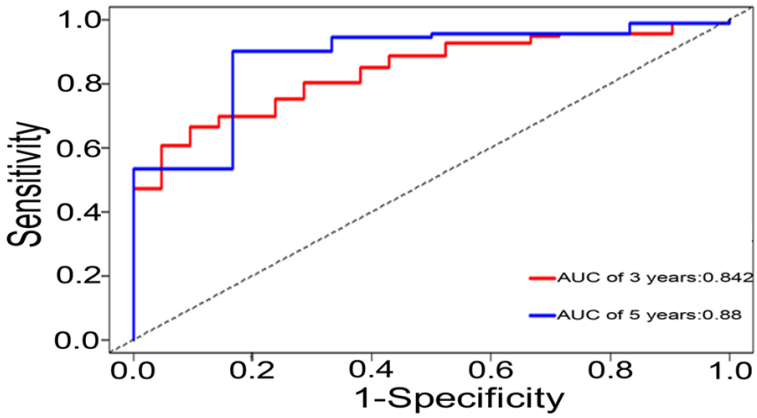

Figure 7. The 3-year and 5-year ROC curves. The predicted 3-year survival AUC value was 0.842, and the predicted 5-year survival AUC value was 0.88.

+ (0.27674 * AL132822.1 expression) + (0.08086 * BNC2-AS1 expression) + (0.22342 * HOXC-AS3 expression) + (0.26874 * NALT1 expression) + (0.14461 * SNHG18 expression). Prediction of patient survival times using a nomogram (**Figure 6**).

We evaluated the prognostic performance using 3-year and 5-year ROC curves, and the results showed that the 3-year survival AUC area was 0.842 and the 5-year survival area under the curve (AUC) was 0.88 (**Figure 7**), indicating that the model has good accuracy.

The lncRNA model predicted the score, and the patients were divided into a high-risk group (score > median) and a low-risk group (score < median) according to the median score, and then risk score charts were drawn (**Figure 8**).

We analyzed the patient survival in the high-risk and low-risk groups and confirmed that the overall survival times of the patients in the high-risk group were significantly shorter than the survival times in the low-risk group, P=3.193e-11, and the difference was statistically significant (**Figure 9**).

lncRNAs coexpression analysis predicts the target genes and functional enrichment

According to the Pearson correlation coefficient, we predicted the protein-coding target genes of 7 lncRNAs in the model, selected the expression values of the lncRNAs and mRNAs related to GBM in the TCGA database, and analyzed the co-expressions of the Pearson correlation. There was a significant correlation between |cor| >0.40, and P<0.05. AC010327.4 had 206 related protein-coding genes, AC080038.3 had 692 related protein-coding genes, BNC2-AS1 had 90 related protein-coding genes, HOXC-AS3 had 89 related protein-coding genes, NALT1 had 2829 related protein-coding genes, and SNHG18 had 143 related protein-coding genes. Only one protein-coding gene was obtained through the AL132822.1 coexpression analysis. Except for the co-expressed protein-coding genes of the positive and negative correlations in BNC2-AS1, the other 6 lncRNAs only had positively related genes. Here we listed the eight lncRNAs with the highest positive and negative correlation coefficients (CCNA1, HYPK, PABPC1L2A, BCAN, BNC2, HOXC10, MST1, and AL162231.1) (**Figure 10**).

The "clusterProfiler" package was selected in R, and the target genes predicted using ln-

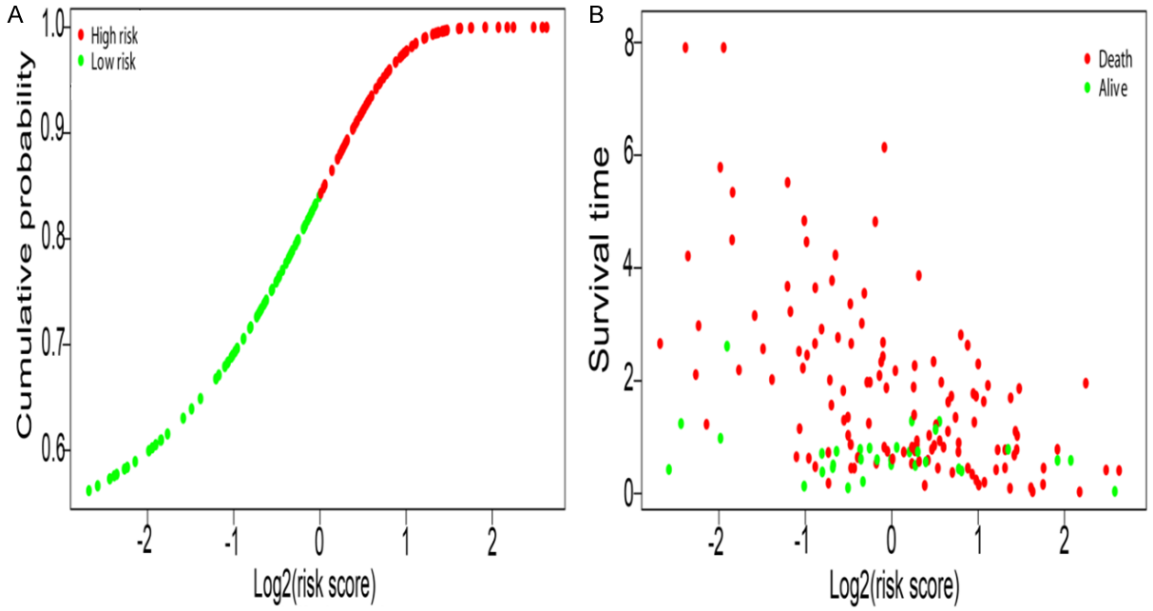

Figure 8. A risk score distribution chart. A. For this cumulative function distribution chart, we calculated the score of each patient, chose the median, and then sorted from low to high. B. The survival status of each patient corresponding to the scatter plot.

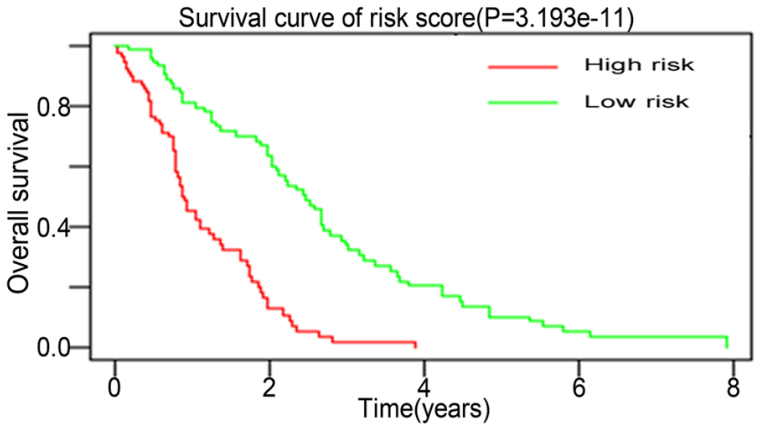

Figure 9. Survival curves of the patients in the high and low-risk groups. The survival times of the patients in the high-risk score group were significantly shortened, P<0.01, with a significant difference.

cRNAs related to the prognosis were analyzed using the GO functional annotation and the KEGG pathway enrichment analysis, and a possible mechanism was speculated. The results of the GO analysis mainly include biological process (BP), molecular function (MF), and cellular component (CC). Our results include RNA methyltransferase activity, DNA-dependent ATPase activity, ATPase activity, DNA-binding transcription activator activity, tRNA methyltransferase activity, protein serine/threonine kinase activity, activating transcription factor binding, and so on. The results of the GO enrichment analysis of the target genes predicted by AC080038.3, HOXC-AS3, and NALT1 are shown in (**Figures 11**-**13**). The other 6 lncRNAs are shown in (**Figures 14**-**17**). The predicted target genes were analyzed using the KEGG pathway, and a total of 12 paths were obtained (**Table 2**). We could see that the pathways were enriching the tumor-related genes, including transcriptional misregulation in cancer, choline metabolism in cancer, and the mRNA surveillance pathway, etc.

Discussion

Glioma is the most common intracranial tumor, among which GBM is the most common and has the worst prognosis, with related gene heterogeneity and tumor susceptibility, high mortality and easy recurrence after treatment, and a short survival time [4]. We intended to construct a clinical risk prediction model for GBM patients through data mining. Clinicians can use the prediction model to develop individual diagnoses and treatment strategies for pati-

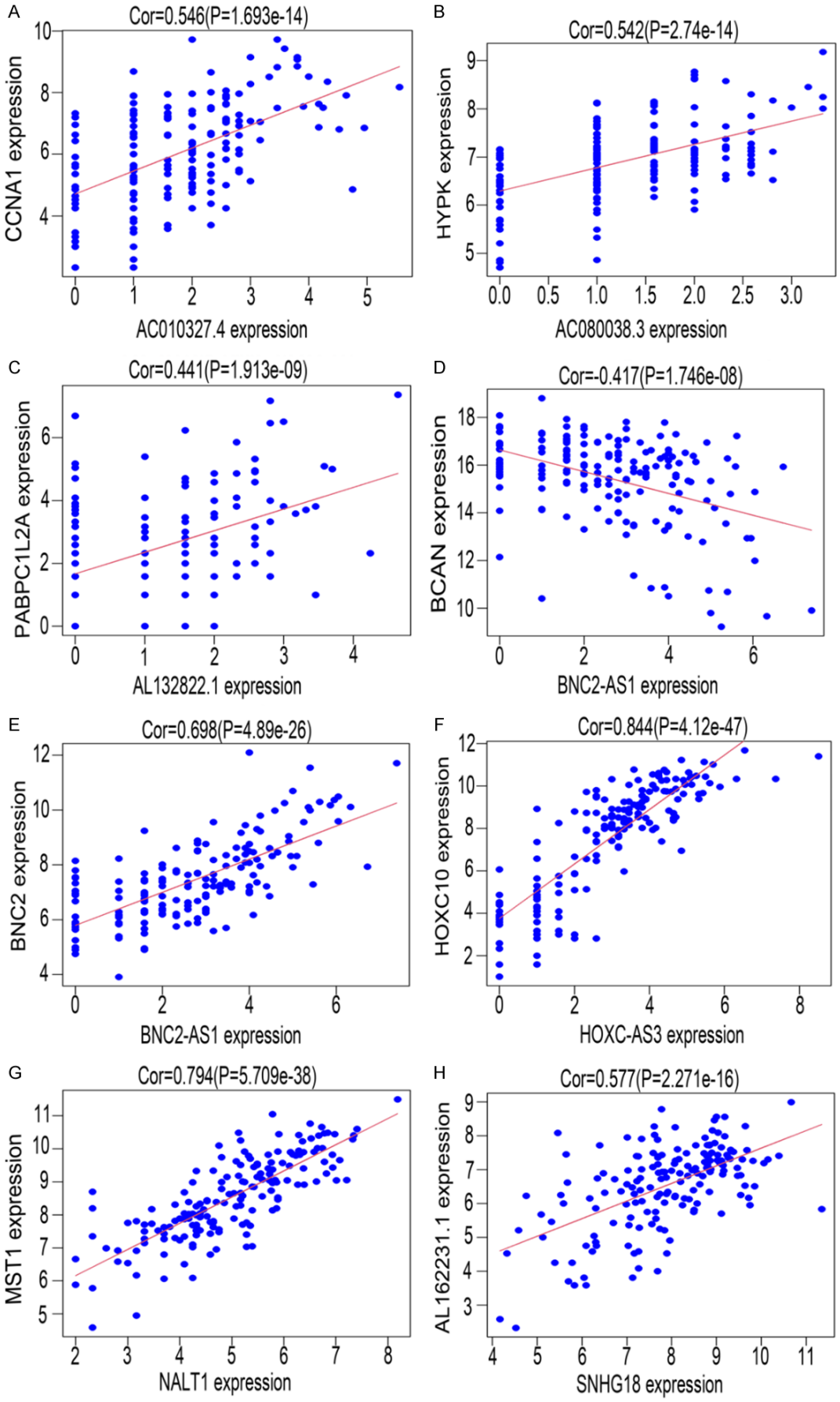
A
Cor=0.546(P=1.693e-14)
CCNA1 expression
AC010327.4 expression
B
Cor=0.542(P=2.74e-14)
HYPK expression
AC080038.3 expression
C
Cor=0.441(P=1.913e-09)
PABPC1L2A expression
AL132822.1 expression
D
Cor=-0.417(P=1.746e-08)
BCAN expression
BNC2-AS1 expression
E
Cor=0.698(P=4.89e-26)
BNC2 expression
BNC2-AS1 expression
F
Cor=0.844(P=4.12e-47)
HOXC10 expression
HOXC-AS3 expression
G
Cor=0.794(P=5.709e-38)
MST1 expression
NALT1 expression
H
Cor=0.577(P=2.271e-16)
AL162231.1 expression
SNHG18 expression

Figure 10. Pearson correlation analysis (|Pearson correlation coefficient| >0.40 and P<0.01) we selected the 7 lncRNA-related mRNAs. The most significantly positive and negative correlated mRNAs are shown. A. AC010327. 4's most positive relevant protein coding genes was CCNA1 (Cor=0.546, P=1.693e-14). B. AC080038.3's most positive relevant protein coding genes was HYPK (Cor=0.542, P=2.74e-14). C. AL132822.1's most positive relevant protein coding genes was PABPC1L2A (Cor=0.441, P=1.913e-09). D, E. The most negatively and positively related protein coding genes of BNC2-AS1 were BCAN (Cor=-0.417, P=1.746e-08) and BNC2 (Cor=0.698, P=4.89e-26), respectively. F. The most positively relevant protein coding gene of HOXC-AS3 was HOXC10 (Cor=0.844, P=4.12e-47). G. While NALT1's protein coding gene was MST1 (Cor=0.794, P=5.709e-38). H. And SNHG18's protein coding gene was AL162231.1 (Cor=0.577, P=2.271e-16).

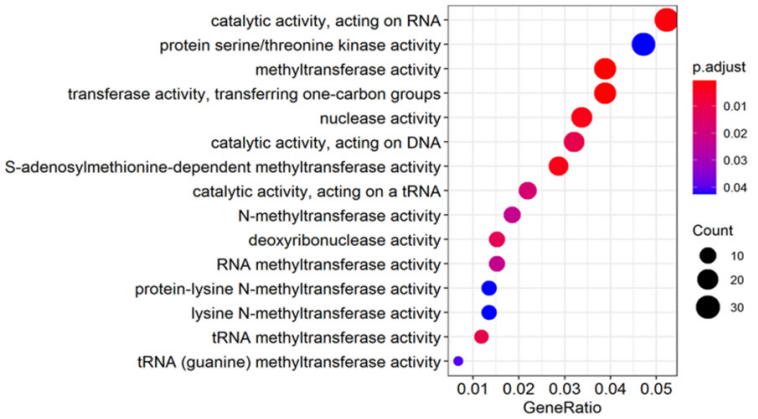

Figure 11. Dotplot of the target gene GO enrichment predicted by AC080038.3.

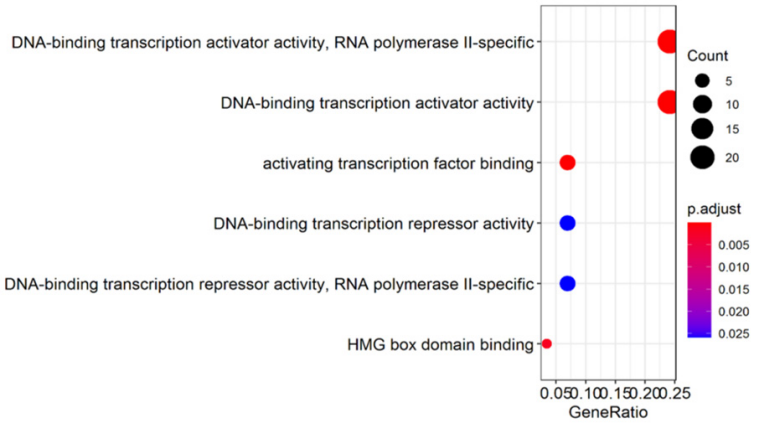

Figure 12. Dotplot of the target gene GO enrichment predicted by HOXC-AS3.

ents and effectively evaluate their prognoses. Compared with a single biomarker, multiple sets of biomarkers can improve the accuracy of the model prediction.

Gene chip technology and high-throughput sequencing technology are the most effective ways to detect and discover lncRNA at present [15]. The limitation of gene chip technology is that specific probes must be designed in the experiment, and these must be helped by the existing lncRNA database. High-throughput sequencing is needed for some unknown lncRNAs. Many large-scale multidimensional gene databases have been established, especially the tumor Genome Map (The Cancer Genome Atlas, TCGA), which includes a variety of gene data, including expression profiles, copy numbers, methylation, and other information, as well as relatively complete clinical data, which further shows that abnormal gene expression is closely related to the occurrence and development of tumors. At the same time, these differential genes can be associated with patients' clinical information to find out the genes that have a significant impact on survival, which is helpful to mine and find valuable cancer survival prediction models. This greatly reduces the scope of the vast research but also improves the efficiency of the research. In addition, the effective information in the database can guide the laboratory research, and the experimental research can verify the database information. In our study, according to the TCGA database, we found the GBM data, and we obtained the gene expression data and the sample clinical data. Then we extracted the expression profiles of the lncRNAs, analyzed the differences between the normal samples and the GBM samples using R, and found the related differentially expressed genes. Then a univariate cox analysis, lasso regression, and a multivariate

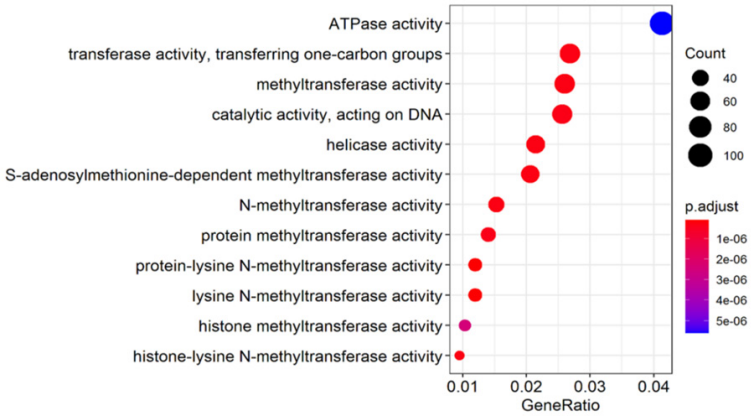

Figure 13. Dotplot of the target gene GO enrichment predicted by NALT1.

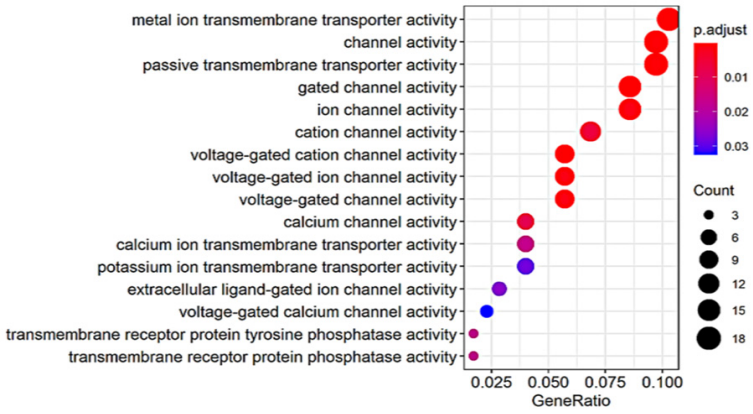

Figure 14. Dotplot of the target gene GO enrichment predicted by AC010327.4.

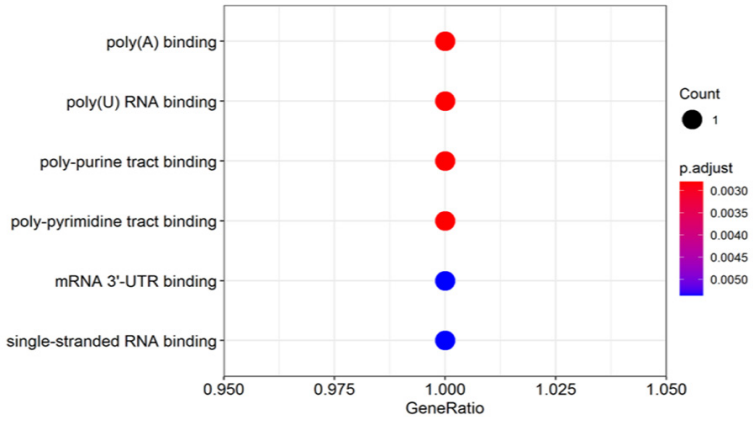

Figure 15. Dotplot of the target gene GO enrichment predicted by AL132822.1.

cox analysis were carried out to establish a GBM clinical prognostic risk model. Finally, new 7-lncRNA prediction models were constructed to predict the overall survival of patients with GBM. A time-dependent ROC analysis showed that our model had a good survival prediction. Since lncRNA does not encode proteins, to understand the function of the 7-lncRNAs, we analyzed the Pearson correlation between the lncRNA and protein-coding genes. Through the lncRNA-related mRNAs, we could indirectly explain the potential mechanism of the lncRNAs. We found that there are tumor-related pathways in the analysis of the functional enrichment of the GO and KEGG pathways, suggesting that the lncRNAs in this model are related to the tumors.

Among the 7-lncRNAs in the model, BNC2-AS1, HOXC-AS3, NALT1, SNHG18 have been described in the literature. Liu [16] found that BNC2-AS1 can promote the proliferation and growth of gastric cancer cells, and the BNC2-AS1 gene knockout can significantly inhibit the proliferation, migration, and invasion of gastric cancer cells. Yang [17] found that HOXC-AS3 is highly expressed in invasive mucinous adenocarcinoma of the lung. This team also found that the significant down-regulation of the HOXC-AS3 gene hinders the proliferation and migration of IMA cells. Shi [18] found that HOXC-AS3 is abnormally overexpressed in breast cancer, especially HER2. HOXC-AS3 upregulates the expression of the PPP1R1A protein, thus promoting the metastasis of breast cancer. Wang [19] indicated that HOXB13 is highly expressed in GBM cells U87 and U251, and HOXC-AS3 is involved in the HOXB13-induced proliferation, migration, and invasion of GBM cells.

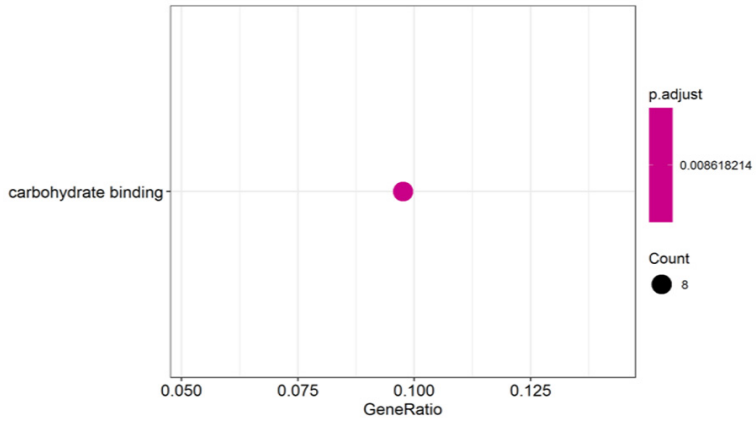

Figure 16. Dotplot of the target gene GO enrichment predicted by BNC2-AS1.

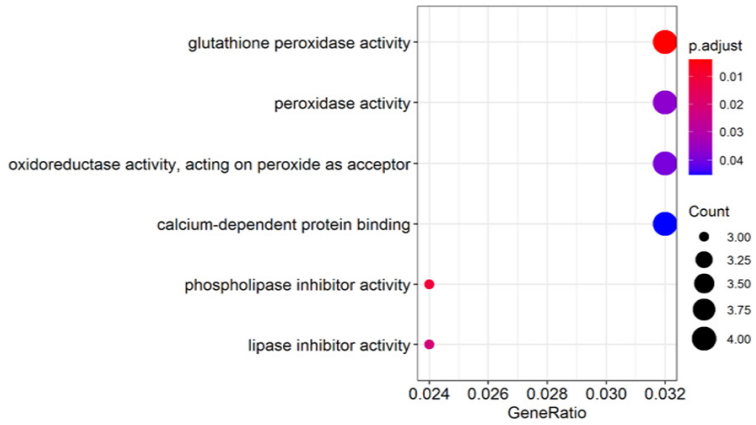

Figure 17. Dotplot of the target gene GO enrichment predicted by SNHG18.

Li [20] found that HOXC-AS3 is related to the occurrence and development of multiple myeloma. Zhang [21] found that HOXC-AS3 is significantly increased in gastric cancer and is related to the clinical prognosis of gastric cancer. Moreover, HOXC-AS3 regulates cell proliferation and migration both in vitro and in vivo. Piao [22] pointed out that the overexpression of NALT1 promotes the invasion and metastasis of gastric cancer, and the mechanism may be related to the regulation of NOTCH1 by NALT1 and its effect on the expression of the NOTCH signal pathway. Zheng [23] found that the expression of SNHG18 in clinical glioma tissues is significantly higher than it is in normal brain tissues. The expression of SNHG18 is correlated with the clinical tumor grade and negatively correlated with the mutation of isocitrate dehydrogenase 1. The knockout of SNHG18 can inhibit the radiation resistance of glioma cells, while the overexpression of SNHG18 has the opposite effect, as it can increase the tumor's resistance to radiotherapy.

So far, there are few studies covering AC010327.4, AC080038.3, or AL132822.1. However, in our study of the lncRNAs and mRNAs in the TCGA database, we found that the expressions of AC010327.4 and CCNA1 had the greatest correlation. Yang's meta-analysis indicates that there is a significant relationship between CCNA1 methylation and the slow progression of human malignant tumors [24]. Some studies have also shown the mechanism of CCNA1 in cervical cancer, nasopharyngeal carcinoma, breast cancer, and squamous cell carcinoma of the head and neck [25-29]. It has been reported that the most significant protein-coding gene, MAN2C1 of AC080038.3, promotes the formation of prostate cancer xenografts by activating AKT [30]. Jiang found that the MAN2C1 gene can promote the growth, invasion, and metastasis of transplanted tumors in mice [31]. Interestingly, AL132822.1 and its most significant protein-coding gene PABPC1L2A have not been reported in PubMed, so it is impossible to have an in-depth discussion on them.

However, our research also has some limitations. We tried to verify the prediction performance of the 7-lncRNA model in other large GBM datasets. Unfortunately, it was difficult to find datasets with complete GBM expression profile information and prognosis information such as TCGA, that is to say, there is no suitable dataset for us to verify in the next step.

Conclusion

In summary, we have constructed a clinical survival prediction model based on a 7-lncRNA expression profile. According to this model, any

Table 2. 7 lncRNAs co-expressed in the 12 pathways of the target gene KEGG enrichment

ID	Description	GeneRatio	Bg Ratio	p.adjust	Count
hsa05202	Transcriptional misregulation in cancer	4/19	192/8047	0.04411713	4
hsa05231	Choline metabolism in cancer	22/885	98/8047	0.03368915	22
hsa03015	mRNA surveillance pathway	25/885	91/8047	0.0004953	25
hsa00480	Glutathione metabolism	4/43	57/7104	0.02877192	4
hsa00514	Other types of O-glycan biosynthesis	4/38	47/8047	0.00658516	4
hsa00512	Mucin type O-glycan biosynthesis	3/38	32/8047	0.01912408	3
hsa04662	B cell receptor signaling pathway	4/38	82/8047	0.01912408	4
hsa05168	Herpes simplex virus 1 infection	110/885	491/8047	1.30E-11	110
hsa00310	Lysine degradation	22/885	61/8047	3.12E-05	22
hsa03460	Fanconi anemia pathway	20/885	54/8047	3.80E-05	20
hsa03440	Homologous recombination	17/885	41/8047	3.80E-05	17
hsa03040	Spliceosome	37/885	149/8047	7.25E-05	37

GBM patient can be scored and divided into a high prognostic risk group or a low prognostic risk group. It is used to predict the overall survival time of 3 years or 5 years, and the model is proved to be accurate and reliable. And we also made an indirect preliminary analysis of its potential mechanism and found that these lncRNAs may be related to the occurrence and development of tumors.

Acknowledgements

The author thanks all the staff of the second Department of Neurology, the Second Affiliated Hospital of Kunming Medical University. This study was supported by the National Natural Science Foundation of China (Grant No. 81960459).

Disclosure of conflict of interest

None.

Address correspondence to: Ji Luo, Neurosurgery Department, The Second Affiliated Hospital of Kunming Medical University, No. 374, Yunnan Dianmian Avenue, Kunming, Yunnan, China. Tel: +86-087163402467; Fax: +86-087163402467; E-mail: 1426814367@qq.com

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Wen PY and Kesari S. Malignant gliomas in adults. N Engl J Med 2008; 359: 492-507.

[3] Cuddapah VA, Robel S, Watkins S and Sontheimer H. A neurocentric perspective on glioma invasion. Nat Rev Neurosci 2014; 15: 455-65.

[4] Molinaro AM, Taylor JW, Wiencke JK and Wrensch MR. Genetic and molecular epidemiology of adult diffuse glioma. Nat Rev Neurol 2019; 15: 405-417.

[5] Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P and Ellison DW. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 2016; 131: 803-20.

[6] Wang L, Liu Y, Sun S, Lu M and Xia Y. Regulation of neuronal-glial fate specification by long noncoding RNAs. Rev Neurosci 2016; 27: 491-9.

[7] Jarroux J, Morillon A and Pinskaya M. History, discovery, and classification of lncRNAs. Adv Exp Med Biol 2017; 1008: 1-46.

[8] Peng WX, Koirala P and Mo YY. LncRNA-mediated regulation of cell signaling in cancer. Oncogene 2017; 36: 5661-5667.

[9] Bhan A, Soleimani M and Mandal SS. Long noncoding RNA and cancer: a new paradigm. Cancer Res 2017; 77: 3965-3981.

[10] Pan YB, Zhu Y, Zhang QW, Zhang CH, Shao A and Zhang J. Prognostic and predictive value of a long non-coding RNA signature in Glioma: a lncRNA expression analysis. Front Oncol 2020; 10: 1057.

[11] Dahai Z, Daliang C, Famu L, Xiang W, Lenian L, Jianmin C and Xiaobing X. Lowly expressed lncRNA PVT1 suppresses proliferation and advances apoptosis of glioma cells through up-regulating microRNA-128-1-5p and inhibiting PTBP1. Brain Res Bull 2020; 163: 1-13.

[12] Galasso M, Dama P, Previati M, Sandhu S, Palatini J, Coppola V, Warner S, Sana ME, Zanella R, Abujarour R, Desponts C, Teitell MA, Garzon R, Calin G, Croce CM and Volinia S. A large scale expression study associates uc.283-plus lncRNA with pluripotent stem cells and human glioma. Genome Med 2014; 6: 76.
[13] Guardia GDA, Correa BR, Araujo PR, Qiao M, Burns S, Penalva LOF and Galante PAF. Proneural and mesenchymal glioma stem cells display major differences in splicing and lncRNA profiles. NPJ Genom Med 2020; 5: 2.
[14] Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-7.
[15] Hoffmann S. Computational analysis of high throughput sequencing data. Methods Mol Biol 2011; 719: 199-217.
[16] Liu C, Yang G, Liu N, Zhou Z, Cao B, Zhou P and Yang B. Effect of lncRNA BNC2-AS1 on the proliferation, migration and invasion of gastric cancer cells. Clin Lab 2018; 64.
[17] Yang Z and Hu T. Long noncoding RNA HOXC-AS3 facilitates the progression of invasive mucinous adenocarcinomas of the lung via modulating FUS/FOXM1. In Vitro Cell Dev Biol Anim 2020; 56: 15-23.
[18] Shi SH, Jiang J, Zhang W, Sun L, Li XJ, Li C, Ge QD and Zhuang ZG. A novel lncRNA HOXC-AS3 acts as a miR-3922-5p sponge to promote breast cancer metastasis. Cancer Invest 2020; 38: 1-12.
[19] Wang X, Sun Y, Xu T, Qian K, Huang B, Zhang K, Song Z, Qian T, Shi J and Li L. HOXB13 promotes proliferation, migration, and invasion of glioblastoma through transcriptional upregulation of lncRNA HOXC-AS3. J Cell Biochem 2019; 120: 15527-15537.
[20] Li B, Han H, Song S, Fan G, Xu H, Zhou W, Qiu Y, Qian C, Wang Y, Yuan Z, Gao Y, Zhang Y and Zhuang W. HOXC10 regulates osteogenesis of mesenchymal stromal cells through interaction with its natural antisense transcript lncHOXC-AS3. Stem Cells 2019; 37: 247-256.
[21] Zhang E, He X, Zhang C, Su J, Lu X, Si X, Chen J, Yin D, Han L and De W. A novel long noncoding RNA HOXC-AS3 mediates tumorigenesis of gastric cancer by binding to YBX1. Genome Biol 2018; 19: 154.
[22] Piao HY, Guo S, Wang Y and Zhang J. Long noncoding RNA NALT1-induced gastric cancer invasion and metastasis NOTCH signaling pathway. World J Gastroenterol 2019; 25: 6508-6526.
[23] Zheng R, Yao Q, Ren C, Liu Y, Yang H, Xie G, Du S, Yang K and Yuan Y. Upregulation of long noncoding RNA small nucleolar RNA host gene 18 promotes radioresistance of Glioma by repressing semaphorin 5A. Int J Radiat Oncol Biol Phys 2016; 96: 877-887.
[24] Yang B, Miao S, Zhang LN, Sun HB, Xu ZN and Han CS. Correlation of CCNA1 promoter methylation with malignant tumors: a meta-analysis introduction. Biomed Res Int 2015; 2015: 134027.
[25] Klajic J, Busato F, Edvardsen H, Touleimat N, Fleischer T, Bukholm I, Børresen-Dale AL, Lønning PE, Tost J and Kristensen VN. DNA methylation status of key cell-cycle regulators such as CDKNA2/p16 and CCNA1 correlates with treatment response to doxorubicin and 5-fluorouracil in locally advanced breast tumors. Clin Cancer Res 2014; 20: 6357-66.
[26] Chujan S, Kitkumthorn N, Siriangkul S and Mutirangura A. CCNA1 promoter methylation: a potential marker for grading Papanicolaou smear cervical squamous intraepithelial lesions. Asian Pac J Cancer Prev 2014; 15: 7971-5.
[27] Zuo Q, Zheng W, Zhang J, Pan Z, Liu Y, Long H, Fan P, Guo C, Li F and Shao R. Methylation in the promoters of HS3ST2 and CCNA1 genes is associated with cervical cancer in Uygur women in Xinjiang. Int J Biol Markers 2014; 29: e354-62.
[28] Rettori MM, de Carvalho AC, Longo AL, de Oliveira CZ, Kowalski LP, Carvalho AL and Vettore AL. TIMP3 and CCNA1 hypermethylation in HNSCC is associated with an increased incidence of second primary tumors. J Transl Med 2013; 11: 316.
[29] Yanatatsaneejit P, Chalermchai T, Kerekhanjanarong V, Shotelersuk K, Supiyaphun P, Mutirangura A and Sriuranpong V. Promoter hypermethylation of CCNA1, RARRES1, and HRASLS3 in nasopharyngeal carcinoma. Oral Oncol 2008; 44: 400-6.
[30] He L, Fan C, Kapoor A, Ingram AJ, Rybak AP, Austin RC, Dickhout J, Cutz JC, Scholey J and Tang D. α-Mannosidase 2C1 attenuates PTEN function in prostate cancer cells. Nat Commun 2011; 2: 307.
[31] Jiang DD, Liu YQ, Gu B, Xiang ZG, Tian Y, Zhou YQ, Ju JY, Liu Y, Zhang LF and Zhu LP. Effects of human alpha-mannosidase Man2c1 transgene on growth and metastasis of transplanted tumor in mice. Zhongguo Yi Xue Ke Xue Yuan Xue Bao 2007; 29: 528-32.