

Identification and validation of a novel eight mutant-derived long non-coding RNAs signature as a prognostic biomarker for genome instability in low-grade glioma

Aierpati Maimaiti¹, Xixian Wang¹, Yinan Pei¹, Nuerbiye Nuermaiti², Abudireheman Tuersunniyazi¹, Yaeraili Abula¹, Zhaohai Feng¹, Lei Jiang¹, Xin Shi¹, Maimaitijiang Kasimu¹

¹Department of Functional Neurosurgery, Neurosurgery Centre, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang 830054, China

²State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Disease in Central Asia, Department of Biochemistry and Molecular Biology, Basic Medicine College, Xinjiang Medical University, Urumqi, Xinjiang 830054, China

Correspondence to: Maimaitijiang Kasimu; **email:** 13999969488@139.com, <https://orcid.org/0000-0003-4810-8187>

Keywords: genomic instability-associated lncRNAs signature (GILncSig), long non-coding RNA (lncRNA), low-grade glioma (LGG), risk score, prognosis

Received: February 18, 2021

Accepted: May 11, 2021

Published: June 3, 2021

Copyright: © 2021 Maimaiti et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Long non-coding RNAs (lncRNAs) comprise an integral part of the eukaryotic transcriptome. Alongside proteins, lncRNAs modulate lncRNA-based gene signatures of unstable transcripts, play a crucial role as antisense lncRNAs to control intracellular homeostasis and are implicated in tumorigenesis. However, the role of genomic instability-associated lncRNAs in low-grade gliomas (LGG) has not been fully explored. In this study, lncRNAs expression and somatic mutation profiles in low-grade glioma genome were used to identify eight novel mutant-derived genomic instability-associated lncRNAs including H19, FLG-AS1, AC091932.1, AC064875.1, AL138767.3, AC010273.2, AC131097.4 and ISX-AS1. Patients from the LGG gene mutagenome atlas were grouped into training and validation sets to test the performance of the signature. The genomic instability-associated lncRNAs signature (GILncSig) was then validated using multiple external cohorts. A total of 59 novel genomic instability-associated lncRNAs in LGG were used for least absolute shrinkage and selection operator (Lasso), single and multifactor Cox regression analysis using the training set. Furthermore, the independent predictive role of risk features in the training and validation sets were evaluated through survival analysis, receiver operating feature analysis and construction of a nomogram. Patients with IDH1 mutation status were grouped into two different risk groups based on the GILncSig score. The low-risk group showed a relatively higher rate of IDH1 mutations compared with patients in the high-risk group. Furthermore, patients in the low-risk group had better prognosis compared with patients in the high-risk group. In summary, this study reports a reliable prognostic prediction signature and provides a basis for further investigation of the role of lncRNAs on genomic instability. In addition, lncRNAs in the signature can be used as new targets for treatment of LGG.

INTRODUCTION

Low-grade gliomas are a diverse group of heterogeneous brain tumors that originate from glial cells, and their aggressiveness varies depending on subtype and grade.

LGG is a grade II-III glioma based on the World Health Organization classification system, and is different from high-grade glioblastoma (GBM) (grade IV glioma) [1, 2]. LGG and GBM exhibit diverse molecular and clinical features [3]. LGG is more prevalent in younger

patients unlike other tumors with a mean age of 41 years [4]. LGG accounts for approximately 15% of all brain and CNS tumors, with an incidence of approximately 1 in 100,000 [5]. Currently, the conventional treatment approaches of LGG are surgical treatment and postoperative radiotherapy, however, these methods do not prevent transformation from LGG to GBM or recurrence [6]. Therefore, diagnosis at an early stage of low-grade gliomas can improve clinical prognosis of patients. Further, studies should explore effective biomarkers based on the underlying mechanisms of LGG progression for development of personalized treatment approaches.

Genomic instability changes and malignant proliferation are hallmarks for cancer development [7]. Previous studies report that genomic instability plays a key role in cancer prognosis, and high levels of genomic instability are implicated in survival and progression of tumors [8]. Although the underlying molecular-underpinnings of genomic instability have not been fully explored, abnormalities in transcription and post-transcriptional modulation are linked to genomic instability, indicating that molecular features can be used to quantify genomic instability. For instance, a study by Bao et al. [9] analysed 128 gene expression profiles of breast cancer specimens and identified genomic instability signatures for two genes and lncRNA signatures associated with genomic instability and breast cancer outcomes.

Long noncoding RNAs (lncRNAs) are a class of RNA molecules with transcripts longer than 200 nt that do not encode proteins. They modulate expression levels of transcribed genes at multiple layers including epigenetic modulation, transcriptional modulation and post-transcriptional modulation [10]. Several studies have explored various lncRNAs, their roles in transcriptional interference, post-transcriptional gene silencing (PTGS), genomic imprinting and induction of chromatin remodelling and nucleosome modification, regulation of variable splicing patterns, generation of endogenous siRNAs, differentiation, and regulation of cis and transgene expression. LncRNAs play integral roles in development of various diseases. For instance, dysregulation of lncRNAs expression is implicated in tumor proliferation, tumor progression and metastasis. The lncRNA, Norad which is a recently described noncoding RNA, is a Noncoding RNA activated by DNA damage (Norad) [11]. Previous studies developed lncRNA-Norad-deficient mouse models using CRISPR/Cas9 gene-editing technology [12]. In these models, Norad-deficient mice exhibit a multisystem degenerative phenotype significantly resembling premature aging compared to controls. This phenotype is characterized by PUM overactivity and increased repression of genes essential for normal mitosis,

resulting in genomic instability [13]. Notably, NORAD deletion causes severe mitochondrial dysfunction due to upregulation of PUM targets for multiple genes that regulate mitochondrial homeostasis *in vivo* [14]. Previous studies on classification, differentiation and prognosis prediction of gliomas mainly focused on high-grade gliomas or glioblastomas. Therefore, biological markers for prognostic stratification of LGG patients have not been fully explored. The aim of this study was to explore the key role of lncRNAs in maintaining genomic instability in LGG patients.

Genomic stability maintenance is fundamental to all life activities, and DNA damage and replication stress as a result of multiple exogenous and endogenous factors are implicated in genomic instability. A previous study reports that lncRNA NORAD regulates activity of a protein complex composed of RBMX-TOPI and other proteins that prevent genomic instability by binding to RBMX proteins [15]. In addition, lncRNA NORAD enhances genomic stability by separating PUMILIO proteins from their target mRNAs [16]. Genomic instability and the resulting mutagenicity can cause genetic alterations in cancer cells, promoting tumor progression. Studies report that copy and allelic imbalance of two TP53 mutations in MDS patients results in genomic instability which is significantly associated with lower survival rates [17]. Prognosis of gliomas with IDH1/2 mutations is generally good, however, some gliomas exhibiting IDH1/2 mutations have similar prognosis as glioblastomas without IDH1/2 mutations [18, 19]. Most glioblastomas have a large number of gene copy number and sequence variants at the genomic level. Some glioma samples exhibit highly complex karyotypic, gene copy number and sequence variants in addition to shared copies, leading to heterogeneity between and within glioma samples. Although studies have several lncRNAs involved in genomic stability, lncRNAs associated with genomic instability and the resulting clinical implications in low-grade gliomas have not been fully explored.

In this study, we explored lncRNAs associated with genomic instability in LGG based on lncRNA expression and somatic mutation profiles in the genome of low-grade gliomas. Further, we developed risk models for prognosis prediction and development of novel treatment approaches for LGG patients.

RESULTS

Identification of lncRNAs associated to genomic instability in Low-grade glioma patients

The flow chart of our work was shown in Figure 1. Frequency of the cumulative number of somatic mutations

was calculated for each patient and then sorted in descending order to determine lncRNAs associated with genomic instability. Clinical and pathological characteristics of TCGA-LGG patients were used to explore the clinical outcome of these patients and are presented in Table 1. Patients were divided into (GU) and (GS) groups based on the cumulative number of somatic mutations. lncRNA expression profiles of 137 patients in the GS group and 133 patients in the GU group were compared, and lncRNAs with significantly different expression levels were identified. Analysis using SAM method showed that 59 lncRNAs were differentially expressed ($|\log_{2}FC| > 1$, FDR adj p -value < 0.05 , Supplementary Table 1). Unsupervised hierarchical clustering analysis was performed using the 59 differentially expressed lncRNAs for the 529 samples retrieved from TCGA. The 529 samples were divided into two groups based on the levels of the 59 differentially expressed lncRNAs. Analysis showed that the two groups had significantly different somatic mutation patterns (Figure 2A). The GU group showed significantly higher cumulative somatic mutations compared with the GS group. Notably, the GU group showed significantly higher median number of cumulative somatic mutations compared with the GS group ($P < 0.001$) (Figure 2B). In addition, the expression level of UBQLN4 gene [20] (a genomic instability-associated gene overexpressed in aggressive tumors) was compared in the GU and GS groups (Figure 2C).

The top 10 most relevant mRNAs among the differential lncRNAs in the GS and GU groups were screened out for each lncRNA as its target genes. This dataset of lncRNA-related mRNAs was used to

construct a lncRNA-mRNA co-expression network (Figure 2D). GO analysis of lncRNA-related target genes showed that the mRNAs in this network were highly involved in formation and progression of genomic instability, including mitosis, maintenance of synaptic machinery, transporter complexes and activity of various ion channels (Figure 2E and 2F). KEGG pathway analysis of lncRNA-associated target genes identified several signalling pathways associated with genomic instability, tumorigenesis, progression and treatment of low-grade glioma including B cell receptor signalling pathway, T cell receptor signalling pathway, Rap1 signalling pathway, JAK-STAT signalling pathway, PD-L1 expression and PD-1 checkpoint signalling pathway, the Wnt signalling pathway, cAMP signalling pathway and cGMP-PKG signalling pathway (Figure 2G and 2H).

Analysis showed that 59 lncRNAs are implicated in genomic instability. Variations in their expressions may destabilize the cellular genome and disrupt cellular homeostasis of regulatory networks associated with lncRNAs thus inducing various signalling pathways that promote cancer development. The 59 differentially expressed lncRNAs were then referred as genomic instability-associated lncRNAs (GILncRNAs).

Identification of a genomic instability-mutant lncRNA signature for prognosis using the training set

To further explore the predictive prognosis role of the candidate lncRNAs associated with genomic instability, 477 patients with low-grade glioma from the TCGA

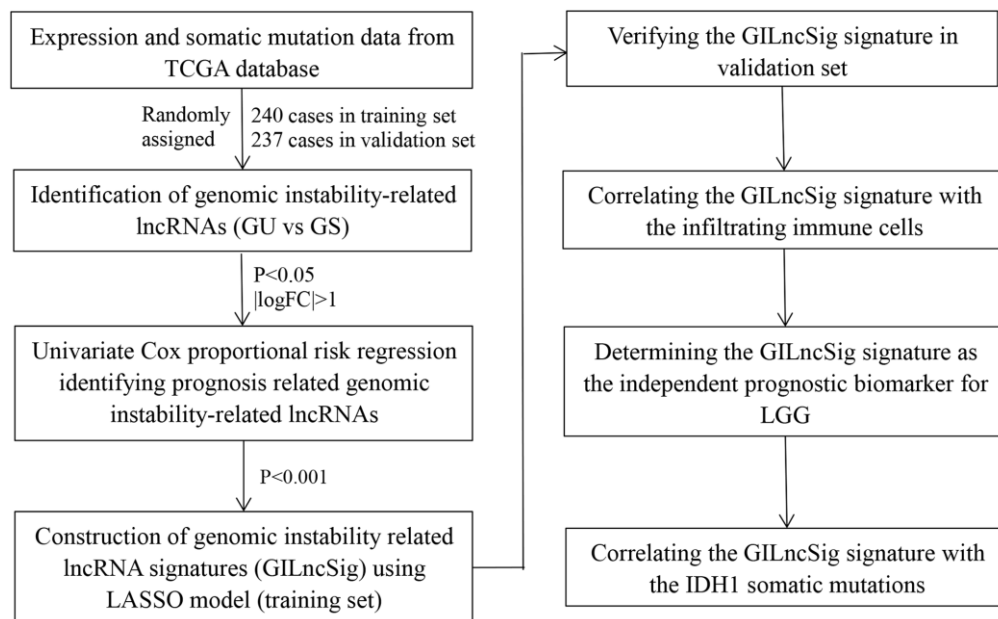


Figure 1. Study flow chart of genomic instability-related lncRNAs construction.

Table 1. Clinical information for three LGG patients sets in this study.

Covariates		Training set (n = 240)	Testing set (n = 237)	TCGA set (n = 477)	P-value
Age	<=41 years	129 (53.75%)	117 (49.37%)	246 (51.57%)	0.3864 ^a
	>41 years	111 (46.25%)	120 (50.63%)	231 (48.43%)	
Gender	female	113 (47.08%)	103 (43.46%)	216 (45.28%)	0.4821 ^a
	male	127 (52.92%)	134 (56.54%)	261 (54.72%)	
First presenting symptom	Headaches	48 (20%)	46 (19.41%)	94 (19.71%)	0.9366 ^b
	Mental Status Changes	19 (7.92%)	19 (8.02%)	38 (7.97%)	
	Motor/Movement Changes	19 (7.92%)	16 (6.75%)	35 (7.34%)	
	Sensory Changes	7 (2.92%)	9 (3.8%)	16 (3.35%)	
	Visual Changes	7 (2.92%)	4 (1.69%)	11 (2.31%)	
	Seizures	116 (48.33%)	116 (48.95%)	232 (48.64%)	
	unknown	24 (10%)	27 (11.39%)	51 (10.69%)	
	unknown	24 (10%)	27 (11.39%)	51 (10.69%)	
First presenting symptom longest duration	0–30 Days	106 (44.17%)	97 (40.93%)	203 (42.56%)	0.7409 ^b
	31–90 Days	33 (13.75%)	39 (16.46%)	72 (15.09%)	
	91–180 Days	19 (7.92%)	15 (6.33%)	34 (7.13%)	
	>181 Days	51 (21.25%)	51 (21.52%)	102 (21.38%)	
	unknown	31 (12.92%)	35 (14.77%)	66 (13.84%)	
Diagnoses	Astrocytoma, anaplastic	58 (24.17%)	63 (26.58%)	121 (25.37%)	0.8172 ^b
	Astrocytoma, NOS	29 (12.08%)	29 (12.24%)	58 (12.16%)	
	Mixed glioma	69 (28.75%)	57 (24.05%)	126 (26.42%)	
	Oligodendroglioma, anaplastic	34 (14.17%)	38 (16.03%)	72 (15.09%)	
	Oligodendroglioma, NOS	50 (20.83%)	50 (21.1%)	100 (20.96%)	
Radiation therapy	YES	137 (57.08%)	135 (56.96%)	272 (57.02%)	0.6421 ^a
	NO	83 (34.58%)	73 (30.8%)	156 (32.7%)	
	unknown	20 (8.33%)	29 (12.24%)	49 (10.27%)	
Seizure history	YES	142 (59.17%)	140 (59.07%)	282 (59.12%)	0.9998 ^a
	NO	83 (34.58%)	82 (34.6%)	165 (34.59%)	
	unknown	15 (6.25%)	15 (6.33%)	30 (6.29%)	
Sample type	Primary Tumor	230 (95.83%)	231 (97.47%)	461 (96.65%)	0.4609 ^a
	Recurrent Tumor	10 (4.17%)	6 (2.53%)	16 (3.35%)	
Grade	G2	124 (51.67%)	107 (45.15%)	231 (48.43%)	0.2069 ^a
	G3	116 (48.33%)	130 (54.85%)	246 (51.57%)	
IDH1 mutation status	Mutant	42 (17.5%)	48 (20.25%)	90 (18.87%)	0.6728 ^a
	Wildtype	18 (7.5%)	16 (6.75%)	34 (7.13%)	
	unknown	180 (75%)	173 (73%)	353 (74%)	

^aChi square test.

^bWilcoxon rank sum test.

project were divided into a training group (n = 240) and a testing group (n = 237). Expression levels of 59 genomic instability-associated lncRNAs were analyzed

using univariate Cox proportional risk regression to identify prognosis associated lncRNAs. Analysis showed 54 genomic instability-associated lncRNAs

significantly correlated with prognosis of low-grade gliomas in the training set ($P < 0.001$). Lasso regression analysis was performed using these lncRNAs to avoid

over-fitting and 16 lncRNAs associated with genomic instability in LGG were identified (Figure 3A). Notably, the optimal value of the penalty parameter was

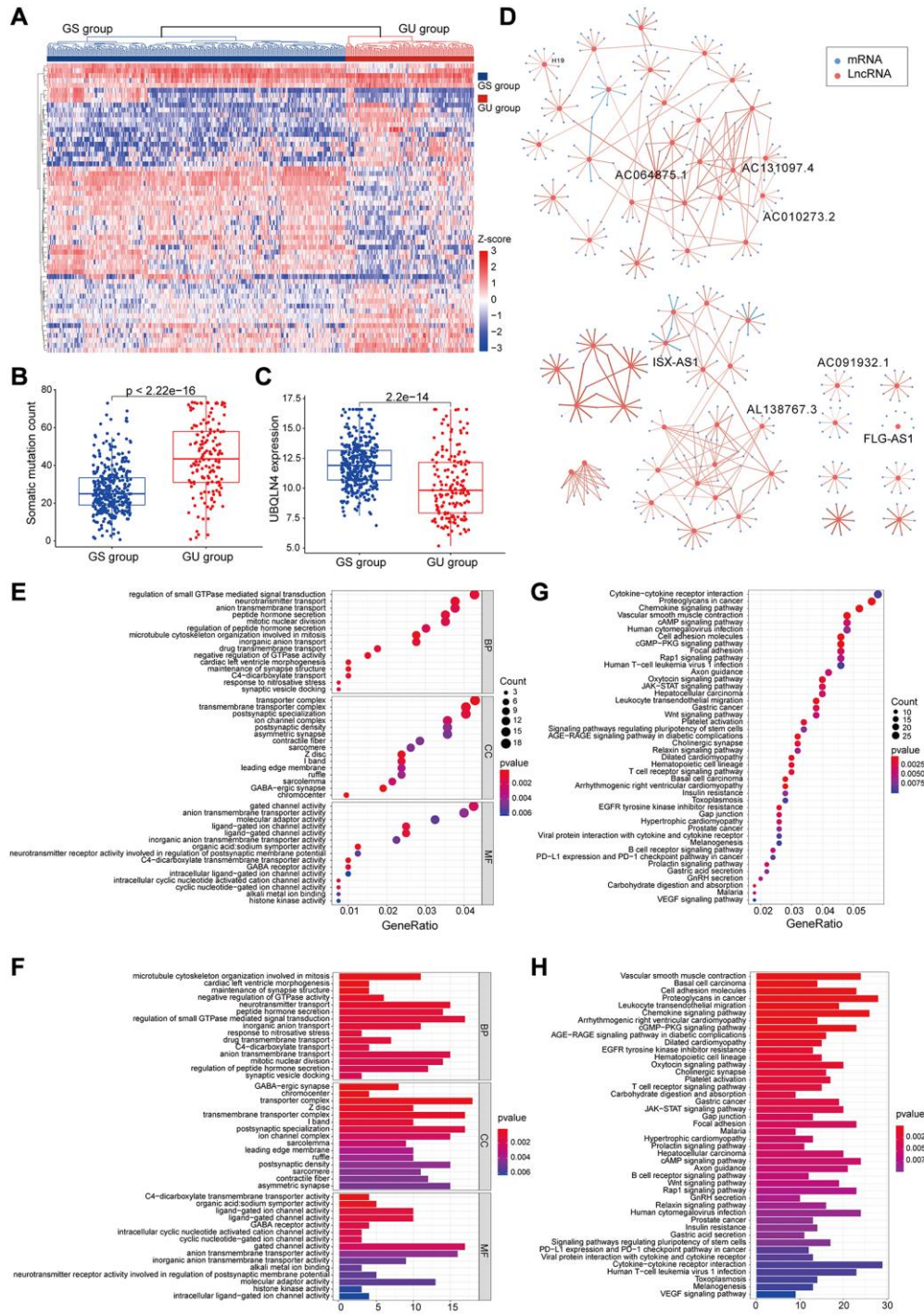


Figure 2. Identified and functionally interpreted genomic instability-associated lncRNAs in patients with low-grade gliomas. (A) An unsupervised clustering among 529 patients with low-grade glioma was performed based on the expression patterns of 59 candidate genomic instability-associated lncRNAs. The GS group is shown in blue on the left, whereas the GU group is shown in red on the right. (B) Box plots for somatic mutations of GS and GU groups. Cumulative somatic mutations in the GU group were significantly higher compared with those in the GS group. (C) Box plots of the expression levels of UBQLN4 in the GU and GS groups. Expression levels of UBQLN4 were significantly lower in the GU group compared with the levels in the GS group. (D) Pearson correlation coefficient analysis based genomic instability-associated lncRNA and mRNA co-expression network. (E–H) GO and KEGG functional enrichment analysis of lncRNA co-expression mRNA through.

determined by performing 1000 replicates of cross-validation (Figure 3B). Further, the 16 candidate LncRNAs were screened to identify those with independent prognostic value. A total of eight genomic instability-associated LncRNAs including H19, FLG-

AS1, AC091932.1, AC064875.1, AL138767.3, AC010273.2, AC131097.4, and ISX-AS1 were identified as independent prognostic risk factors using stepwise regression multi-factor Cox regression analysis identified (Table 2). A GILncSig was constructed using

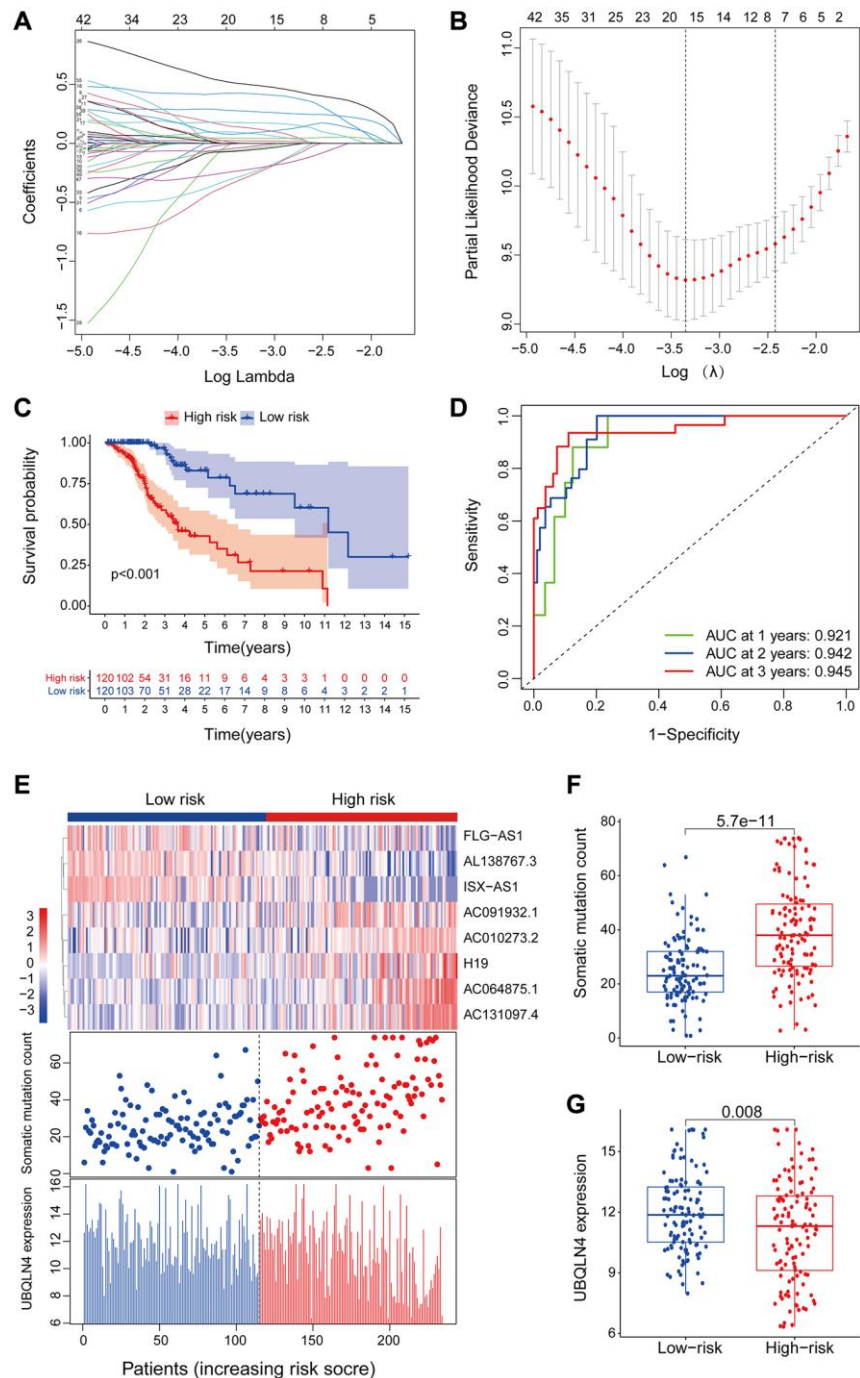


Figure 3. LncRNAs signature of the genomic instability used to predict outcomes in the training set. (A) Lasso Cox analysis identified 16 lncRNAs associated with genomic instability that were highly associated with prognosis. **(B)** Determination of the optimal value of penalty parameters through 1000 replicates of cross-validation. **(C)** Kaplan-Meier estimation of GILncSig-predicted overall survival of low- or high-risk patients in the training set. **(D)** Time-dependent ROC curves of GILncSig at 1, 2 and 3 years. **(E)** Distribution of cumulative somatic mutations and expression of UBQLN4 in high- and low-risk groups in the GILncSig model of low-grade glioma patients. **(F)** Box plot for distribution of cumulative somatic mutations in the low- and high-risk groups of LGG patients. **(G)** Box plots for UBQLN4 gene expression in low- and high-risk groups of LGG patients.

Table 2. Multivariate Cox regression analyses of the 8 of 59 genome instability-related lncRNAs associated with overall survival in LGG.

Ensembl ID	Gene Symbol	Genomic location	Coefficient	HR	95% CI	P-value
ENSG00000130600	H19	chr11:1,995,130–2,001,710	0.029	1.030	1.015–1.045	<0.001
ENSG00000237975	FLG-AS1	chr1:152,168,125–152,445,460	–0.596	0.551	0.276–1.100	0.091
ENSG00000260981	AC091932.1	chr5:8,785,042–8,785,468	0.336	1.399	1.082–1.808	0.010
ENSG00000225649	AC064875.1	chr2:12,780,593–13,007,029	0.369	1.446	1.113–1.879	0.006
ENSG00000225913	AL138767.3	chr10:87,607,985–87,659,383	–0.629	0.533	0.260–1.093	0.086
ENSG00000248664	AC010273.2	chr5:69,113,112–69,159,496	0.691	1.995	1.322–3.012	0.001
ENSG00000235151	AC131097.4	chr2:241,844,380–241,845,036	0.703	2.019	1.258–3.242	0.004
ENSG00000286592	ISX-AS1	chr22:34,756,665–34,997,916	–0.528	0.590	0.399–0.872	0.008

the coefficients of multifactorial cox analysis and expression levels of the eight independent prognosis-associated lncRNAs for assessing the prognostic risk in low-grade glioma patients. the score was calculated using the formula: GILncSig Risk score = [H19*0.0293] + [FLG-AS1*–0.5959] + [AC091932.1*0.3356] + [AC064875.1*0.3690] + [AL138767.3*–0.6290] + [AC010273.2*0.6908] + [AC131097.4*0.7026] + [ISX-AS1*–0.5280].

The risk score for an individual invalid in the training set was obtained using GILncSig and LGG patients were then grouped into low- and high-risk groups using the median risk score as the threshold. Kaplan-Meier analysis showed that overall survival of patients in the low-risk group was significantly higher compared with that of patients in the high-risk group. This finding indicates that the prognostic biomarker for predicting risk scores was accurate ($P < 0.001$; Gehan-Breslow-Wilcoxon test: $P = 4.0207E-9$) (Figure 3C). Survival analysis curves showed that the three-year survival rate among LGG patients with high-risk LGG was approximately 56.8% at 95% CI confidence interval of [46.41%–69.6%], and the three-year survival rate for patients in the low-risk LGG group was approximately 94.9% at 95% CI confidence interval of [89.3%–100%]. Furthermore, the five-year survival rate for patients with high-risk LGG was approximately 39.0% at 95% CI confidence interval of [27.2%–55.9%], and the five-year survival rate for patients with low-risk LGG was approximately 78.6%, at 95% CI confidence interval [66.3%–93.3%]. The ten-year survival rate for patients with high-risk LGG was approximately 10.7%, at 95% CI confidence interval [2.25%–50.8%], whereas the low-risk LGG patients had a ten-year survival rate of approximately 60.2% at 95% CI confidence interval of [41.8%–86.6%]. ROC curve analysis of GILncSig over time for 1, 2, and 3 years showed an area under the curve of 0.921, 0.942, and 0.945, respectively (Figure 3D).

Patients in the training set were grouped based on the score and expression level of GILncSig was determined to represent the somatic mutation counts of the patients. A risk heat map, mutation scatters plot and gene expression map were generated to show the relationship between the risk score and gene expression levels of each LGG sample (Figure 3E). The heat map of the expression profiles of the 8 lncRNAs showed that lncRNA FLG-AS1, AL138767.3, and ISX-AS1 were significantly highly expressed in the low-risk group compared with the levels in the high-risk group. On the other hand, lncRNA H19, AC091932.1, AC064875.1, AC010273.2, and AC131097.4 were significantly highly expressed in the high-risk group compared with the low-risk group. Comparative analysis showed significant differences in somatic mutation patterns and UBQLN4 gene expression patterns between invalids in the low- and high-risk groups.

High-risk group patients showed a higher number of somatic mutations compared with the patients in the low-risk group ($P < 0.001$, Figure 3F). Interestingly, UBQLN4 expression level was significantly higher in low-risk patients compared with the level in high-risk patients ($P = 0.008$, Figure 3G).

Validation of a genomic instability-mutant lncRNA signature for prognosis using the test set and TCGA set

The GILncSig model was validated using a testing set of 237 patients to test its prognostic performance. Using the same GILncSig and risk thresholds as the training set, the 237 patients in the testing set were grouped into a low-risk group ($n = 120$) and high-risk group ($n = 117$). Kaplan-Meier curves showed that the high-risk group sample had significantly lower OS compared with the low-risk group ($p < 0.001$; Gehan-Breslow-Wilcoxon test: $P = 2.0904E-9$) (Figure 4A). Survival analysis curves showed that the 3- and 5-year survival

rates of LGG patients in the high-risk group were approximately 53.8% and 31.1%, at 95% CI confidence intervals of [42.42%–68.3%] and [18.65%–51.8%], respectively. The 3-, 5-, and 10-year survival rates of LGG patients in the low-risk LGG group were approximately 94.1%, 81.7%, and 48.4% at 95% CI

confidence intervals of [88.49%–100%], [69.97%–95.4%] and [29.00%–80.6%], respectively. Analysis of the 1-year, 2-year, and 3-year ROC curves for GILncSig in the test group over time showed AUC values of 0.920, 0.909, and 0.886, respectively (Figure 4B). Expression of GILncSig and somatic mutation counts in

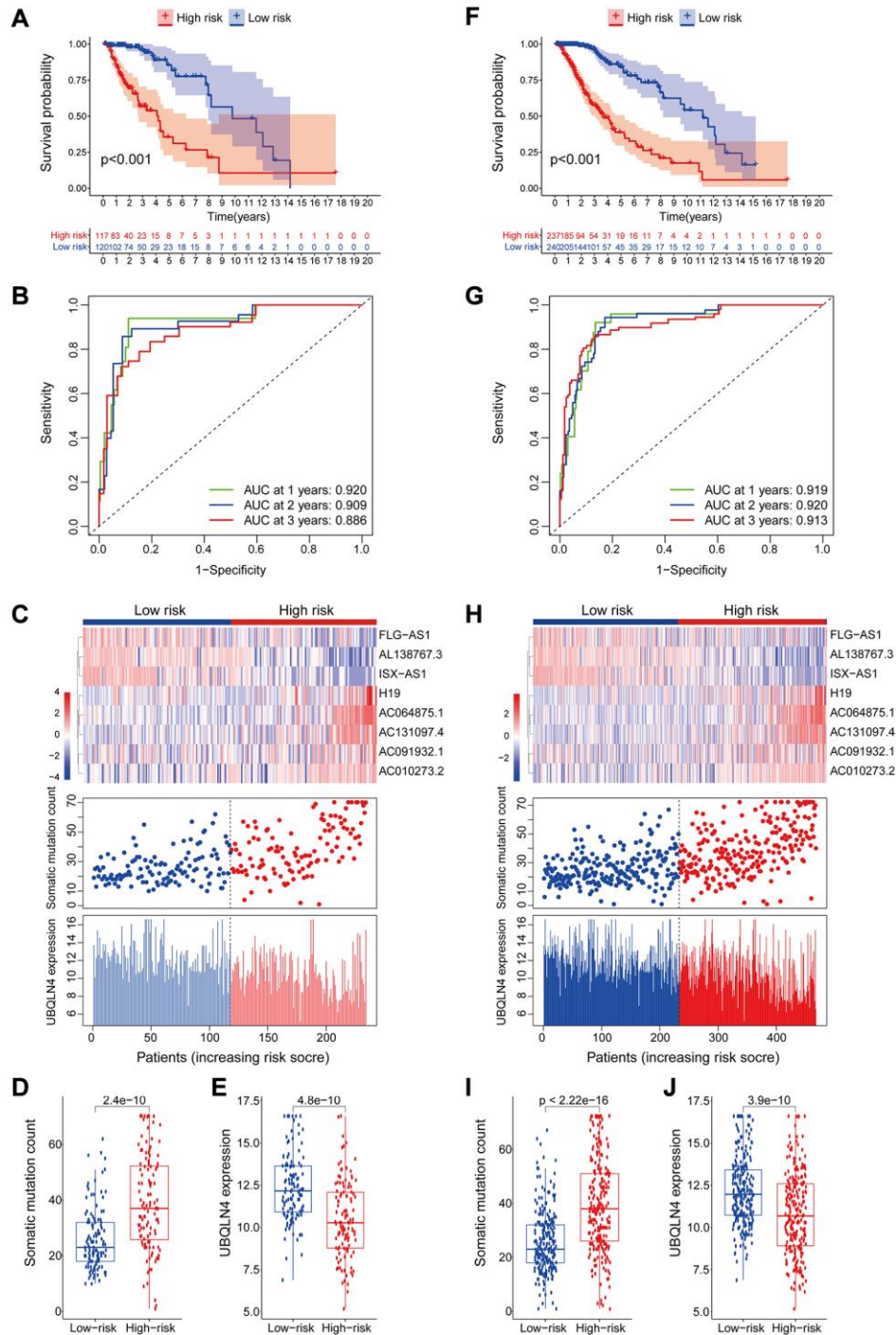


Figure 4. Validation of the lncRNA signature for genomic instability used to predict outcomes in the testing and TCGA set. (A) Validation of overall survival in low- or high-risk patients predicted by pooling GILncSig with Kaplan-Meier estimates. (B) Time-dependent ROC curves of GILncSig at 1, 2 and 3 years in the testing group. (C) Verification of lncRNA expression patterns, the profile of somatic mutations and UBQLN4 expression in patients in low- and high-risk groups. (D–E) Box plots for the distribution of somatic mutations and UBQLN4 expression in high- and low-risk groups of patients. (F–J) Verification of the above results using the TCGA set.

the test samples are presented in Figure 4C. Significant differences in somatic mutation counts were observed between patients in the low- and high-risk groups ($p < 0.001$, Figure 4D). The high-risk group showed significantly lower expression levels of UBQLN4 compared with the low-risk group ($p < 0.001$, Figure 4E).

GILncSig's prognosis in the TCGA group was similar to the findings of the test group. Patients in the TCGA group were divided into low- and high-risk groups. Overall survival analysis showed that the overall survival time of patients in the high-risk group was significantly less compared with that of the low-risk group ($p < 0.001$, Figure 4F). The survival rates at three, five and ten years for patients with high-risk LGG were approximately 56.65%, 36.74%, and 11.64%, at 95% CI confidence intervals of [48.85%–65.7%], [27.57%–49.0%], and [4.21%–32.2%], respectively. The survival rates at 3, 5, and 10 years for patients with low-risk LGG were approximately 95.5%, at 95% CI confidence intervals of [91.94%–99.2%], [73.98%–91.2%], and [39.74%–73.8%], respectively. The 1-year, 2-year, and 3-year ROC curve analysis of the TCGA group over time showed AUC values at 0.919, 0.920, and 0.913, respectively (Figure 4G). GILncSig expression levels and somatic mutation count in TCGA samples are presented in Figure 4H. Analysis showed a significant difference in somatic mutation pattern between patients in the high- and low-risk groups ($p < 0.001$, Figure 4I). The high-risk group showed significantly lower expression levels of UBQLN4 compared with those for the low-risk group ($p < 0.001$, Figure 4J).

Construction of 8 genomic instability-mutant lncRNAs signature and correlation with clinical features for Lower-grade glioma

Correlation between overall survival and the eight genomic instability-associated lncRNA models was determined using Kaplan-Meier curves using the training set (Figure 5A). Analysis showed that lncRNA H19, AC091932.1, AC064875.1, AC010273.2, and AC131097.4 were negatively correlated with overall survival, indicating that they are risk factors for low-grade glioma, therefore, high expression levels of the lncRNAs are associated with poor prognosis. On the contrary, lncRNA FLG-AS1, AL138767.3, and ISX-AS1 were positively correlated with overall survival ($p < 0.001$). These findings show that lncRNA FLG-AS1, AL138767.3, and ISX-AS1 play a protective role in low-grade glioma. Similar findings were obtained through validation using the validation set (Figure 5B). A nomogram was constructed based on tumor grade, gender, age and the eight genomic instability-related lncRNA risk scores in the training group using multivariate Cox regression results (Figure 5C). Analysis of the nomogram,

showed that the prognosis-related factors can be used to accurately predict 1-, 3- and 5-year survival of patients. The consistency index (C-index) was used for validation to evaluate the predictive value of the nomogram. Analysis showed that the C-index of the nomogram was 0.8693 and the 95% CI confidence interval of the C-index was [0.8356–0.9030]. The calibration curve (Figure 5D) and clinical decision curve (Figure 5E) showed that the nomogram had a good predictive effect.

Predictive performance of the nomogram was significantly higher compared to that of the risk score model. Similar results were obtained using the validation set (Figure 5F–5H).

The 8 genomic instability-associated lncRNAs were correlated with infiltration of immune cell subtypes in LGG

To explore the relationship between the expression levels of eight genomic instability-associated lncRNA models and the tumor microenvironment, the abundance of tumor-infiltrating immune subpopulations was determined using the CiberSort algorithm. Furthermore, the correlation between the risk scores of the eight lncRNA-constructed models and infiltration of immune cell subtypes in LGG was analyzed using the training set. The correlation coefficients between macrophages M0, macrophages M1, memory resting CD4 T cells, and CD8 T cells and risk scores were 0.23, 0.26, 0.27, and 0.22, respectively ($P < 0.001$, Figure 6A). These findings indicate that infiltration of tumor immune cells is positively correlated with the risk scores of the eight genomic instability-associated lncRNAs. Activated mast cells ($R = -0.21$, $p < 0.05$) and monocytes ($R = -0.29$, $p < 0.001$) were negatively correlated with risk scores of the eight lncRNAs. In addition, verification using the validation set showed that macrophages M1 ($R = 0.31$, $p < 0.001$) and memory resting CD4 T cells ($R = 0.29$, $p < 0.001$) were positively correlated with the risk scores of the eight genomic instability-associated lncRNAs. On the other hand, analysis using the validation set showed that activated mast cells ($R = -0.27$, $p < 0.001$) and monocytes ($R = -0.2$, $p < 0.05$) immune cell subtypes were negatively correlated with risk scores of the eight genomic instability-associated lncRNAs (Figure 6B). The above correlations between GILncSig signature and the infiltrating immune cells were confirmed with xCell platform in both the training set and validation set (Figure 7A–7B).

Verification of the genomic instability-mutant lncRNA signature model genes using two external independent LGG datasets

Cross-platform validation of genomic instability-associated lncRNA models using independent datasets

from different platforms showed that only four of the eight lncRNAs in GILncSig (AC064875.1, AC131097.4, FLG-AS1, H19) retrieved from CGGA mRNA-seq-693 represented a large sample size and were matched clinicopathological features. Therefore, we explored the relationship between lncRNA AC064875.1, AC131097.4, FLG-AS1, H19 and low-

grade glioma and genomic instability using the CGGA mRNA-seq-693 dataset. Analysis showed that expression levels of AC064875.1 and H19 were significantly correlated with age (≤ 41 and >41 years), tumor grade, IDH1 mutation status and chromosome 1p19q joint deletion ($p < 0.05$), however, expression levels of these lncRNAs were not correlated with

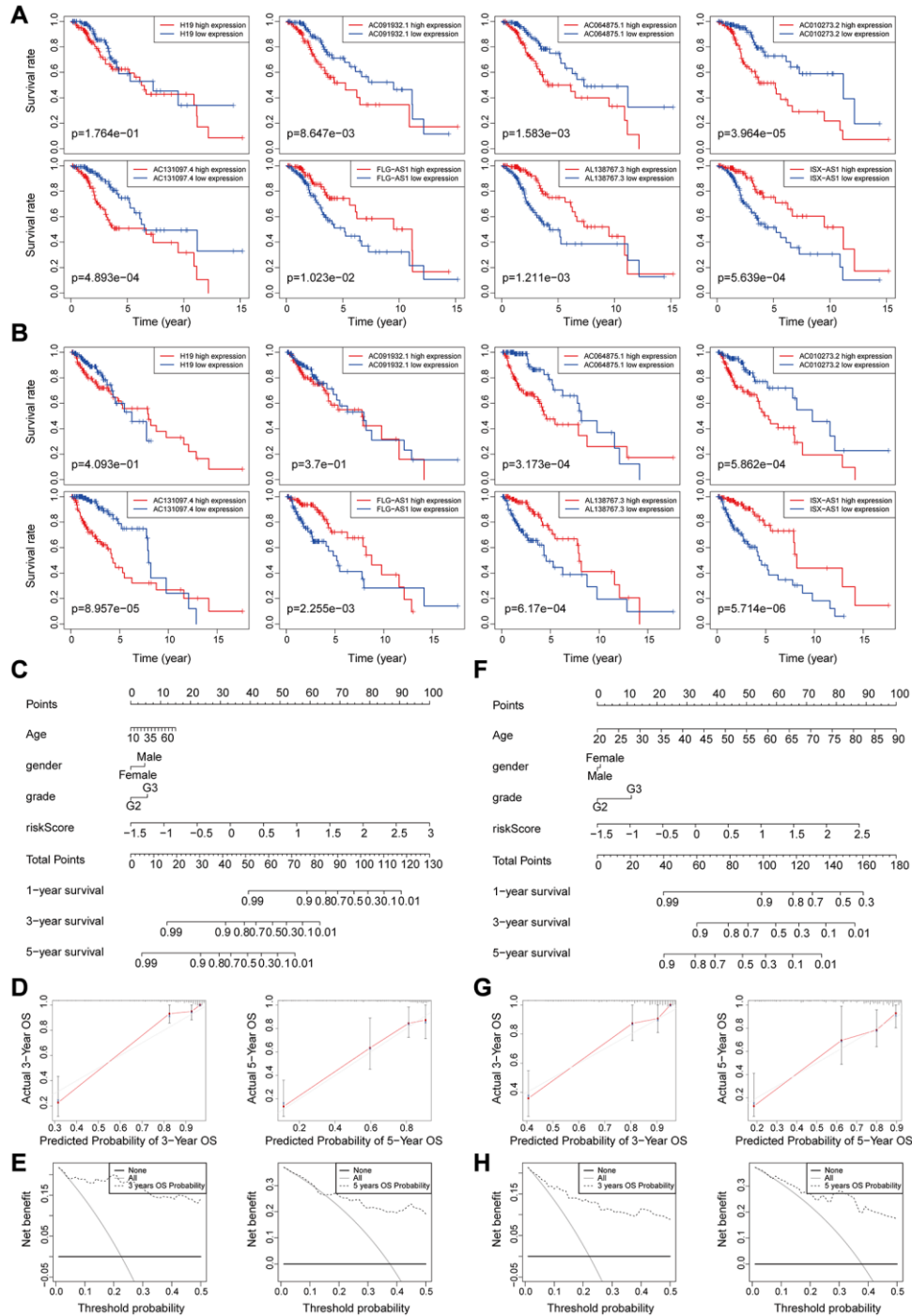


Figure 5. (A–B) Kaplan-Meier survival curves analysis for the eight genomic instability-associated lncRNAs using the training set (A) and validation set (B) of patients with low-grade glioma. (C–E) Nomograms for the eight genomic instability-associated lncRNAs for each factor in the training set, predictions of patient survival at 1, 3, and 5 years. Nomograms were evaluated using calibration curves and DCA curves. (F–H) Plot nomogram plots in the validation set and evaluation of nomograms using calibration curves and DCA curves.

gender and MGMT methylation status (Figure 8A and 8B). AC131097.4 expression levels were significantly different between age subgroups ($p < 0.05$, Figure 8C). Moreover, FLG-AS1 was significantly correlated with gender and tumor grade (especially between grade II vs IV, III vs IV) ($p < 0.05$, Figure 8D).

Similar findings were obtained using the GSE16011 dataset for lncRNA H19. Expression level of H19 was significantly correlated with tumor grade and IDH1(R132) mutation status ($p < 0.05$, Figure 8E). Survival analysis was performed using lncRNA in the model obtained from CGGA mRNA-seq-693 and GSE16011 datasets to determine their correlation with prognosis. In the CGGA mRNA-seq-693 datasets, the OS of patients with high expression of AC064875.1 and H19 was significantly lower compared with that of patients with low expression levels of AC064875.1 and H19 ($p < 0.001$, Supplementary Figure 1A–1B). On the contrary, high expression level of FLG-AS1 was correlated with a higher overall survival compared with low expression levels of FLG-AS1 ($p < 0.05$,

Supplementary Figure 1C). In GSE16011 dataset, patients with low H19 expression level showed significantly higher overall survival compared with the OS of patients with high expression level of H19 ($p < 0.001$, Supplementary Figure 1D). Notably, these findings were consistent with the findings obtained using the TCGA training group and test group.

Genomic instability-associated lncRNA signature performance compared with existing lncRNA-related signatures in survival prediction

The predictive performance of the genomic instability-associated lncRNA model was compared with three glioma-associated lncRNA signatures reported in previous studies (Table 3). The three signatures included a 6-lncRNA signature reported by Lin's study [21] (hereafter referred to as LinlncSig) and the 8-lncRNA signature from Li's study [22] (hereafter referred to as LilncSig), which used the same cohort of TCGA patients. The third study was a 3-lncRNA signature reported by Qiu's study [23] (hereafter

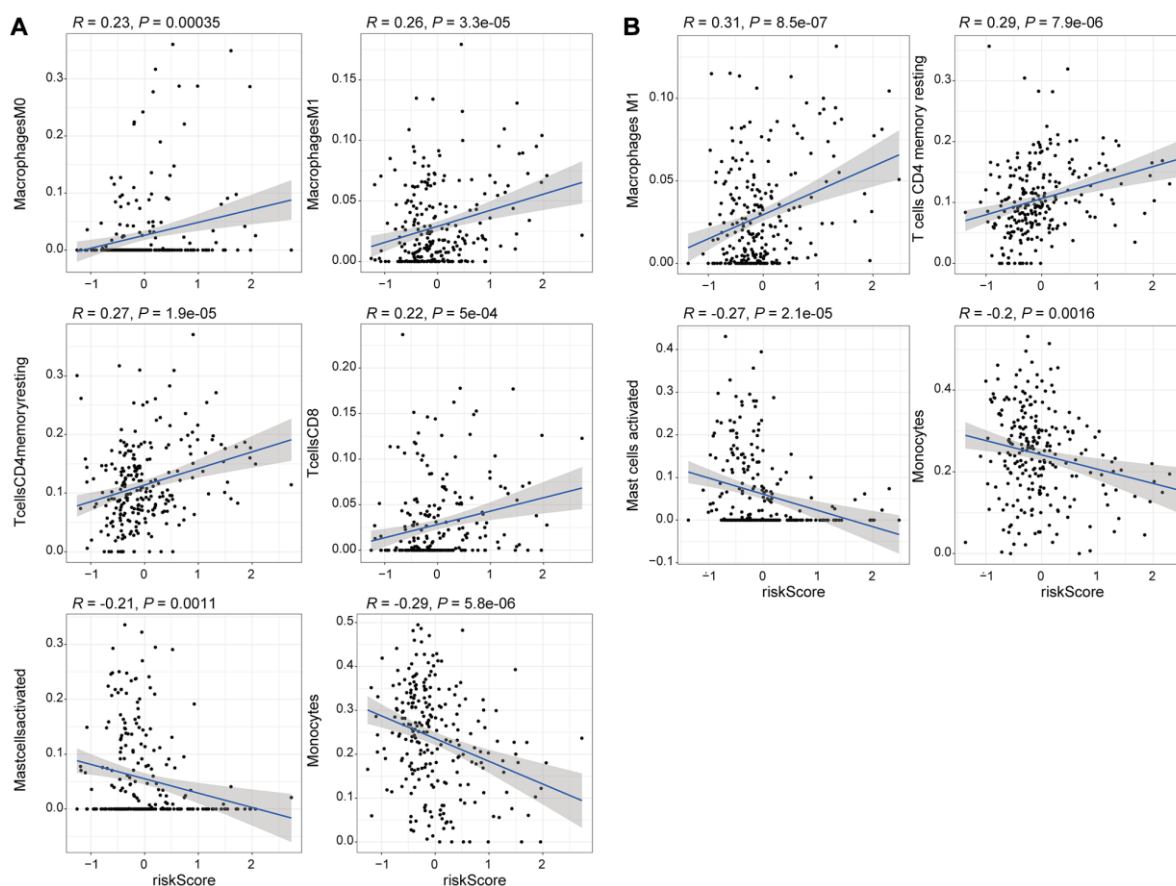


Figure 6. Correlation analysis of the 8 genomic instability-associated lncRNAs with infiltration of each subtype of immune cells. (A) Correlation coefficients for M0 macrophages, M1 macrophages, memory resting CD4 T cells, and CD8 T cells for the training set were 0.23, 0.26, 0.27, 0.22 ($p < 0.001$), whereas the correlation coefficient for activated mast cells was $R = -0.21$ ($p < 0.05$) and the correlation coefficient for monocytes was $R = -0.29$ ($p < 0.001$). **(B)** The correlation coefficients among the risk scores for M1 macrophages, memory resting CD4 T cells, activated mast cells and monocytes for the validation set were 0.31, 0.29, -0.27 and -0.2 , respectively ($p < 0.05$).

referred to as QiuLncSig) which used 167 TCGA-LGG patients with radiological response information. The AUC of GILncSig for one-year OS was 0.919, which was significantly higher compared with the AUC for LinLncSig (AUC = 0.854), LilncSig (AUC = 0.796) and QiuLncSig (AUC = 0.833) (Figure 9A). The AUC of GILncSig for three years OS was 0.913 which was significantly higher compared with the AUC for LinLncSig (AUC = 0.775), LilncSig (AUC = 0.769) and

QiuLncSig (AUC = 0.760) (Figure 9B). Further, the AUC of GILncSig for five years OS was 0.851 which was significantly higher compared with the AUC for LinLncSig (AUC = 0.699), LilncSig (AUC = 0.644) and QiuLncSig (AUC = 0.757) (Figure 9C). We turned the GILncSig, LinLncSig, LiLncSig and QiuLncSig signature into the dichotomous variables in order to combine the signatures, which was named as CombinedSig signature. As the ROC curves showed

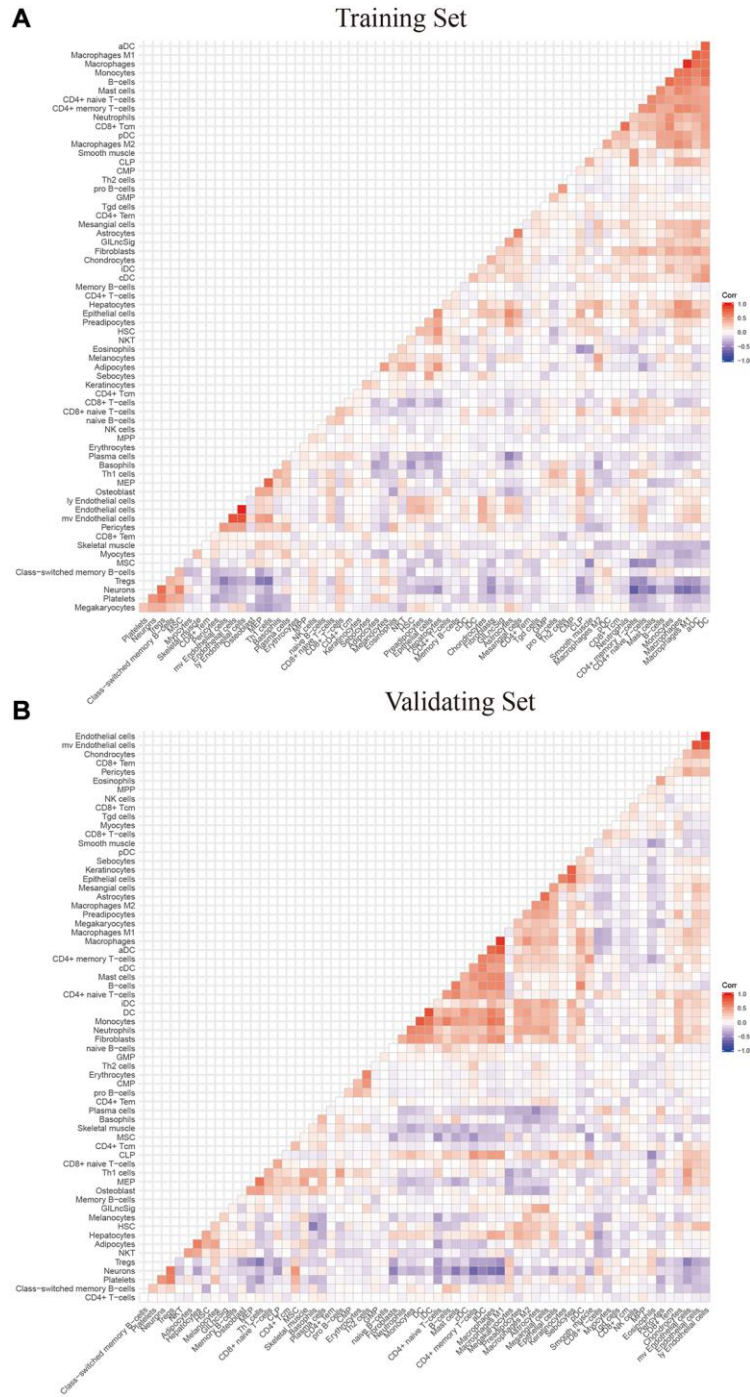


Figure 7. The correlation between GILncSig signature and 64 microenvironment infiltrating immune cells using xcell platform (A) training set (B) validation set.

(Supplementary Figure 2A–2C), we found the CombinedSig signature could increase the prognostic power for 1 year and 3 year survival rate compared to other four independent signatures. Although the combined signature could not increase the predictive power for 5 year rate, the AUC values of the GILncSig signature we constructed in this study were still higher than the other three recently published lncRNA

prognostic signature as well as CombinedSig signature, indicating that the GILncSig signature had better prognostic performance in predicting survival in LGG patients. We also compared the GILncSig signature with the other clinical factors including Age, Gender, IDH status and Grade. The results showed that whether the prediction of 1 year, 3 year and 5 year, the power of GILncSig signature for predicting prognosis was better

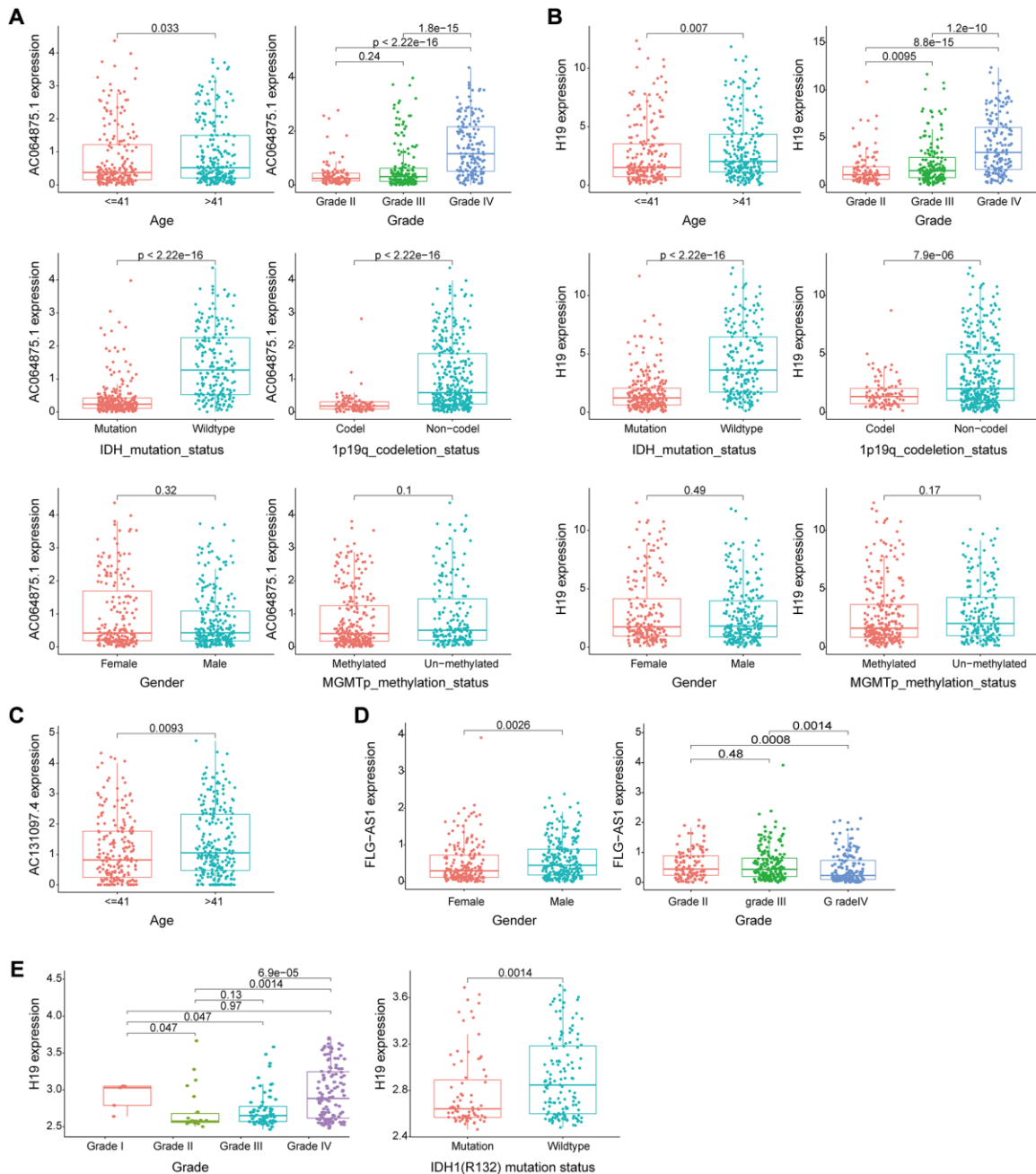


Figure 8. Evaluation of the performance of the GILncSig partial gene using two external independent CGGA mRNA-seq-693 and GSE16011 datasets. (A–B) Box plots for gene expression levels of AC064875.1 and H19 for patients at different ages (< = 41 and > 41 years), tumor grade, IDH1 mutation status, 1p19q chromosome union deletion status, gender and MGMT methylation status in patients from the CGGA mRNA-seq-693 set. **(C)** Box plots for expression of AC131097.4 for patients of different ages in the CGGA mRNA-seq-693 set. **(D)** Box plots for gene expression levels of FLG-AS1 for patients of different tumor grades and gender in the CGGA mRNA-seq-693 set. **(E).** Box plots expression level of lncRNA H19 in patients with different tumor grades and IDH1(R132) mutation status in the GSE16011 dataset.

Table 3. Survival predictive value of the GILncSig, LinlncSig, LilncSig and QiuLncSig.

	AUC of 1-year	AUC of 3-year	AUC of 5-year
GILncSig	0.919	0.913	0.851
LinlncSig	0.854	0.775	0.699
LilncSig	0.796	0.769	0.644
QiuLncSig	0.833	0.760	0.757

Abbreviation: AUC: area under curve.

than the Age, Gender, IDH status and Grade factors (Supplementary Figure 2D–2F). These findings indicate that GILncSig had significantly higher prognostic performance for survival prediction compared with the three recently published lncRNA markers. The sensitivity and specificity of each GILncSig cut-off value for predicting 5-year survival were shown in Supplementary Table 2.

Identification of the independence of genomic instability-associated lncRNAs signature in prognostic prediction

A multivariate Cox regression analysis for age, gender, tumor grade, tumor type (primary and recurrent), LGG

diagnostic types and GILncSig risk score was performed to see if the GILncSig signature was an independent prognostic predictor. Multivariate analysis showed that GILncSig was significantly correlated with OS in each group in the training set, validation set and TCGA set ($p < 0.001$). In the multivariate analysis, two other clinical factors, namely age and diagnostic types were significant ($p < 0.05$) in addition to GILncSig (Table 4 and Figure 10A–10D).

Therefore, we performed a stratified analysis to determine whether GILncSig prognostic value was independent of age and tumor grade. Patients in the TCGA set were grouped as a young ≤ 41 years group ($n = 246$) and an age > 41 years group ($n = 231$) based

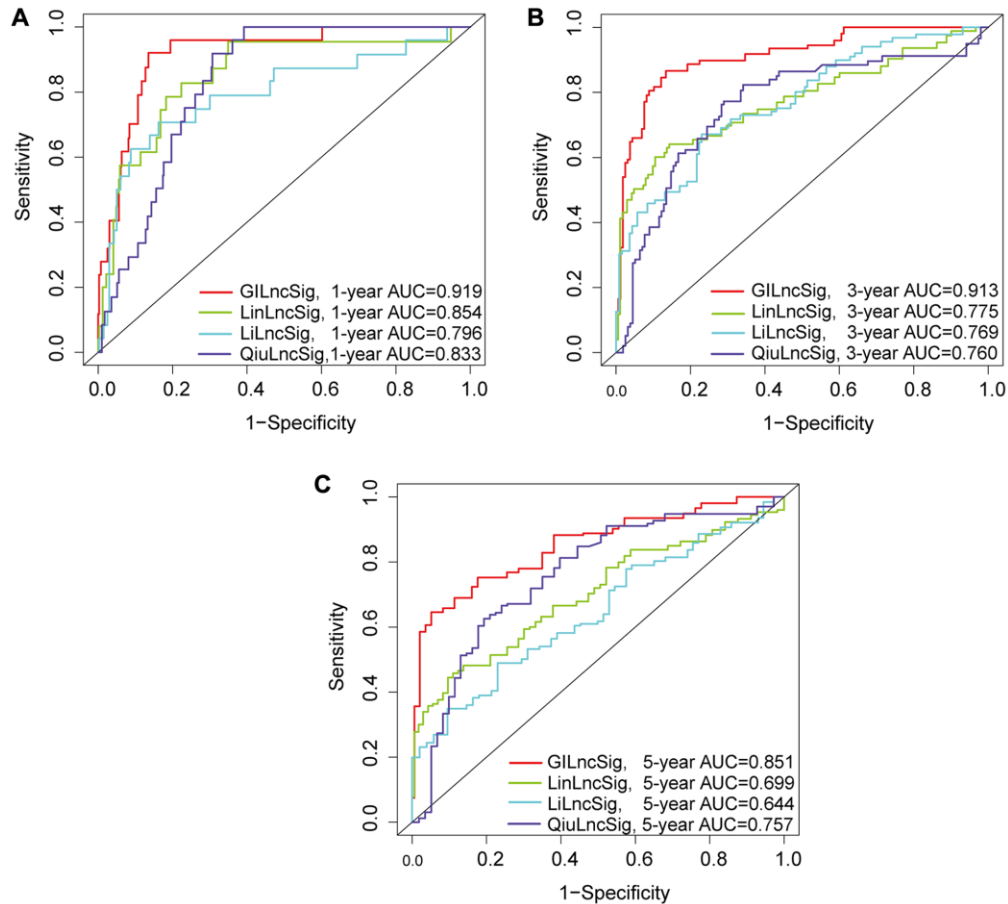


Figure 9. (A–C) ROC analysis of overall survival at 1-, 2- and 3- years for GILncSig, LilncSig, LinlncSig and QiuLncSig.

Table 4. Univariate and Multivariate Cox regression analysis of the GILncSig and overall survival in different LGG patient sets.

Variables		Univariable model			Multivariable model		
		HR	95%CI	P-value	HR	95%CI	P-value
Training set (<i>n</i> = 240)							
GILncSig	High/Low	1.010	1.006–1.014	<0.001	1.01	1.00–1.01	<0.001
Age		1.047	1.027–1.068	<0.001	1.05	1.02–1.07	<0.001
Gender	Female/Male	1.427	0.855–2.383	0.174	1.11	0.65–1.88	0.708
Grade	G2/G3	3.510	2.016–6.110	<0.001	2.25	1.23–4.13	0.008
Diagnoses.type		0.715	0.589–0.867	<0.001	0.74	0.60–0.92	0.006
Primary/Recurrent		1.070	0.449–2.550	0.879	1.94	0.76–4.98	0.168
Testing set (<i>n</i> = 237)							
GILncSig	High/Low	1.013	1.009–1.017	<0.001	1.01	1.00–1.01	0.002
Age		1.068	1.046–1.090	<0.001	1.08	1.06–1.11	<0.001
Gender	Female/Male	0.783	0.472–1.299	0.343	0.99	0.59–1.68	0.983
Grade	G2/G3	3.181	1.808–5.596	<0.001	1.75	0.90–3.43	0.101
Diagnoses.type		0.764	0.647–0.902	0.001	0.74	0.61–0.91	0.004
Primary/Recurrent		1.728	0.621–4.809	0.295	5.23	1.71–15.93	0.004
TCGA set (<i>n</i> = 477)							
GILncSig	High/Low	1.011	1.008–1.014	<0.001	1.01	1.00–1.01	<0.001
Age		1.057	1.042–1.073	<0.001	1.06	1.04–1.08	<0.001
Gender	Female/Male	1.066	0.746–1.524	0.726	1.08	0.75–1.56	0.679
Grade	G2/G3	3.338	2.254–4.944	<0.001	1.95	1.25–3.02	0.003
Diagnoses.type		0.745	0.658–0.844	<0.001	0.76	0.66–0.87	<0.001
Primary/Recurrent		1.233	0.642–2.368	0.529	2.80	1.42–5.55	0.003

on the median age (41 years). Each age group of patients was further classified as a low- or high-risk group based on GILncSig. The OS of low- and high-risk groups in the young ≤ 41 years group was significantly different ($p < 0.001$, Figure 11A). In addition, the OS in the age > 41 years group was statistically significant for the low- and high-risk groups ($p < 0.001$) (Figure 11B). All patients with low-grade glioma were then stratified by tumor grade, and patients in the TCGA set were stratified into Grade II ($n = 231$) and Grade III groups ($n = 246$). Patients with pathological grade of Grade II were further grouped into high- and low- groups based on GILncSig score. Analysis showed significant differences in OS between the high- and low- groups ($p < 0.001$, Figure 11C). Furthermore, GILncSig was used to classify pathologically graded Class III patients into high- and low-risk groups, and analysis showed significant difference in OS between the two groups ($p < 0.001$, Figure 11D). These findings indicate that GILncSig is an independent prognostic factor in predicting the overall survival in patients with low-grade gliomas.

Identification of the relationship between genomic instability-associated lncRNAs signature and IDH1 somatic mutations

Chi-square test analysis showed that a significantly higher proportion of patients in the low-risk group possessed IDH1 mutant phenotype compared with the proportion in the high-risk group of the training set, testing set, and TCGA set. In the training set, 90% of patients in the low-risk group had an IDH1 mutation, which was significantly higher compared with the proportion in the high-risk group (63%) ($p < 0.001$, Figure 12A). In the testing set, 92% of patients in the low-risk group had the IDH1 mutation, which was significantly higher compared with the proportion in the high-risk group (63%) ($p < 0.001$, Figure 12B). In the TCGA group, a significantly lower number of patients in the high-risk group (63%) had the IDH1 mutation compared with the number of patients with the IDH1 mutation in the low-risk group (91%) ($p < 0.001$, Figure 12C). These findings indicate that GILncSig is associated with IDH1 mutation status and can be used as a

mutational marker for the IDH1 gene. The IDH1 gene mutant is associated with high genomic stability, whereas an increase in the number of wild-type IDH1 genes induces genomic instability, impairs non-homologous end-joining DNA repair, and increases susceptibility to DNA damage [24, 25]. This implies that IDH1 wild type effect on genomic instability can be used to improve the clinical prognosis of low-grade gliomas.

Previous studies report that IDH1 mutations are linked to more prolonged survival and may serve as an independent prognosis biomarker for low-grade gliomas [26, 27]. Therefore, we compared the prognosis performance of GILncSig on patient outcome with prediction of outcome using the IDH1 mutation status. The log-ranch test was used to group the patients into four groups, namely IDH1 Mutation/GS group, IDH1 Mutation/GU group, IDH1 Wild/GS group, and IDH1 Wild/GU group ($p < 0.001$). Use of GILncSig for patients with IDH1 wild-type and IDH mutation showed

longer overall survival in the IDH1 Mutation/GU group compared with that for the IDH1 Wild/GU group (Figure 12D). Interestingly, the IDH1 Wild/GS group showed a higher OS rate and had a better outcome compared with that of the IDH1 Mutation/GS group. These findings indicate that GILncSig combined with IDH1 mutation status has greater prognostic significance compared with use of IDH1 mutation status alone.

DISCUSSION

Low-grade gliomas are complex, heterogeneous intracranial tumors associated with multiple genetic mutations, epigenetic alterations, chromosomal deletions, amplifications, and ectopics. Current personalized treatment for low-grade glioblastoma, comprising surgical treatment combined with radiotherapy with/without temozolomide, increases the two-year survival rate of patients to approximately 27%

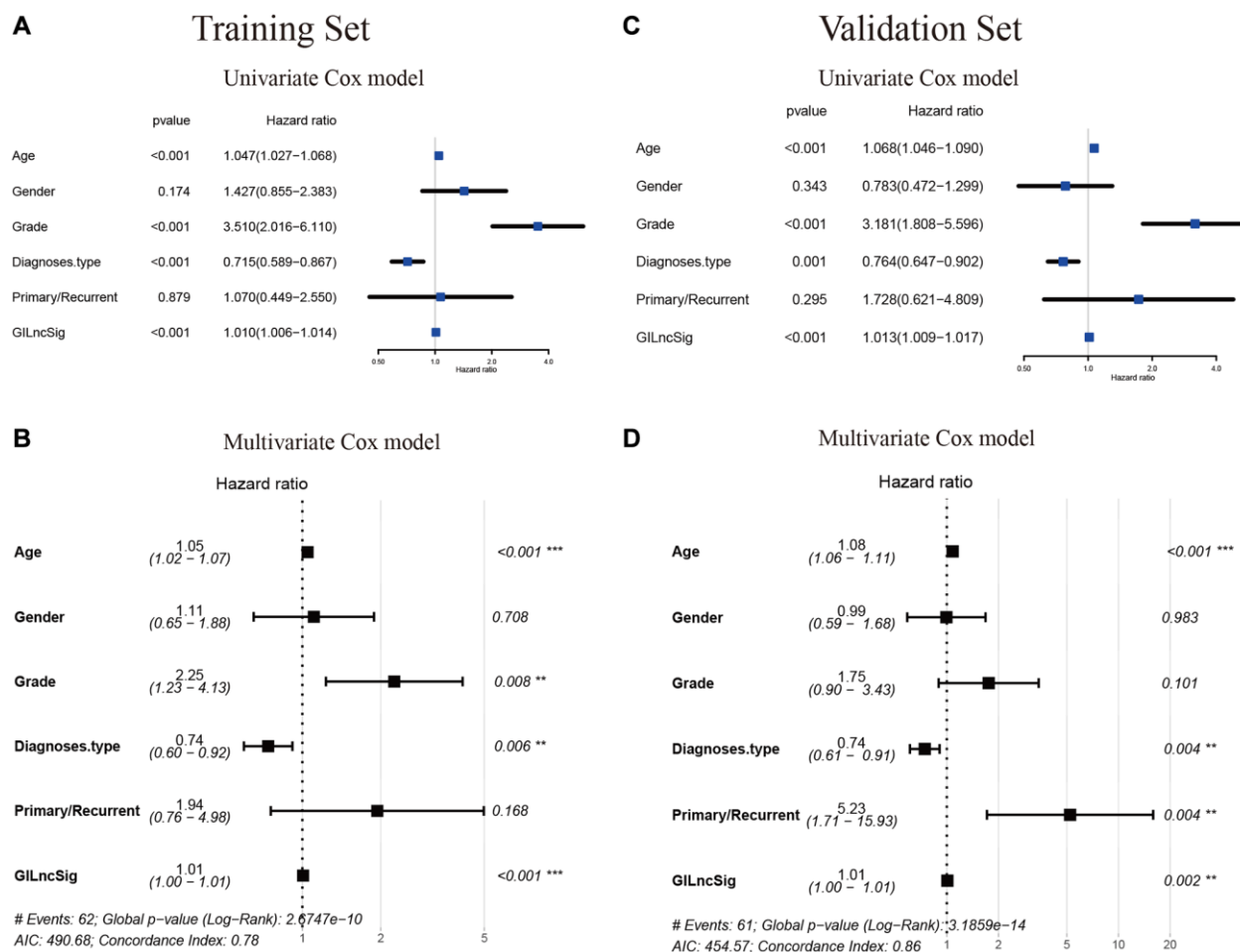


Figure 10. Verification that the GILncSig is an independent prognostic factor. (A–C) The result of univariate Cox regression showed that the age, grade, diagnostic types of gliomas and GILncSig signature were significant prognostic factors for LGG patients. (B–D) while only the factor of age, diagnostic types and GILncSig signature were also associated with overall survival in the multivariate Cox regression model, indicating that the GILncSig signature was the independent prognostic biomarker for predicting the survival of LGG patients.

[28]. Prognostic factors in patients with low-grade gliomas include complex genetic, molecular markers of cancer progression and pathological mechanisms, have been explored as reported in the 2016 World Health Organization (WHO) classification of LGG. These markers group LGG as IDH mutated, 1p/19q co-deficient or 1p/19q preserved, TERT mutated or P53 and ATRX mutated oligodendroglioma cell tumors or astrocytomas; simple TERT mutations and triple-negative tumors, including oligodendroglioma or astrocytoma with no IDH mutation, no co-deletion of 1p/19q and no TERT mutation that is not associated with IDH mutation [29]. IDH mutations are present in nearly 80% of low-grade glioma tumors [30]. IDH mutations are stable markers for LGG progression and prognostic classification, and LGG patients with IDH mutations have significantly longer OS and progression-free survival (PFS) [5]. However, stratification of LGG patients using IDH mutation status is not an effective

biomarker for early diagnosis and therapeutic target for LGG. Therefore, there is an urgent need for a robust prognostic model to predict the survival of patients with low-grade gliomas.

Genomic instability drives progression of many tumors [31] and affects prognosis of low-grade gliomas. Genomically unstable tumor cells generate new genetic variants that directly contribute to tumor heterogeneity and resistance to radiotherapy. Poor prognosis of tumors carrying unstable genomes under conventional treatment paradigms implies that the patterns and extent of genomic instability have great prognostic and diagnostic significance for predicting progression and recurrence of tumor. In addition, genomic instability can be used to design novel therapeutic targets. Genomically unstable tumors have a expression level of Neoantigen, a potential target for immunotherapy which can further induce genomic DNA damage in

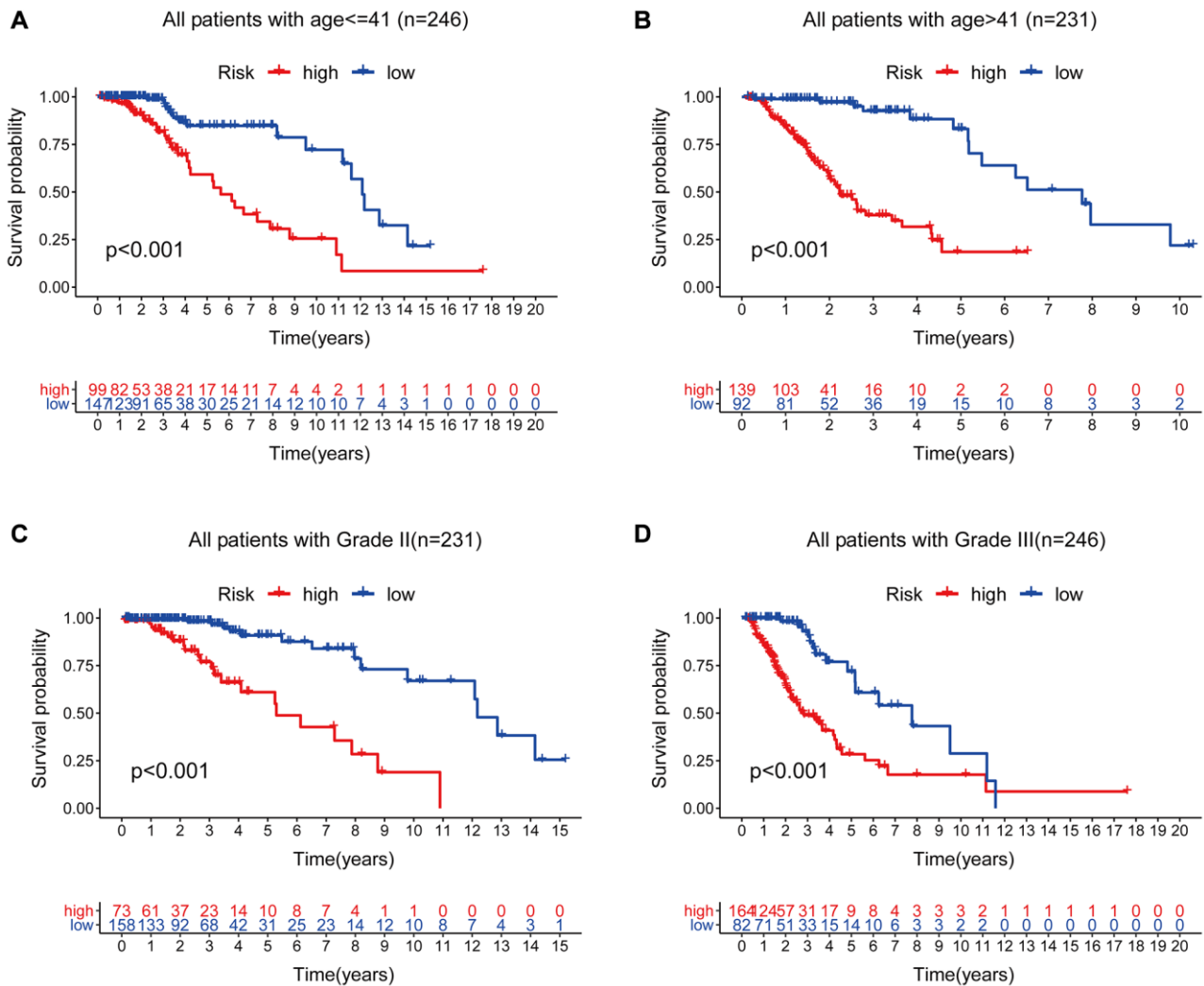


Figure 11. Stratified analysis by age and tumor grade. (A–B) Kaplan-Meier curve analysis of OS in the high- and low-risk groups for patients in the two age groups. (≤ 41 and > 41 years). (C–D) Kaplan-Meier curve analysis for OS in high- and low-risk groups for Grade II and Grade III groups.

genomically unstable low-grade gliomas, thus it can be used to establish new chemotherapeutic regimens [32].

lncRNAs are widely involved in cancer pathways and several studies are exploring them as potential tumor biomarkers. In addition, lncRNAs are potential therapeutic targets as they are associated with progression of many tumors and prognosis of cancer patients [33]. However, only a few studies have explored the role of genomic instability-associated lncRNAs on prognosis of patients with low-grade gliomas. Therefore, the aim of this study was to model genomic instability-associated lncRNAs for predicting overall patient survival and clinical outcome of LGG patients. In addition, a model containing eight genomic instability-associated lncRNAs (H19, FLG-AS1, AC091932.1, AC064875.1, AL138767.3, AC010273.2, AC131097.4 and ISX-AS1) composed of lncRNA signatures (GILncSig) was constructed. GILncSig

grouped patients into low- and high-risk groups with statistically significant differences in overall survival for the training set which was validated using the independent testing set. The GILncSig model predicted significantly better OS in low-risk patients compared with the high-risk patients. Nomogram plots showed that the model was a good predictor of prognosis for 1-, 3- and 5-year OS of low-grade glioma patients. The C-index calibration curve and DCA curve further showed that the model was accurate. The risk-risk model based on eight genomic instability-associated lncRNAs stratified well in LGG was validated by KM curve, ROC curve and risk map analysis. In this study, samples with high-risk scores were significantly correlated with high-risk propensity. In addition, the high-risk group exhibited lower OS compared with the low-risk group, indicating that risk scores are highly correlated with LGG progression and poor prognosis. Further, univariate and multivariate Cox

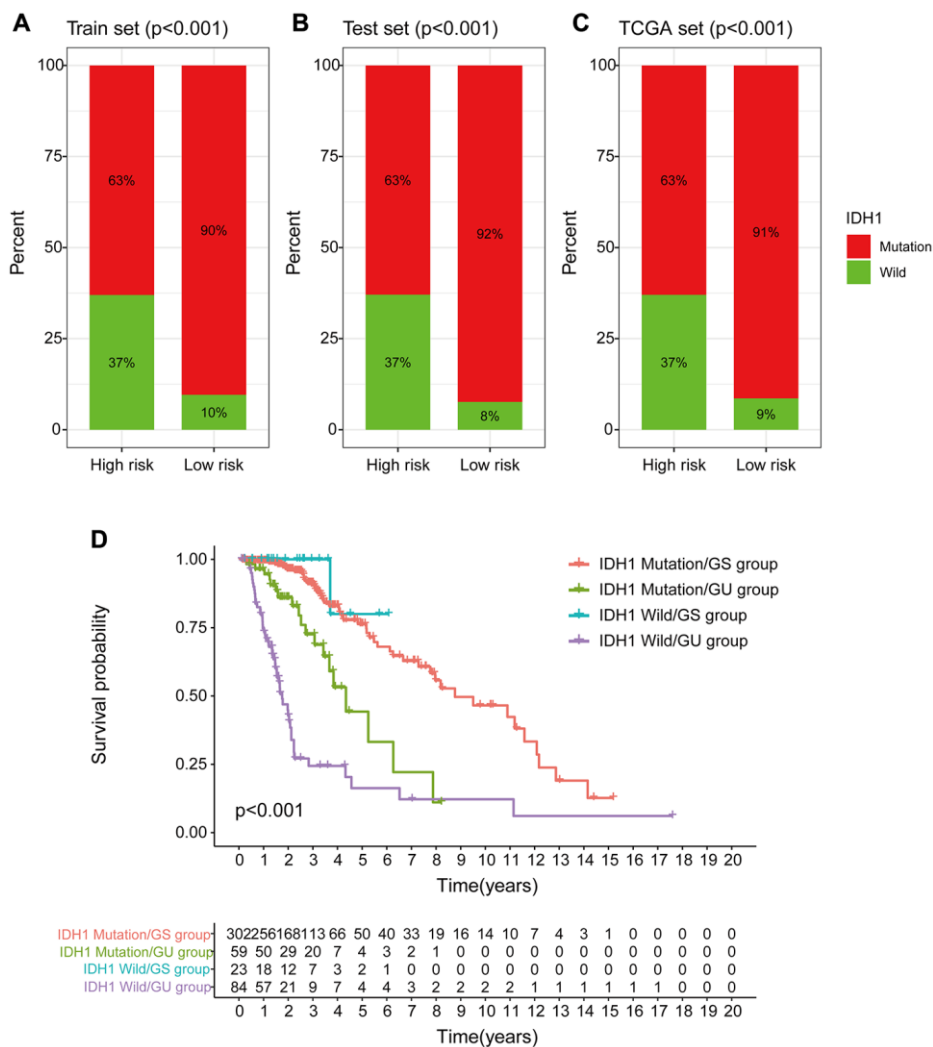


Figure 12. Correlation between GILncSig and IDH1 somatic mutations. (A–C) Proportion of IDH1 mutations in the high- and low-risk groups using the training set, testing set and TCGA set. (D) Kaplan-Meier curve analysis of OS of patients with IDH1 mutant status and wild-type status for the combined GS and GU groups.

regression analyses showed that the signature was an accurate prognostic factor. The findings of this study showed that lncRNA H19, AC091932.1, AC064875.1, AC010273.2, and AC131097.4 were negatively correlated with overall survival, whereas lncRNA FLG-AS1, AL138767.3, and ISX-AS1 were positively correlated with overall survival ($p < 0.001$), indicating they play a protective role in low-grade gliomas. In addition, genomic instability-associated lncRNA genes were correlated with clinical features of two external datasets.

lncRNA H19 was the first identified RNA regulator implicated in multiple steps of tumorigenesis [34] and is a potential tumorigenic lncRNA for glioma. The expression of H19 non-coding RNA is induced by c-Myc product, a member of the MYC proto-oncogene family, which promotes development of glioma [35]. In addition, H19 acts as a miRNA precursor gene to promote glioma growth, which induces the production of miRNA-675 that further modulates expression of cancer-associated calmodulin 13 (CDH13) [36]. Expression level of H19 is highly correlated with drug resistance in glioma cells. Notably, treatment of gliomas with the drug Temozolomide showed survival of fewer glioma cells and analysis showed low expression levels of H19 [37, 38]. Further studies showed that high expression levels of H19 in glioma cells, promotes development of glioma and invasive metastasis which is consistent with the findings of our study on genomic instability showing a distinct signature. In addition, H19 plays an integral role in evolution of several types of cancer. For example, lncRNA H19 expression is significantly upregulated in primary and metastatic foci of colorectal cancer and is associated with poor prognosis in colorectal cancer. Ectopic H19 expression increases *ex vivo* metastasis of colorectal cancer cells and induces epithelial to mesenchymal transition (EMT) [39]. In addition, lncRNA H19 knockdown inhibits breast cancer cell proliferation and induces apoptosis by regulating miR-130a-3p/SATB1. Furthermore, H19 acts as miRNA-130a-3p, sponge leading to upregulation of SATB1, thus promoting breast cancer progression [40]. Interestingly, H19 interacts with 4E-BP1 at the TOS motif and inhibits 4E-BP1 binding to Raptor competitively, implying that inhibition of pituitary tumors by H19 is more effective compared with carte blanche treatment [41]. AC064875.1 has been used as a new prognostic marker for glioma in recent studies, and high expression level of AC064875.1 was correlated with poor prognosis of patients [42]. Another study reports that lncRNA FLG-AS1 predicts the pathological response and prognosis of neoadjuvant radiotherapy for esophageal squamous cell carcinoma and plays a positive role in prognosis [43]. However, the roles of several other lncRNAs (AC091932.1, AC010273.2,

AC131097.4, AL138767.3, ISX-AS1) identified in our GILncSig have not been reported in previous studies. Our validation results on multiple datasets from different databases, findings from previous studies, indicate that GILncSig is a good predictor of the prognosis of cancer patients and serve as an indicator of genomic instability in cancer patients.

Analysis showed that GILncSig significantly distinguishes IDH1 gene wild-type and mutant status and implying that GILncSig can be used to predict the IDH1 mutant status based on the significantly higher IDH1 mutation rate in the low-risk group of patients with genomic instability-associated lncRNAs compared with the high-risk group. In addition, the mutation rate of IDH1 mutant patients was significantly higher compared with that of IDH1 wild-type patients in the low- and high-risk groups ($p < 0.001$). Survival curve analysis after use of GILncSig on IDH1 wild-type and IDH1 mutant patients separately, showed that prognosis of IDH1 Mutation/GS group > IDH1 Mutation/GU group > IDH1 Wild/GU group ($p < 0.001$). The outcomes showed that prognosis of IDH1 mutation genomic instability group was significantly better compared with that of IDH1 wild genomic instability group. These findings indicate that IDH1 mutation status, combined with genomic instability-associated lncRNA, is a more effective prognostic marker compared with use of IDH1 mutation status alone.

The prognostic value of the eight genomic instability-associated lncRNAs identified in this study has not been fully explored in various cancers. Although our study offers essential insights for determining genomic instability and prognosis in patients with low-grade gliomas, the study had some limitations. Genomic instability plays an essential role in tumor biology, as it increases DSB formation through uncontrolled DNA replication during oncogene activation. Uncontrolled DNA replication results in continued DDR activation allowing cells to virtually enter a state of protective senescence where further proliferation is halted [44]. Oncogene-induced senescence is accompanied by significant alterations in chromatin organization, causing most genome to enter a silent heterochromatin state. Recent studies reported an acutely transformed cDNA from the transcriptome of mouse melanoma cells in clone M3 which was transfected to form foci of tumor-transformed cells, and this RNA was referred as genomic instability-inducing RNA (Ginir). High expression of Ginir in mouse fibroblasts induces genomic instability and oncogenic transformation [45].

Although genomic instability-associated lncRNAs were validated in the TCGA database and the GSE16011 dataset in the CGGA and GEO databases, more

independent datasets are needed to validate GILncSig to ensure its reproducibility and robustness. Additionally, roles of these lncRNAs should be further explored through flow cytometry, PCR or IHC. Moreover, to assess the early diagnostic utility of this lncRNA model, the differential expression of the eight lncRNAs in normal and tumor tissues should be evaluated. In addition, further animal studies and cellular experiments should be conducted to test the predictive accuracy of our signature and to explore the mechanisms of genomic instability-associated lncRNAs. Despite these limitations, only a few genomic instability-associated lncRNA signatures have been reported in low-grade gliomas, therefore this signature plays an important role in prediction of survival of LGG patients.

In this study, we constructed a robust and independent prognostic signature based on eight genomic instability-associated lncRNAs. Our results demonstrated that the GILncSig signature was significantly associated with microenvironment infiltrating immune cells and IDH1 mutation status. This study may help improve the power of existing diagnosis and prognosis prediction of low-grade glioma.

MATERIALS AND METHODS

Acquisition and processing of low-grade glioma data

Low-grade glioma gene expression profile data, clinical features and somatic mutation information were obtained from TCGA database (<https://portal.gdc.cancer.gov>), a cancer genome mapping database, and VarScan2, a genomic variant-based investigation platform (GenomeVIP) [46, 47]. VarScan2 is used for analysis of somatic mutations and copy number alterations (CNA) in exome data using tumor-normal tissue pairs. TCGA-LGG cohort comprised 529 low-grade glioma samples, which we identified using the Ensembl database and GENCODE Release 29 (GRCh38.p12) (<https://www.genecodegenes.org/human/>) annotations were used to discriminate the pattern of low-grade glioma lncRNA copy number alterations and expression [48]. We postulated that most GENCODE lncRNAs (97%) are located in the covered regions. All low-grade glioma samples obtained were randomly divided into training and testing groups. A total of 477 patients were included after excluding patients with incomplete clinical information and survival time less than 30 days. The training set comprised 240 patients, which were used for analysis of lncRNA signatures and generation of a prognostic risk model. A test set of 237 cases was used to independent validation of the performance of the prognostic risk model. Two additional sets for

independent verifications of glioma, GSE16011 and CGGA mRNA-seq-693 were acquired from the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16011>) and Chinese Glioma Genome Atlas (<http://www.cgga.org.cn/download.jsp>), respectively, characterized by large sample sizes and multiple clinicopathological features. All the data we used were quoted from the TCGA and GEO databases, and had obtained the consent of the data submitter. And the patient data in this work were also acquired from the publicly available datasets whose informed consent of patients were complete. All the authors declared that there was no conflict of interest in this study.

Identification of lncRNAs associated with genome instability in LGG

lncRNA somatic mutation profiles and expression profiles in low-grade glioma genomes were retrieved for identification of mutation-derived binding genomic instability-associated lncRNAs. The proportions of somatic mutations for each LGG individual were calculated then patients were sorted in descending order based on the number of somatic mutation frequencies using Wilcoxon rank-sum test. Further, gene mutation Unsupervised clustering was performed to define the genomic unstable group (GU) which comprised the first 25% of patients and the genomic stable group (GS) was defined as the last 25% of patients. The lncRNAs expression profiles of the GU group and GS group were compared by SAM method ($|\log FC| > 1$ and FDR-adjusted with $p < 0.05$) and low-expressed genes were excluded (The mean value of lncRNA expression in all samples was less than 0.5) to determine genomic instability-related lncRNAs.

Identification of genome instability-associated lncRNAs prognostic signature for low-grade glioma

A univariate Cox proportional risk regression profiling was used to determine the correlation among expression levels of genomic instability-associated lncRNAs and overall survival ($p < 0.001$) of patients with low-grade gliomas. The cox regression analysis was downscaled using the Lasso analysis to further select the best lncRNAs for construction of risk models. A total of 1000 cross-validations were performed to prevent overfitting before identifying the genomic instability-related lncRNAs associated with OS. Multivariate cox regression analysis was then performed to determine the risk coefficients of prognostic markers for lncRNAs associated with genomic instability.

Genomic instability-associated lncRNAs signature (GILncSig) was constructed based on the multiple

regression analysis of coefficients and expression levels of lncRNAs associated with genomic instability for prognosis prediction using the following equation:

$$\text{GILncSig Riskscore} = \sum_{i=1}^n \text{coefGILncRNA}_i \times \text{Exp GILncRNA}_i$$

where, GILncSig Risk score represents the prognostic risk score of patients with low-grade glioma. Coef represents multivariate cox regression coefficient, GILncRNA_i represents the *i*th genomic instability-associated lncRNA, whereas Exp GILncRNA_i represents the expression level of genomic instability-associated lncRNAs. Further, univariate and multivariate Cox regression analysis were used to determine whether genomic instability-associated lncRNAs are independent predictors for low-grade gliomas.

Identification and verification of genome instability-associated lncRNAs survival analysis for Low-grade glioma

Patients were grouped into GILncSig low- and high- risk groups using the median scores of LGG patients in training set as risk cutoff values. Kaplan-Meier survival curve analysis was performed to compare differences in overall survival (OS) between the low- and high- groups. Hazard ratios (HRs) and 95% confidence interval (CI) were used to analyze whether the prognoses of the two groups were significantly different. A time-dependent subject receiver operating characteristic curve (ROC) analysis was utilized to compare the specificity and sensitivity of GILncSig risk score for prognosis of patients with low-grade glioma. The above results were then validated using a validation set.

Construction of a nomogram based on genome instability-associated lncRNAs

Genomic instability-associated lncRNA risk scores and clinical indices were incorporated into the construction of the model to optimize its predictive capability. Nomogram plots were generated to predict patient survival at 1-, 3- and 5- years based on the multivariate Cox regression results. Calibration curves were used to verify its predictive value. The C index and 95% CI confidence interval were evaluated to determine the prediction performance of the nomogram using clinical decision curve analysis (DCA). These results were then validated using the validation set.

Association of genome instability-associated lncRNAs signature with the immune infiltration

The relationship between lncRNA models and immune cell infiltration was evaluated using expression profiles

using ESTIMATE and CIBERSORT algorithms to perform correlation with levels of immune cell subpopulations in the samples [49, 50]. Unsupervised hierarchical clustering of the samples was performed. The infiltrative expression of 22 immune cells in the LGG training set was determined using CiberSort R package. Further, the correlation among immune cell infiltration and risk score was determined using Spearman test. The results were then validated with a validation set to explore the potential mechanisms of different risk groups reflected by GILncSig.

Functional enrichment analysis

Pearson correlation analysis was performed to establish the correlation between lncRNA and mRNA paired expression, using the top 10 mRNAs that co-expressed lncRNA-associated chaperone proteins. Further, to explore the potential function of lncRNAs, functional enrichment analysis of lncRNA-related mRNA chaperone proteins was performed. The significantly enriched gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were then identified.

External validation of genes in the genome instability-associated lncRNAs signature model

Validation of the signature model genes was performed using CGGA mRNAseq-693 tool in the CGGA database and the GSE16011 dataset from GEO database. The equation used to calculate risk scores was used for validation. Kaplan-Meier curve analysis was performed, and expression of genes of different ages, genders, grades and IDH1 mutation status were presented in boxplots.

Statistical analysis

All statistical analyzes and plots were performed or generated using the R software version 4.0.3. We used the Pearson correlation analysis to determine the relationship between the two. "Limma" R package was used for differential gene screening. The "glmnet" package was used for LASSO Cox regression model analysis. The "survivor" and "survminer" packages were used for perform survival data analysis and for generation of Kaplan-Meier plots. A two-tailed log-rank test was used to determine statistical significance of survival curves. ROC analysis was performed using the "TimeROC" package. The "rms" package was used to construct the Nomogram and for generation of calibration curves. The "clusterProfiler" package was used for GO and KEGG enrichment analysis. $P < 0.05$ was statistically significant. We used non-parametric tests to deal with non-normal distributions, and used parametric tests to deal with normal distributions.

Abbreviations

LGG: low-grade glioma; lncRNA: Long non-coding RNA; GILncSig: Genomic instability-associated lncRNA signature; Lasso: least absolute shrinkage and selection operator; ROC: receiver operating characteristic; AUC: area under the curve; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; OS: overall survival; TCGA: The Cancer Genome Atlas; CGGA: Chinese Glioma Genome Atlas; GEO: Gene Expression Omnibus; GU: genomic unstable; GS: genomic stable.

AUTHOR CONTRIBUTIONS

AierpatiMaimaiti: Conceptualization, Methodology, Validation, Investigation, Supervision, Software, Formal analysis, Visualization, Writing - original draft, Writing- Reviewing and Editing. Xixian Wang: Resources, Supervision, Investigation. Yinan Pei: Conceptualization, Methodology. Nuerbiye Nuermaiti: Software, Validation. Abudirehman Tuersunniyazi: Formal analysis, Visualization. Yaeraili Abula: Data curation, Validation. Zhaohai Feng: Software, Validation. Lei Jiang: Data curation, Validation. Xin Shi: Methodology, Validation. Maimaitijiang Kasimu: Formal analysis, Investigation, Writing- Reviewing and Editing.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

REFERENCES

1. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* 2016; 131:803–20.
<https://doi.org/10.1007/s00401-016-1545-1>
PMID:27157931
2. Dong F, Li Q, Yang C, Huo D, Wang X, Ai C, Kong Y, Sun X, Wang W, Zhou Y, Liu X, Li W, Gao W, et al. PRMT2 links histone H3R8 asymmetric dimethylation to oncogenic activation and tumorigenesis of glioblastoma. *Nat Commun.* 2018; 9:4552.
<https://doi.org/10.1038/s41467-018-06968-7>
PMID:30382083
3. Gittleman H, Sloan AE, Barnholtz-Sloan JS. An independently validated survival nomogram for lower-grade glioma. *Neuro Oncol.* 2020; 22:665–74.
<https://doi.org/10.1093/neuonc/noz191>
PMID:31621885
4. Jones PS, Carroll KT, Koch M, DiCesare JAT, Reitz K, Frosch M, Barker FG 2nd, Cahill DP, Curry WT Jr. Isocitrate Dehydrogenase Mutations in Low-Grade Gliomas Correlate With Prolonged Overall Survival in Older Patients. *Neurosurgery.* 2019; 84:519–28.
<https://doi.org/10.1093/neuros/nyy149>
PMID:29846690
5. Lu VM, O'Connor KP, Shah AH, Eichberg DG, Luther EM, Komotar RJ, Ivan ME. The prognostic significance of CDKN2A homozygous deletion in IDH-mutant lower-grade glioma and glioblastoma: a systematic review of the contemporary literature. *J Neurooncol.* 2020; 148:221–29.
<https://doi.org/10.1007/s11060-020-03528-2>
PMID:32385699
6. Duffau H, Taillandier L. New concepts in the management of diffuse low-grade glioma: Proposal of a multistage and individualized therapeutic approach. *Neuro Oncol.* 2015; 17:332–42.
<https://doi.org/10.1093/neuonc/nou153>
PMID:25087230
7. Yoshioka KI, Matsuno Y. Genomic destabilization and its associated mutagenesis increase with senescence-associated phenotype expression. *Cancer Sci.* 2021; 112:515–22.
<https://doi.org/10.1111/cas.14746>
PMID:33222327
8. Condoluci A, Rossi D. Genomic Instability and Clonal Evolution in Chronic Lymphocytic Leukemia: Clinical Relevance. *J Natl Compr Canc Netw.* 2020; 19:227–33.
<https://doi.org/10.6004/jnccn.2020.7623>
PMID:33383567
9. Bao S, Zhao H, Yuan J, Fan D, Zhang Z, Su J, Zhou M. Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Brief Bioinform.* 2020; 21:1742–55.
<https://doi.org/10.1093/bib/bbz118>
PMID:31665214
10. Huang Y, Guo Q, Ding XP, Wang X. Mechanism of long noncoding RNAs as transcriptional regulators in cancer. *RNA Biol.* 2020; 17:1680–92.
<https://doi.org/10.1080/15476286.2019.1710405>
PMID:31888402
11. He H, Yang H, Liu D, Pei R. lncRNA NORAD promotes thyroid carcinoma progression through targeting miR-202-5p. *Am J Transl Res.* 2019; 11:290–99.
PMID:30787987
12. Wu Y, Shen QW, Niu YX, Chen XY, Liu HW, Shen XY. lncNORAD interference inhibits tumor growth and

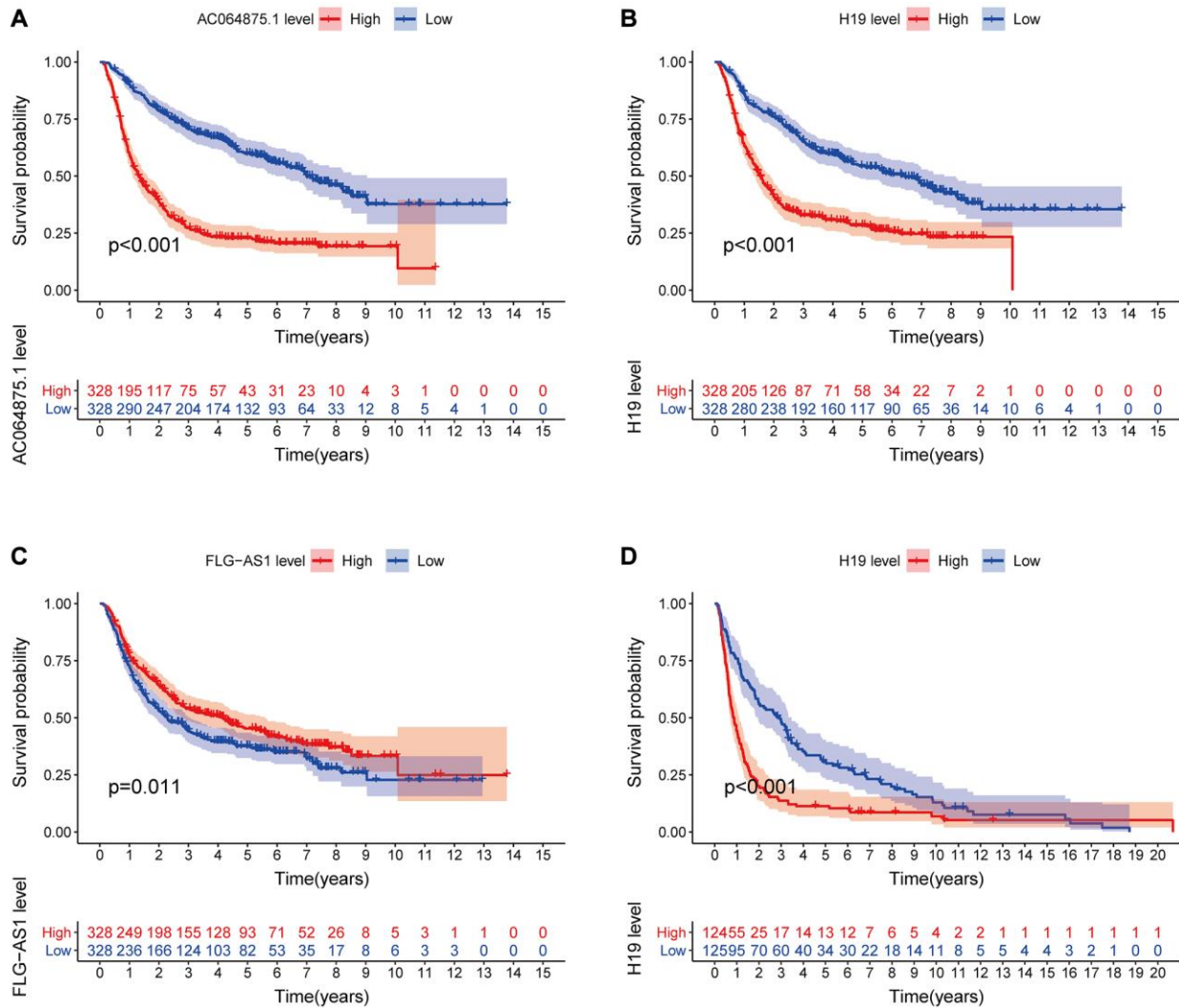
- lung cancer cell proliferation, invasion and migration by down-regulating CXCR4 to suppress RhoA/ROCK signaling pathway. *Eur Rev Med Pharmacol Sci*. 2020; 24:5446–55.
https://doi.org/10.26355/eurrev_202005_21329
PMID:32495879
13. Kopp F, Elguindy MM, Yalvac ME, Zhang H, Chen B, Gillett FA, Lee S, Sivakumar S, Yu H, Xie Y, Mishra P, Sahenk Z, Mendell JT. PUMILIO hyperactivity drives premature aging of *Norad*-deficient mice. *Elife*. 2019; 8:e42650.
<https://doi.org/10.7554/eLife.42650>
PMID:30735131
 14. Marques FC, Ulitsky I. Aging well with *Norad*. *Elife*. 2019; 8:e45974.
<https://doi.org/10.7554/eLife.45974>
PMID:30848246
 15. Munschauer M, Nguyen CT, Sirokman K, Hartigan CR, Hogstrom L, Engreitz JM, Ulirsch JC, Fulco CP, Subramanian V, Chen J, Schenone M, Guttman M, Carr SA, Lander ES. The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature*. 2018; 561:132–36.
<https://doi.org/10.1038/s41586-018-0453-z>
PMID:30150775
 16. Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y, Mendell JT. Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell*. 2016; 164:69–80.
<https://doi.org/10.1016/j.cell.2015.12.017>
PMID:26724866
 17. Bernard E, Nannya Y, Hasserjian RP, Devlin SM, Tuechler H, Medina-Martinez JS, Yoshizato T, Shiozawa Y, Saiki R, Malcovati L, Levine MF, Arango JE, Zhou Y, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat Med*. 2020; 26:1549–56.
<https://doi.org/10.1038/s41591-020-1008-z>
PMID:32747829
 18. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, Morozova O, Newton Y, Radenbaugh A, Pagnotta SM, Anjum S, Wang J, Manyam G, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell*. 2016; 164:550–63.
<https://doi.org/10.1016/j.cell.2015.12.028>
PMID:26824661
 19. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, Pekmezci M, Rice T, Kosel ML, Smirnov IV, Sarkar G, Caron AA, Kollmeyer TM, et al. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med*. 2015; 372:2499–508.
<https://doi.org/10.1056/NEJMoa1407279>
PMID:26061753
 20. Jachimowicz RD, Beleggia F, Isensee J, Velpula BB, Goergens J, Bustos MA, Doll MA, Shenoy A, Checa-Rodriguez C, Wiederstein JL, Baranes-Bachar K, Bartenhagen C, Hertwig F, et al. UBQLN4 Represses Homologous Recombination and Is Overexpressed in Aggressive Tumors. *Cell*. 2019; 176:505–19.e22.
<https://doi.org/10.1016/j.cell.2018.11.024>
PMID:30612738
 21. Lin JZ, Lin N, Zhao WJ. Identification and validation of a six-lncRNA prognostic signature with its ceRNA networks and candidate drugs in lower-grade gliomas. *Genomics*. 2020; 112:2990–3002.
<https://doi.org/10.1016/j.ygeno.2020.05.016>
PMID:32447005
 22. Li X, Meng Y. Survival analysis of immune-related lncRNA in low-grade glioma. *BMC Cancer*. 2019; 19:813.
<https://doi.org/10.1186/s12885-019-6032-3>
PMID:31419958
 23. Lin W, Huang Z, Xu Y, Chen X, Chen T, Ye Y, Ding J, Chen Z, Chen L, Qiu X, Qiu S. A three-lncRNA signature predicts clinical outcomes in low-grade glioma patients after radiotherapy. *Aging (Albany NY)*. 2020; 12:9188–204.
<https://doi.org/10.18632/aging.103189>
PMID:32453707
 24. Mirchia K, Sathe AA, Walker JM, Fudym Y, Galbraith K, Viapiano MS, Corona RJ, Snuderl M, Xing C, Hatanpaa KJ, Richardson TE. Total copy number variation as a prognostic factor in adult astrocytoma subtypes. *Acta Neuropathol Commun*. 2019; 7:92.
<https://doi.org/10.1186/s40478-019-0746-y>
PMID:31177992
 25. Núñez FJ, Mendez FM, Kadiyala P, Alghamri MS, Savelieff MG, Garcia-Fabiani MB, Haase S, Koschmann C, Calinescu AA, Kamran N, Saxena M, Patel R, Carney S, et al. IDH1-R132H acts as a tumor suppressor in glioma via epigenetic up-regulation of the DNA damage response. *Sci Transl Med*. 2019; 11:eaq1427.
<https://doi.org/10.1126/scitranslmed.aaq1427>
PMID:30760578
 26. Arita H, Matsushita Y, Machida R, Yamasaki K, Hata N, Ohno M, Yamaguchi S, Sasayama T, Tanaka S, Higuchi F, Iuchi T, Saito K, Kanamori M, et al. TERT promoter mutation confers favorable prognosis regardless of 1p/19q status in adult diffuse gliomas with IDH1/2 mutations. *Acta Neuropathol Commun*. 2020; 8:201.
<https://doi.org/10.1186/s40478-020-01078-2>
PMID:33228806

27. Houillier C, Wang X, Kaloshi G, Mokhtari K, Guillevin R, Laffaire J, Paris S, Boisselier B, Idbaih A, Laigle-Donadey F, Hoang-Xuan K, Sanson M, Delattre JY. IDH1 or IDH2 mutations predict longer survival and response to temozolomide in low-grade gliomas. *Neurology*. 2010; 75:1560–66. <https://doi.org/10.1212/WNL.Ob013e3181f96282> PMID:20975057
28. Izquierdo C, Alentorn A, Idbaih A, Simó M, Kaloshi G, Ricard D, Barritault M, Meyronet D, Bruna J, Honnorat J, Delattre JY, Ducray F. Long-term impact of temozolomide on 1p/19q-codeleted low-grade glioma growth kinetics. *J Neurooncol*. 2018; 136:533–39. <https://doi.org/10.1007/s11060-017-2677-4> PMID:29143276
29. Ruff MW, Uhm JH, Benarroch EE. Neuro-oncology: Implications of the molecular era. *Neurology*. 2019; 92:568–74. <https://doi.org/10.1212/WNL.00000000000007126> PMID:30760635
30. Acquaviva G, Visani M, de Biase D, Marucci G, Franceschi E, Tosoni A, Brandes AA, Rhoden KJ, Pession A, Tallini G. Prevalence of the single-nucleotide polymorphism rs11554137 (IDH1^{105GGT}) in brain tumors of a cohort of Italian patients. *Sci Rep*. 2018; 8:4459. <https://doi.org/10.1038/s41598-018-22222-y> PMID:29535392
31. Lin SY, Chang CH, Wu HC, Lin CC, Chang KP, Yang CR, Huang CP, Hsu WH, Chang CT, Chen CJ. Proteome Profiling of Urinary Exosomes Identifies Alpha 1-Antitrypsin and H2B1K as Diagnostic and Prognostic Biomarkers for Urothelial Carcinoma. *Sci Rep*. 2016; 6:34446. <https://doi.org/10.1038/srep34446> PMID:27686150
32. Zhang Y, Li J, Yi K, Feng J, Cong Z, Wang Z, Wei Y, Wu F, Cheng W, Samo AA, Salomoni P, Yang Q, Huang Y, et al. Elevated signature of a gene module coexpressed with CDC20 marks genomic instability in glioma. *Proc Natl Acad Sci U S A*. 2019; 116:6975–84. <https://doi.org/10.1073/pnas.1814060116> PMID:30877245
33. Sheng J, He X, Yu W, Chen Y, Long Y, Wang K, Zhu S, Liu Q. p53-targeted lncRNA ST7-AS1 acts as a tumour suppressor by interacting with PTBP1 to suppress the Wnt/ β -catenin signalling pathway in glioma. *Cancer Lett*. 2021; 503:54–68. <https://doi.org/10.1016/j.canlet.2020.12.039> PMID:33476649
34. Cao T, Jiang Y, Wang Z, Zhang N, Al-Hendy A, Mamillapalli R, Kallen AN, Kodaman P, Taylor HS, Li D, Huang Y. H19 lncRNA identified as a master regulator of genes that drive uterine leiomyomas. *Oncogene*. 2019; 38:5356–66. <https://doi.org/10.1038/s41388-019-0808-4> PMID:31089260
35. Barysytte-Lovejoy D, Lau SK, Boutros PC, Khosravi F, Jurisica I, Andrulis IL, Tsao MS, Penn LZ. The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res*. 2006; 66:5330–37. <https://doi.org/10.1158/0008-5472.CAN-06-0037> PMID:16707459
36. Lo Dico A, Costa V, Martelli C, Diceglie C, Rajata F, Rizzo A, Mancone C, Tripodi M, Ottobrini L, Alessandro R, Conigliaro A. MiR675-5p Acts on HIF-1 α to Sustain Hypoxic Responses: A New Therapeutic Strategy for Glioma. *Theranostics*. 2016; 6:1105–18. <https://doi.org/10.7150/thno.14700> PMID:27279905
37. Jia L, Tian Y, Chen Y, Zhang G. The silencing of lncRNA-H19 decreases chemoresistance of human glioma cells to temozolomide by suppressing epithelial-mesenchymal transition via the Wnt/ β -Catenin pathway. *Onco Targets Ther*. 2018; 11:313–21. <https://doi.org/10.2147/OTT.S154339> PMID:29391808
38. Jiang P, Wang P, Sun X, Yuan Z, Zhan R, Ma X, Li W. Knockdown of long noncoding RNA H19 sensitizes human glioma cells to temozolomide therapy. *Onco Targets Ther*. 2016; 9:3501–09. <https://doi.org/10.2147/OTT.S96278> PMID:27366087
39. Zhang Y, Huang W, Yuan Y, Li J, Wu J, Yu J, He Y, Wei Z, Zhang C. Long non-coding RNA H19 promotes colorectal cancer metastasis via binding to hnRNPA2B1. *J Exp Clin Cancer Res*. 2020; 39:141. <https://doi.org/10.1186/s13046-020-01619-6> PMID:32698890
40. Zhong G, Lin Y, Wang X, Wang K, Liu J, Wei W. H19 Knockdown Suppresses Proliferation and Induces Apoptosis by Regulating miR-130a-3p/SATB1 in Breast Cancer Cells. *Onco Targets Ther*. 2020; 13:12501–13. <https://doi.org/10.2147/OTT.S280142> PMID:33324070
41. Wu ZR, Yan L, Liu YT, Cao L, Guo YH, Zhang Y, Yao H, Cai L, Shang HB, Rui WW, Yang G, Zhang XB, Tang H, et al. Inhibition of mTORC1 by lncRNA H19 via disrupting 4E-BP1/Raptor interaction in pituitary tumours. *Nat Commun*. 2018; 9:4624. <https://doi.org/10.1038/s41467-018-06853-3> PMID:30397197

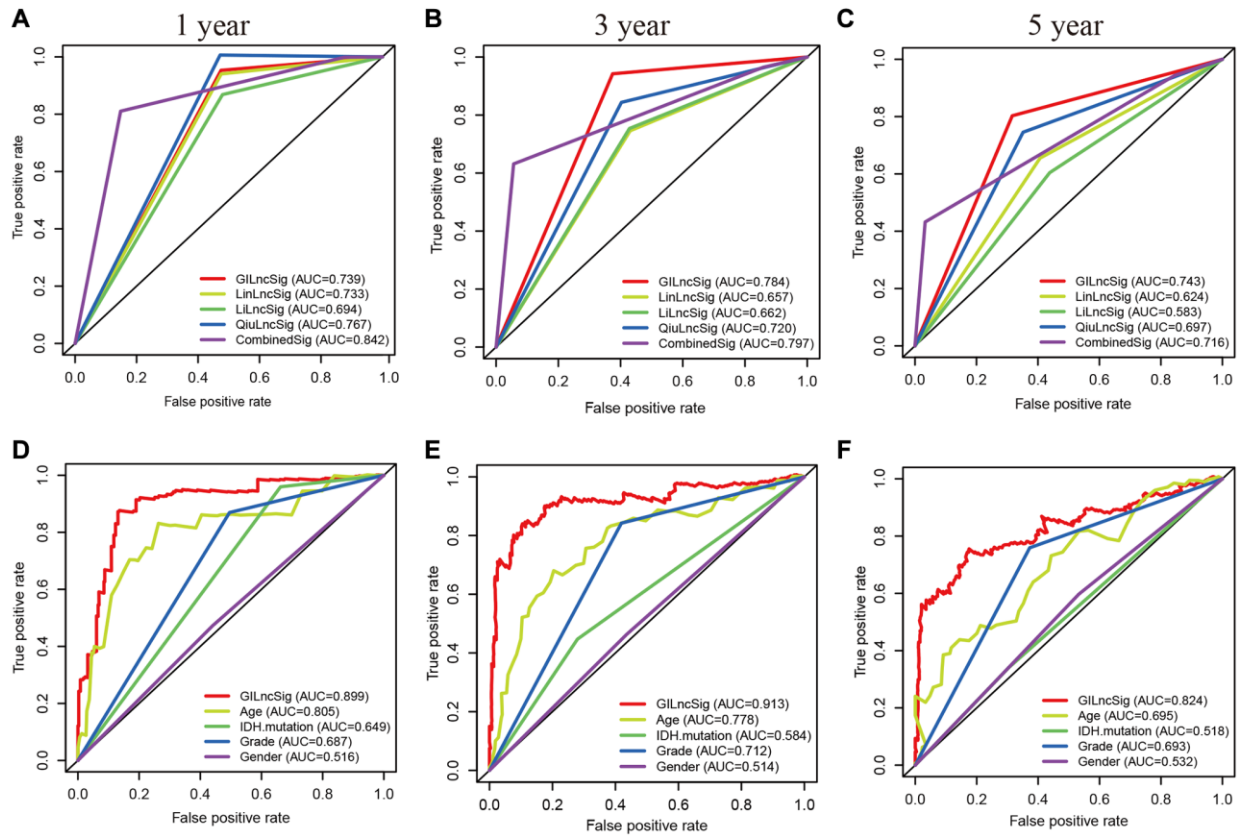
42. Song L, Zhang S, Duan C, Ma S, Hussain S, Wei L, Chu M. Genome-wide identification of lncRNAs as novel prognosis biomarkers of glioma. *J Cell Biochem.* 2019; 120:19518–28.
<https://doi.org/10.1002/jcb.29259>
PMID:[31297871](https://pubmed.ncbi.nlm.nih.gov/31297871/)
43. Zhang C, Zhang Z, Zhang G, Xue L, Yang H, Luo Y, Zheng X, Zhang Y, Yuan Y, Lei R, Yang Z, Zheng B, Zhang Z, et al. A three-lncRNA signature of pretreatment biopsies predicts pathological response and outcome in esophageal squamous cell carcinoma with neoadjuvant chemoradiotherapy. *Clin Transl Med.* 2020; 10:e156.
<https://doi.org/10.1002/ctm2.156>
PMID:[32898328](https://pubmed.ncbi.nlm.nih.gov/32898328/)
44. Perkhofe L, Gout J, Roger E, Kude de Almeida F, Baptista Simões C, Wiesmüller L, Seufferlein T, Kleger A. DNA damage repair as a target in pancreatic cancer: state-of-the-art and future perspectives. *Gut.* 2021; 70:606–17.
<https://doi.org/10.1136/gutjnl-2019-319984>
PMID:[32855305](https://pubmed.ncbi.nlm.nih.gov/32855305/)
45. Panda S, Setia M, Kaur N, Shepal V, Arora V, Singh DK, Mondal A, Teli A, Tathode M, Gajula R, Padhy LC, Shiras A. Noncoding RNA Glnr1 functions as an oncogene by associating with centrosomal proteins. *PLoS Biol.* 2018; 16:e2004204.
<https://doi.org/10.1371/journal.pbio.2004204>
PMID:[30296263](https://pubmed.ncbi.nlm.nih.gov/30296263/)
46. Mashl RJ, Scott AD, Huang KL, Wyczalkowski MA, Yoon CJ, Niu B, DeNardo E, Yellapantula VD, Handsaker RE, Chen K, Koboldt DC, Ye K, Fenyö D, et al. GenomeVIP: a cloud platform for genomic variant discovery and interpretation. *Genome Res.* 2017; 27:1450–59.
<https://doi.org/10.1101/gr.211656.116>
PMID:[28522612](https://pubmed.ncbi.nlm.nih.gov/28522612/)
47. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012; 22:568–76.
<https://doi.org/10.1101/gr.129684.111>
PMID:[22300766](https://pubmed.ncbi.nlm.nih.gov/22300766/)
48. Gravendeel LA, Kouwenhoven MC, Gevaert O, de Rooij JJ, Stubbs AP, Duijm JE, Daemen A, Bleeker FE, Bralten LB, Kloosterhof NK, De Moor B, Eilers PH, van der Spek PJ, et al. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res.* 2009; 69:9065–72.
<https://doi.org/10.1158/0008-5472.CAN-09-2307>
PMID:[19920198](https://pubmed.ncbi.nlm.nih.gov/19920198/)
49. Pang Z, Chen X, Wang Y, Wang Y, Yan T, Wan J, Wang K, Du J. Long non-coding RNA C5orf64 is a potential indicator for tumor microenvironment and mutation pattern remodeling in lung adenocarcinoma. *Genomics.* 2021; 113:291–304.
<https://doi.org/10.1016/j.ygeno.2020.12.010>
PMID:[33309768](https://pubmed.ncbi.nlm.nih.gov/33309768/)
50. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol.* 2018; 1711:243–59.
https://doi.org/10.1007/978-1-4939-7493-1_12
PMID:[29344893](https://pubmed.ncbi.nlm.nih.gov/29344893/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. (A–B) Overall survival probability of patients with high expression levels of AC064875.1 and H19 in the CGGA mRNA-seq-693 dataset was significantly lower compared with that of patients with low expression levels ($p < 0.001$). (C) Overall survival probability of patients with high expression level of FLG-AS1 in the CGGA mRNA-seq-693 dataset was significantly higher compared with the low expression level of FLG-AS1. ($P < 0.05$). (D) The overall survival probability of patients with low expression level of H19 in the GSE16011 dataset was significantly higher compared with the OS of patients with high expression of H19. ($p < 0.001$).



Supplementary Figure 2. (A–C) As the ROC curves showed, the CombinedSig signature could increase the prognostic power for 1 year and 3 year survival rate compared to other four independent signatures. (D–F) Receiver operating characteristic (ROC) curve analysis shows that the prediction of 1 year, 3 year and 5 year in the prognostic accuracy of clinicopathological parameters such as age, gender, IDH status and grade and GILncSig signature prognostic risk score.

Supplementary Tables

Supplementary Table 1. The differentially expressed genomic unstable and genomic stable related lncRNAs in lower-grade glioma patients.

lncRNA	conMean	treatMean	logFC	p Value	FDR
LINC01007	0.740307871	0.168018	-2.139513597	0.0009411	0.00340202
AC109439.2	2.395596545	0.644303	-1.894574393	7.50E-20	1.00E-16
AC002428.2	1.333921383	0.431929	-1.62680791	1.02E-15	2.48E-13
LINC02217	0.66787853	0.240895	-1.471180853	0.0005697	0.002244542
DNMBP-AS1	1.066865411	0.401631	-1.40943686	8.01E-16	2.48E-13
LINC02058	1.58744	0.617127	-1.363063185	3.27E-16	1.25E-13
AL354863.1	1.089323271	0.432545	-1.332510567	7.06E-07	7.61E-06
ISX-AS1	1.021477208	0.414219	-1.302192683	1.43E-09	4.14E-08
AC023421.1	1.608994161	0.668245	-1.267711123	5.76E-06	4.52E-05
AL138767.3	0.814245354	0.340586	-1.257443277	3.59E-14	5.64E-12
ANKRD62P1-PARP4P3	0.776817632	0.330067	-1.234818558	2.04E-09	5.15E-08
AL391845.2	1.148031042	0.488924	-1.231479834	4.52E-14	6.71E-12
AC125616.1	1.074019887	0.457844	-1.23009128	6.29E-08	9.84E-07
AL031710.1	0.809037217	0.347418	-1.219532567	5.44E-07	6.22E-06
AC097641.1	0.577981177	0.249994	-1.209126788	4.43E-12	3.01E-10
AC104024.2	0.780439243	0.338927	-1.203313713	0.00698	0.017782305
AC087442.1	1.129076825	0.516902	-1.127179583	1.52E-15	3.39E-13
AL139246.1	0.776730928	0.358061	-1.117209634	6.76E-11	2.82E-09
AC061961.1	1.246362108	0.5815	-1.099872909	5.22E-07	6.02E-06
AC018410.1	0.717255386	0.337371	-1.088150211	2.08E-08	3.83E-07
FLG-AS1	0.566102043	0.266784	-1.08539195	9.14E-13	7.88E-11
AC009118.1	0.8900668	0.42445	-1.068319556	2.72E-09	6.55E-08
AC124854.1	1.705458064	0.820497	-1.055589518	6.39E-18	5.69E-15
AL390786.1	0.848419029	0.409249	-1.051798604	3.34E-12	2.41E-10
LINC00836	3.672622495	1.797571	-1.030761536	1.68E-12	1.32E-10
RNF219-AS1	1.830454824	0.897093	-1.028873419	1.02E-15	2.48E-13
AL139246.4	0.707122635	0.349126	-1.018210715	1.37E-08	2.71E-07
AL121956.4	0.937556362	0.463795	-1.015419647	8.14E-11	3.35E-09
AL121821.1	0.985791003	0.487908	-1.01467392	1.30E-10	4.83E-09
PIK3CD-AS2	0.269837475	0.539894	1.000585509	4.36E-08	7.33E-07
AC110015.1	0.268027521	0.54466	1.022973919	1.12E-06	1.13E-05
AC091057.1	0.291361296	0.593392	1.026175261	2.57E-14	4.29E-12
AC010273.2	0.468871731	0.959781	1.033511818	3.97E-10	1.35E-08
MIR4435-2HG	0.51666879	1.081763	1.066072864	8.51E-11	3.45E-09
AC016168.2	0.261125127	0.558021	1.095578471	2.93E-07	3.61E-06
AL355974.2	3.247179752	7.105989	1.1298482	1.46E-09	4.16E-08
AP000696.1	0.338403992	0.750695	1.149481128	5.88E-08	9.47E-07
SLCO4A1-AS1	0.318637508	0.718064	1.172197009	0.0131875	0.030202932

AL355974.3	1.528651135	3.485717	1.189196285	2.89E-08	5.08E-07
LINC02308	0.529617356	1.305029	1.301059411	1.21E-10	4.55E-09
AC091932.1	0.28513877	0.740863	1.377541821	2.42E-06	2.19E-05
AJ011932.1	0.332752511	0.872696	1.391029184	5.75E-10	1.83E-08
AL512785.1	0.339901779	0.930143	1.452334171	3.45E-05	0.000207188
CRNDE	0.858686284	2.468972	1.523707283	6.96E-15	1.33E-12
FOXD3-AS1	0.690153292	2.027998	1.555067621	5.27E-11	2.35E-09
AC131097.4	0.22318894	0.693529	1.635691708	0.0010957	0.003836379
AC025171.5	0.269306355	0.850237	1.658615922	5.13E-09	1.13E-07
HOXD-AS2	0.298261437	0.954926	1.678811181	7.01E-17	3.75E-14
AL035446.1	0.707730748	2.295653	1.697631989	6.18E-08	9.79E-07
LINC01579	0.401522107	1.311422	1.707580169	1.01E-05	7.28E-05
AGAP2-AS1	1.111059598	3.814722	1.779641793	0.0001462	0.000718734
AL049871.1	0.167635677	0.676832	2.01346857	5.88E-09	1.28E-07
AC002454.1	0.160563184	0.654202	2.026594269	0.0014202	0.004713457
HOTAIRM1	0.59097175	2.463736	2.059686487	3.79E-08	6.49E-07
AC064875.1	0.269906847	1.127583	2.062700604	9.62E-08	1.41E-06
LINC01831	0.154674658	0.734951	2.248411301	2.08E-13	2.42E-11
LINC01956	0.164126221	0.827916	2.334678097	2.35E-21	6.29E-18
LINC02587	0.315083274	2.130638	2.757480259	0.0057771	0.015225469
H19	0.285984353	2.744758	3.262670572	0.0003098	0.001367447

Supplementary Table 2. The specificity and sensitivity of each cut-off value of risk score signature.

Please browse Full Text version to see the data of Supplementary Table 2.