Mutation-derived, genomic instability-associated lncRNAs are prognostic markers in gliomas

Shenglun Li [Equal first author, 1] , Yujia Chen [Equal first author, 1] , Yuduo Guo [1] , Jiacheng Xu [1] , Xiang Wang [1] , Weihai Ning [1] , Lixin Ma [1] , Yanming Qu [1] , Mingshan Zhang [1] , Hongwei Zhang [Corresp. 1]

1 Department of Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing, China

Corresponding Author: Hongwei Zhang
Email address: zhanghongwei@ccmu.edu.cn

Background: Gliomas are the most commonly-detected malignant tumors of the brain. They contain abundant long non-coding RNAs (lncRNAs), which are valuable cancer biomarkers. LncRNAs may be involved in genomic instability; however, their specific role and mechanism in gliomas remains unclear. LncRNAs that are related to genomic instability have not been reported in gliomas.

Methods: The transcriptome data from The Cancer Genome Atlas (TCGA) database were analyzed. The co-expression network of genomic instability-related lncRNAs and mRNA was established, and the model of genomic instability-related lncRNA was identified by univariate Cox regression and LASSO analyses. Based on the median risk score obtained in the training set, we divided the samples into high-risk and low-risk groups and proved the survival prediction ability of genomic instability-related lncRNA signatures. The results were verified in the external data set. Finally, a real-time quantitative polymerase chain reaction assay was performed to validate the signature.

Results: The signatures of 17 lncRNAs (*LINC01579, AL022344.1, AC025171.5, LINC01116, MIR155HG, AC131097.3, LINC00906, CYTOR, AC015540.1, SLC25A21.AS1, H19, AL133415.1, SNHG18, FOXD3.AS1, LINC02593, AL354919.2* and *CRNDE*) related to genomic instability were identified. In the internal data set and Gene Expression Omnibus (GEO) external data set, the low-risk group showed better survival than the high-risk group (p<0.001). In addition, this feature was identified as an independent risk factor, showing its independent prognostic value with different clinical stratifications. The majority of patients in the low-risk group had isocitrate dehydrogenase 1 (*IDH1*) mutations. The expression levels of these lncRNAs were significantly higher in glioblastoma cell lines than in normal cells.

Conclusions: Our study shows that the signature of 17 lncRNAs related to genomic instability has prognostic value for gliomas and could provide a potential therapeutic method for glioblastoma.

1 **Mutation-derived, genomic instability-associated lncRNAs are prognostic markers in**

2 **gliomas**

3 Shenglun Li[1#],Yujia Chen[1#], Yuduo Guo[1], Jiacheng Xu[1], Xiang Wang[1], Weihai Ning[1], Lixin Ma[1],

4 Yanming Qu[1], Mingshan Zhang[1] and Hongwei Zhang[1※]

5 [#]Contributed equally

6 [※]Correspondence: Hongwei Zhang: zhanghongwei@ccmu.edu.cn

7 1. Department of Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing,

8 China.

10 **Abstract**

11 **Background:** Gliomas are the most commonly-detected malignant tumors of the brain. They

12 contain abundant long non-coding RNAs (lncRNAs), which are valuable cancer biomarkers.

13 LncRNAs may be involved in genomic instability; however, their specific role and mechanism in

14 gliomas remains unclear. LncRNAs that are related to genomic instability have not been reported

15 in gliomas.

16 **Methods:** The transcriptome data from The Cancer Genome Atlas (TCGA) database were

17 analyzed. The co-expression network of genomic instability-related lncRNAs and mRNA was

18 established, and the model of genomic instability-related lncRNA was identified by univariate

19 Cox regression and LASSO analyses. Based on the median risk score obtained in the training set,

20 we divided the samples into high-risk and low-risk groups and proved the survival prediction

21 ability of genomic instability-related lncRNA signatures. The results were verified in the external

22 data set. Finally, a real-time quantitative polymerase chain reaction assay was performed to

23 validate the signature.

24 **Results:** The signatures of 17 lncRNAs (*LINC01579, AL022344.1, AC025171.5, LINC01116,*

25 *MIR155HG, AC131097.3, LINC00906, CYTOR, AC015540.1, SLC25A21.AS1, H19,*

26 *AL133415.1, SNHG18, FOXD3.AS1, LINC02593, AL354919.2* and *CRNDE*) related to genomic

27 instability were identified. In the internal data set and Gene Expression Omnibus (GEO) external

28 data set, the low-risk group showed better survival than the high-risk group (p<0.001). In

29 addition, this feature was identified as an independent risk factor, showing its independent

30 prognostic value with different clinical stratifications. The majority of patients in the low-risk

31 group had isocitrate dehydrogenase 1 (*IDH1*) mutations. The expression levels of these lncRNAs

32 were significantly higher in glioblastoma cell lines than in normal cells.

33 **Conclusions:** Our study shows that the signature of 17 lncRNAs related to genomic instability has

34 prognostic value for gliomas and could provide a potential therapeutic method for glioblastoma.

35 **Introduction**

36 Genetic abnormalities and mutations are crucial factors in the development of cancer, resulting in

37 a loss of balance in the nucleotide chain of most human tumors[1]. Genomic instability is defined

38 as an increased likelihood of acquiring chromosomal aberrations due to defects in processes such

39 as DNA repair, DNA damage, replication, or chromosome segregation[2]. Due to genomic

40 disruption, genomic instability is typically subdivided into three categories: nucleotide,

41 microsatellite, and chromosome[3]. The majority of tumor cells exhibit genomic instability. For

42 example, more than two-thirds of human tumors gain or lose whole chromosomes during cell

43 division. The abnormal expression of genes, such as *CDK4*, Ras downstream prokaryotic gene,

44 and BRAF gene can contribute to this instability[3,4]. As a marker of cancer evolution, genomic

45 instability (mainly caused by mutations in DNA repair genes) promotes cancer progression and

46 has been identified as a key prognostic factor[5,6]. In hereditary cancers, genomic instability results

47 from mutations in DNA repair genes and drives cancer development according to the mutator

48 hypothesis. However, the molecular basis of genomic instability in sporadic (non-hereditary)

49 cancers remains unclear. Recent high-throughput sequencing studies suggest that mutations in

50 DNA repair genes are infrequent before therapy, arguing against the mutator hypothesis for these

51 cancers[7]. Additionally, in contrast to the different numbers of genetic mutations, various cancer

52   types exhibit different patterns of somatic cell mutations, indicating specific carcinogenic

53   mechanisms in tissues and cells[8]. It is, therefore, important to identify the potential molecular

54   characteristics associated with genomic instability in cancer and explore their clinical

55   significance.

56   Glioma, commonly known as glioma cerebri (GC), is a malignant primary brain tumor

57   that accounts for approximately 81% of all tumors[9]. Glioblastoma (GBM), the most malignant

58   among them, has an average survival time of just 16 months for patients[10]. GBM is characterized

59   by uncontrolled cell proliferation, diffuse infiltration, necrosis, intense angiogenesis, strong

60   resistance to apoptosis, and rampant gene instability[11]. The etiology of glioma remains unclear;

61   however, exposure to high doses of ionizing radiation, and genetic mutations associated with high

62   penetrance of rare syndromes had been identified as two related risk factors in glioma

63   tumorigenesis[12]. Thus, that it is critical to identify biomarkers associated with genomic instability

64   to accurately evaluate the clinical prognosis of glioma patients.

65   Long non-coding RNAs are a type of RNA with over 200 nucleotides that are incapable of coding

66   proteins[13]. Previous research has demonstrated that lncRNAs play a role in various life events[14,15]

67   and biological processes that are thought to be involved in the development and progression of

68   malignancies, including gliomas. These processes include stemness, angiogenesis, and drug

69   resistance[16]. Their abnormal expression can affect cell proliferation, tumor progression, and

70   metastasis[17]. In addition, abnormal lncRNA expression profiles in clinical glioma samples

71   correlate with the degree of malignancy and histological differentiation, which has important

72   clinical implications for the diagnosis and subclassification of gliomas[18,19]. LncRNAs can act as

73   molecular signaling mediators, regulating the expression of a specific set of genes and

74 corresponding signaling pathways. For example, the *CRNDE-mTOR* pathway can promote the

75 giloma growth[20] and lncRNAs have been found to play a role in various cancers. For example,

76 Ling[21] et al reported a novel lncRNA, *CCAT2*, containing rs6983267 *SNP*. This lncRNA was

77 highly overexpressed in microsatellite-stabilized colorectal cancer and could promote tumor

78 growth, metastasis, or cause chromosome instability. This indicates that lncRNAs involved in the

79 biological process of gene modification contribute to genomic instability and tumor progression.

80 However, studies on the association of lncRNA with genomic instability are still lacking.

81 In this study, we identified a set of lncRNAs associated with genomic instability, and constructed

82 a genomic instability-related lncRNA signature (GIrLncSig). We then validated its prognostic

83 significance in glioma patients. The results showed that this signature had a great prediction role

84 in the prognosis of glioma patients.

85 **Materials and methods**

86 Data collection

87 We collected the clinical characteristics, transcription group data, and somatic cell mutation data

88 from GBM and low-grade glioma (LGG) patients. The datasets generated during this study are

89 available in The Cancer Genome Atlas repository (https://portal.gdc.cancer.gov/). These data

90 were matched using sample names. Patients who had no corresponding information about their

91 survival, or who had less than 30 days of treatment were excluded to eliminate interference from

92 non-cancer causes. We differentiated mRNA and lncRNAs using human genome profiles;

93 mRNAs and lncRNAs were annotated using the HUGO Gene Nomenclature Committee

94 (HGNC2) database (https://www.genenames.org/). A total of 629 samples retained paired

95 lncRNA and mRNA expression profiles, and survival information, somatic mutation information,

96   and common clinicopathological features were obtained for further study. All patients with

97   glioma were randomly divided into two groups: training and test sets. A total of 316 patients in

98   the training set were used to identify the prognostic features of lncRNAs and establish a risk

99   model for the outcome. The test set included 313 patients and was used to independently validate

100 the performance of the prognostic risk model. The GSE43378 dataset generated and analyzed

101 during the current study is available in the Gene Expression Omnibus (GEO) repository

102 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi? acc = GSE43378) for the external validation of

103 the model.

104 Technical route

105 The study process is shown in Figure 1. We collected and then analyzed the data from somatic

106 cell mutations and transcription groups to obtain genomic instability-related lncRNAs

107 (GIrlncRNAs). The relationship between GIrlncRNAs and mRNA was then analyzed using co-

108 expression analysis. Next, we randomly divided the patient cohort into training and testing

109 sets. Cox and Lasso regression analyses of GIrlncRNAs were conducted to construct a prognostic

110 signature. The signature was evaluated using mutation correlation analysis, model comparison,

111 independent prognostic value analysis, clinical stratification, examination of the external dataset

112 and cell line validation.

113 Identification of GIrlncRNAs

114 We used a method from the Mutator Hypothesis [22] to identify GIrlncRNAs. Patients with the

115 highest cumulative gene mutation count and who were at the lowest 25% cumulative gene

116 mutation count were assigned to genomic unstable-like (GU) and genome stable-like (GS)

117 groups, respectively. The average expression of lncRNAs between the two groups was compared

118   using the Wilcoxon rank-sum test in the limma package of the R software. The cut-off thresholds

119   were intended to be |fold change|>2.0 and false discovery rate (FDR)<0.05.

120   Construction of GIrLncSig

121   The "survival" software package of R was used to conduct univariate Cox regression analysis on

122   the training set to assess the relationship between the expression level of GIrlncRNAs and the

123   overall survival time of patients. The least absolute shrinkage and selection operator (LASSO)

124   regression algorithm was used to further screen candidate GIrlncRNAs to construct the

125   GIrlncRNAs prognostic signature (GIrLncSig). The following formula, based on a combination

126   of the Cox coefficient and gene expression, was used to calculate the signature risk score:

127
$$GIrLncSig\ score = \sum_{i=1}^{n} coefi \times Ei$$
128

129   GIrLncSig is the prognostic risk score for patients with glioma. Ei represents the expression level

130   of lncRNAi in patients and coefi represents the coefficient of lncRNAi. The median GIrLncSig

131   score was used as the risk cut-off point to divide glioma patients into low-risk and high-risk

132   groups. The survival curves of the two groups were plotted by Kaplan-Meier method using

133   "Survminer" and "Survival" packages in R language, and a log-rank sum test that obtained

134   p<0.05 was considered significant.

135   Real time-PCR validation

136   The U87, U251, LN229, and U343 cell lines of human glioblastoma and the immortalized cell

137   line SVGp12 were used for the cellular level validation of lncRNA in the model. All cells were

138   cultured in DMEM supplemented with 10% FBS, 100 U/ml penicillin, and 100 U/ml

139   streptomycin. The cultures were maintained at 37 °C in a humidified environment containing 5%

140 carbon dioxide and were confirmed to be mycoplasma-free prior to experimental use.

141   A total of 1 ml Trizol (Invitrogen, Waltham, MA, USA) was added to the collected cells to

142 extract total cellular RNA, and the absorbance value of RNA at 260 nm was measured using a

143 Nanodrop 2000 UV spectrophotometer. Then, 1 ug of RNA was removed and used to synthesize

144 cDNA by reverse transcription (New England Biolabs, Ipswich, MA, USA), the SYBR Green

145 (Applied Biosystems, Foster City, CA, USA) method and CFX96 real-time PCR system (Bio-

146 Rad, Hercules, CA, USA) were used for real-time polymerase chain reaction (RT-PCR), and actin

147 was used as an internal control. Amplification was performed at 95 °C/120 s followed by 39

148 cycles at 95 °C/5 s and 60 °C/30 s. Relative expression of RNA was calculated using the $2^{-\Delta\Delta Ct}$

149 method. The primers were generated by Sangon Biotechnology (Shanghai, China). The primers

150 for lncRNAs in GIrLncSig are listed in Table 1.

151 **Results**

152 **Identified genomic instability-related lncRNAs (GIrlncRNAs) in glioma samples**

153 To identify genomic instability-related lncRNAs, we sorted glioma patients with the number of

154 gene somatic mutation sites and defined the top 25% (n=227) as the genome unstable (GU) group

155 and the last 25% (n=110) as the genome stable (GS) group. Next, we compared the expression

156 profiles of lncRNAs in these two groups to identify the lncRNAs that were significantly different.

157 After screening, we identified 91 differentially expressed lncRNAs (Supplementary Table 1) and

158 selected the top forty significantly different lncRNAs between the GS and GU groups for

159 heatmap plotting (Figure 2A). These 91 differentially expressed lncRNAs were significantly

160 associated with genomic instability; thus, they were defined as genome instability-related

161 lncRNAs (GIrlncRNAs). Consensus cluster analysis was then performed on 896 samples from

162 the cancer genome atlas (TCGA) collection (including 390 GBMs and 506 LGGs), and all

163 samples were divided into two groups based on the differential expression of GIrlncRNAs and

164 the median number of accumulated somatic mutations (Figure 2B). The group with the higher

165 number of accumulated somatic mutations was defined as the GU-like group and the group with a

166 lower number of accumulated somatic mutations was defined as the GS-like group. The two

167 clustered groups had significantly different somatic mutation patterns (Figure 2C). We analyzed

168 the expression correlation between lncRNAs and their target mRNAs. We also selected the top

169 nine strongly correlated mRNAs as target genes since lncRNAs cannot perform direct biological

170 functions but can regulate mRNAs. We then constructed a co-expression network from our results

171 (Figure 2D)

172 **Construction of GIrlncRNAs signature in the training set.**

173 A total of 629 glioma patients from the TCGA project were randomly divided into a training

174 dataset (n=316) and a test dataset (n=313). Patients' clinicopathological characteristics are shown

175 in Table 2. Eighty prognostic related lncRNAs were identified based on univariate Cox

176 proportional risk regression in the training set (Supplementary Table 2, Supplementary Figure 2).

177 Lasso regression analysis and stepwise multifactor Cox proportional risk regression analysis were

178 performed since lncRNAs have distinct biological functions and many lncRNAs interact with

179 each other. These steps were taken because there were large amounts of data and the potential for

180 inaccurate model construction. A total of 17 lncRNAs were screened as independent prognostic

181 factors to build the prognostic model (Figure 3 A-B, Table 3). A prediction model was finally

182 obtained: genomic instability-related lncRNA signature (GIrLncSig) score =

183 $(0.0441217761138621 \times LINC01579) + (-0.197814767909076 \times AL022344.1) +$

184   (0.0387926268692573×*AC025171.5*)+(0.000726291473736049×*LINC01116*)+(0.140438109893

185   274×*MIR155HG*)+(0.205877895933553×*AC131097.3*)+

186   (-.0369769918481854×*LINC00906*)+(0.0588040388194378×*CYTOR*)+

187   (-.0336307964944069×*AC015540.1*)+

188   (-.108512558356276×*SLC25A21.AS1*)+(0.00613294332460668×*H19*)+(0.0819381463964954×

189   *AL133415.1*)+(0.00495065493579931×*SNHG18*)+(0.0180933234213193×*FOXD3.AS1*)+

190   (-.0263113038864061×*LINC02593*)+(0.0103411968758757×*AL354919.2*)+(0.046081456238200

191   4×*CRNDE*). A positive coefficient for lncRNAs suggested that high expressions were associated

192   with long survival times as a protective factor in glioma. In contrast, a negative coefficient for

193   lncRNAs indicates that it is a risk factor for glioma. The risk score of each patient in the training

194   set was obtained using GIrLncSig, and then these patients were divided into high- and low-risk

195   groups based on the median risk score (0.0103411968758757). Survival analysis revealed that the

196   low-risk group had significantly better survival rates than the high-risk group (Figure 3C). The 5-

197   year survival rates were 4.43% in the high-risk group and 14.56% in the low-risk group. The

198   ROC curve showed that the AUC was 0.934, 0.898, and 0.904 at 1 year, 3 years, and 5 years,

199   respectively (Figure 3D). We sorted patients in the training set according to their risk score and

200   there were different expression levels of GIrlncRNAs in the two groups (Figure 3E). Patients

201   with high-risk scores exhibited an increased expression of risk genes and decreased expression of

202   protective genes, whereas those with low-risk scores exhibited the opposite trend. We also found

203   a significant difference in somatic mutation patterns between patients in the high-risk and low-

204   risk groups, implying that the model could accurately reflect the somatic mutation situation in

205   gliomas (Figure 3F).

Independent validation of GIrLncSig in the glioma cerebri data set of transcription data

and external validation in the GEO data set.

An independent TCGA test set of 313 patients was used to examine the performance of GIrLncSig. The test set used the same GIrLncSig and risk thresholds as the training set and the 313 patients in the test set were divided into high-risk (n=173) and low-risk groups (n=140). There was a significant difference in overall survival (Figure 4A). The overall survival rate was significantly lower in the high-risk group than in the low-risk group, which is consistent with the results of the training group. The 5-year survival rate was 7.51% in the high-risk group, which is lower than the 16.4% in the low-risk group (Figure 4A). ROC curve analysis showed that the AUC of GIrLncSig was 0.875 for 1 year and 0.773 for 3 years (Figure 4B). The expression levels of GIrLncSig, patient death distribution counts and model lncRNA expression are shown in Figure 4C. A significant difference was observed in the somatic mutation pattern between patients in the high-risk and low-risk groups (Figure 4D), and this result was similar in the training group. To further validate the prognostic significance of GIrLncSig, a cross-platform was performed in other independent datasets from different platforms. GSE43378, a dataset from the Gene Expression Matrix data set, was downloaded for further analysis because of the large sample size and complete clinicopathological features. We investigated the relationship between glioma and genomic instability in this independent dataset and found that four lncRNAs (Linc01116, CRNDE, Linc00906, and SNHG18) in GIrLncSig were included in GSE43378. The expressions of Linc01116 and CRNDE were positively correlated with tumor grade (Figure 4E-F) and the survival time was significantly different between their high- and low-expression subgroups (Figure 4G-H). These results were consistent with those observed in the training and

228   test sets.

229   **Evaluation of independent prognostic significance of GIrLncSig and clinical stratification**

230   **analysis**

231   In the TCGA dataset, the prognosis of GIrLncSig was analyzed by adjusting for clinical

232   stratification, including age (>65 and ≤65 years), gender (male and female), tumor classification

233   (WHO grade Ⅱ, Ⅲ and Ⅳ), and other clinical factors. In all clinical subgroups, the survival rate

234   in the low-risk group was higher than that in the high-risk group (Figure 5). This demonstrated

235   that GIrLncSig exhibited a significant independent prognostic prediction value on the overall

236   survival of glioma patients under different clinical stratification conditions.

237   **GIrLncSig was associated with IDH1 mutation status and comparison between GIrLncSig**

238   **prediction and other lncRNA model predictions**

239   We observed a difference in the *IDH1* status in a cohort of glioma patients. *IDH1* is a security

240   mutant in the nervous region, as reported, therefore we determined that it was important to

241   evaluate the connection between GIrLncSig and *IDH1* mutation status. We compared the

242   differences between the high-risk and low-risk groups on the training set and test set and the

243   results revealed that the proportion of *IDH1* mutation was significantly higher in the low-risk

244   group than in the high-risk group, regardless of whether the patient was affected by glioma

245   cerebri (GC) (Figure 6A) or low-grade glioma (Figure 6C). These results imply that patients with

246   *IDH1* mutations were at lower risk. We then divided the patients into four subgroups (*IDH1*

247   mutation/GS-like, *IDH1* mutation/GU-like, *IDH1* wild-type/GS-like and *IDH1* wild-type/GU-

248   like) to consider both the GIrLncSig and *IDH1* mutation status. As shown in Figures 6B and 6D,

249   there were significant differences in the survival time among the four groups. This suggests that

250   *IDH1* mutation status might be related to genomic instability and combining them together is

251   better for predicting prognosis in glioma patients.

252   We then compared the predictability of GIrLncSig and two recently reported lncRNA signatures

253   for survival prediction using the same TCGA glioma patients. Compared to the 10-lncRNA

254   prognostic signature reported by PAN[23] (PanlncSig), 4-lncRNA prediction signature reported by

255   Li[24] (LiDlncSig), 5-lncRNA prediction signature reported by Li[25] (LiXlncSig), 10-lncRNA

256   prediction signature reported by Luan[26] (LuanlncSig), and 9-lncRNA prediction signature

257   reported by Tao[27] (TaolncSig) as displayed in Figure 6E, our GIrLncSig had an AUC value of

258   0.889 in 1-year OS. These results were more effective than those of PanLncSig (AUC value =

259   0.852), LiDlncSig (AUC value = 0.835), LiXlncSig (AUC value = 0.790), LuanlncSig (AUC

260   value = 0.842), and TaolncSig (AUC value = 0.862) in predicting patient survival. In summary,

261   the above results confirmed the reliability and effectiveness of GIrLncSig in predicting GC

262   patients.

263   **Validation of GirLncSig**

264   To further validate the prognostic significance of GirLncSig in gliomas, we verified eight

265   lncRNAs in GirLncSig in four glioblastoma cell lines (U87, U251, LN229, and U343) and one

266   normal cell line (SVGp12). The expression levels of these eight lncRNAs were significantly

267   higher in glioblastoma cell lines than in normal cells (Figure 7A). Similarly, the results of clinical

268   samples also proved that the expression of GirLncSig in normal brain tissue (N1-N3) was lower

269   than that in glioma tumor tissue (T1-T6), further validating the prognostic significance of

270   GirLncSig.

271   We then examined the relationship between GirLncSig and cell proliferation and found that in the

272   glioma samples from the TCGA database, the expression of almost all model-related sub-risk

273   lncRNAs was positively correlated with the expression of cell proliferation markers *Ki67* and

274   *PCNA* (Figure 8A-B). Similarly, we also detected *CDH2* and *VIM*, which are markers of cell

275   migration, invasion, and mesenchymal transition. The eight lncRNAs were found to be positively

276   correlated (Figure 8C-D). These findings tentatively show that GIrLncSig can promote tumors by

277   affecting cell proliferation and EMT.

278   **Discussion**

279   Genomic instability is caused by chromosomal segregation errors during mitosis that result in

280   aneuploidy mutations of whole chromosomes in daughter cells, or by DNA damage that causes

281   chromosome structure changes. These changes can result in gene translocations, deletions,

282   inversions, and breaks[28,29]. Genomic changes can occur at different levels, from mutations in a

283   single or few nucleotides to the gain or loss of entire chromosomes. This may trigger abnormal

284   divisions, multinucleation, and tripolar mitosis. Maintaining genetic integrity is critical for cell

285   viability and is accomplished through complex repair processes. When these processes are

286   defective, genomic instability occurs, resulting in the accumulation of chromosomal mutations

287   which can then cause susceptibility to cancer[30,31]. Genomic instability plays a fundamental role in

288   cancer progression and recurrence, suggesting that its pattern and degree have significant

289   diagnostic and prognostic implications[32,33]. LncRNAs have recently been demonstrated to be

290   promising tumor biomarkers. Abnormal expression of lncRNAs in tumors is associated with

291   disease progression and may serve as a prognostic marker for patients[34,35,36]. Furthermore, recent

292   advances in understanding the functional mechanisms underlying lncRNAs have recognized that

293   lncRNAs are essential for genomic stability[37,38].

294   Glioma is the most common primary intracranial tumor. To date, no studies have examined the

295   lncRNA signature of genomic instability in glioma. In this study, we identified a group of

296   GIrlncRNAs in GC and determined their significance for predicting patient survival. We then

297   identified 91 GIrlncRNAs and compared their expression levels in different mutation counts.

298   Systematic clustering analysis and subsequent differential analysis of mutation counts confirmed

299   the association between these lncRNAs with genomic instability. Based on co-expression with 91

300   GIrlncRNAs, we further investigated whether GIrlncRNAs could predict the clinical outcomes of

301   glioma patients by constructing a GIrLncSig consisting of 17 GIrlncRNAs (*LINC01579,*

302   *AL022344.1, AC025171.5. LINC01116, MIR155HG, AC131097.3, CYTOR, AC015540.1,*

303   *SLC25A21.AS1, H19, AL133415.1, SNHG18, FOXD3.AS1, LINC02593, AL354919.2,* and

304   *CRNDE*). Using GIrLncSig we classified patients into two risk groups with a significant

305   difference in survival in the training set, which was validated on an independent test set. Similar

306   results were found in the external GEO dataset, GSE43378. These validation results across

307   multiple datasets and technology platforms indicate that GIrLncSig may be an indicator of

308   genomic instability in cancer patients. Few studies have been conducted on tumor models based

309   on genomic instability lncRNA. Recent studies were selected for glioma lncRNA models,

310   including those by Li et al. (2019) and Pan et al. (2020), to construct lncRNA models using the

311   lncRNA expression data of patients in the CGGA database and GSE16011 dataset, respectively.

312   None of these studies discussed the impact of genomic instability on tumors, nor did they use the

313   TCGA database. In addition, Xiaomen et al. (2021) studied the prognostic model of immune

314   related lncRNA. Tao et al. (2021) explored the relationship between epithelial mesenchymal

315   transition (EMT)-related lncRNA and the prognosis of glioma. Similarly, Luan et al. (2019)

316 studied the autophagy-related lncRNAs. The prognostic model of lncRNAs was also mentioned

317 in Kiran et al. (2019)'s article, but it was for low-grade glioma and did not cover GBM samples.

318 In addition, Zheng et al. (2021)'s article studied the role of barbed wire-related lncRNAs in

319 glioma. However, no study has linked lncRNAs to genomic instability. This study investigates

320 genome instability, since genome instability and ferroptosis affect cancer at different magnitudes.

321 Following the first discovery of glioblastoma (GBM) in 2008 by the Johns Hopkins University

322 team in the United States using whole exome sequencing technology, *IDH1-R132* was detected in

323 18 cases (12%) of 149 tumor samples. The *IDH1* mutations detected in the study are more

324 common in young patients with secondary tumors, and the median overall survival (OS) of

325 patients with mutations reaches 3.8 years, which is much better than the median OS of 1.1 years

326 in IDH1 wild-type patients (HR=3.7, P<0.001)[41,42]. In our study, the frequency of IDH1 mutations

327 was significantly higher in the low-risk group than in the high-risk group, reflecting the reliability

328 of the model.

329 This study creatively constructed a prediction model related to genomic instability through the

330 analysis of patients' genomic mutation data, transcriptome data, and clinical data in the TCGA

331 database. This model has an obvious leading role in predicting the prognosis of glioma patients

332 (Figure 6E). It may play a role in the survival difference of glioma patients in high- and low-risk

333 groups (Figure 4A-C) and in different grades of glioma (Figure 5A-C). The value of this model in

334 the prognosis and survival of glioma patients has been demonstrated. In addition, we validated

335 the expression levels of eight GIrlncRNAs with known sequences in four glioma cell lines, and

336 the results revealed significant differences in the expression levels of eight lncRNAs as risk

337 factors between glioma cell lines and ordinary cells.

338   Although our study showed some links between genomic instability and prognosis in patients

339   with glioma, several limitations remain. We lacked the validation of the function of our

340   GIrLncSig to ensure its accuracy and reproducibility. Therefore, further studies are necessary and

341   its action mechanism in glioma development and progression remains to be further investigated.

342   **Conclusion**

343   In summary, we constructed a risk prediction signature consisting of 17 lncRNAs associated with

344   genomic instability. This signature can predict the prognosis of glioma patients and reveal their

345   genomic instability. It serves as a reference indicator for clinical individualized treatment and has

346   great significance for glioma patients. Therefore, this signature may be an important tool for

347   further investigating the role of lncRNAs in genomic instability. Further research will clarify the

348   relationship between lncRNAs and genomic instability in tumors.

349

350

351

352 **References**

353 1. Hanahan Douglas., Weinberg Robert A.(2011). Hallmarks of cancer: the next generation.

354   Cell, 144(5), 646-74. doi:10.1016/j.cell.2011.02.013

355 2. Tubbs Anthony., Nussenzweig André.(2017). Endogenous DNA Damage as a Source of

356   Genomic Instability in Cancer. Cell, 168(4), 644-656. doi:10.1016/j.cell.2017.01.002

357 3. Pikor Larissa., Thu Kelsie., Vucic Emily., Lam Wan.(2013). The detection and implication of

358   genome instability in cancer. Cancer Metastasis Rev, 32(null), 341-52. doi:10.1007/s10555-

359   013-9429-5

360 4. Carvalho Claudia M B., Lupski James R.(2016). Mechanisms underlying structural variant

361   formation in genomic disorders. Nat Rev Genet, 17(4), 224-38. doi:10.1038/nrg.2015.25

362 5. Kamata Tamihiro., Hussain Jahan., Giblett Susan., Hayward Robert., Marais Richard.,

363   Pritchard Catrin.(2010). BRAF inactivation drives aneuploidy by deregulating CRAF. Cancer

364   Res, 70(21), 8475-86. doi:10.1158/0008-5472.CAN-10-0603

365 6. Cui Yongping., Borysova Meghan K., Johnson Joseph O., Guadagno Thomas M.(2010).

366   Oncogenic B-Raf(V600E) induces spindle abnormalities, supernumerary centrosomes, and

367   aneuploidy in human melanocytic cells. Cancer Res, 70(2), 675-84. doi:10.1158/0008-

368   5472.CAN-09-1491

369 7. Suzuki, K., Ohnami, S., Tanabe, C., Sasaki, H., Y asuda, J., Katai, H., et al. (2003).The

370   genomic damage estimated by arbitrarily primed PCR DNA fingerprinting is useful for the

371   prognosis of gastric cancer. Gastroenterology 125, 1330–1340. doi:

372   10.1016/j.gastro.2003.07.006

373 8. Ottini L., Falchetti M., Lupi R., Rizzolo P., Agnese V., Colucci G., Bazan V., Russo A.(2006).

374   Patterns of genomic instability in gastric cancer: clinical implications and perspectives. Ann

375   Oncol, null(undefined), vii97-102. doi:10.1093/annonc/mdl960

376   9.   Negrini Simona., Gorgoulis Vassilis G., Halazonetis Thanos D.(2010). Genomic instability--

377   an evolving hallmark of cancer. Nat Rev Mol Cell Biol, 11(3), 220-8. doi:10.1038/nrm2858

378   10.   Anandakrishnan Ramu., Varghese Robin T., Kinney Nicholas A., Garner Harold R.(2019).

379   Estimating the number of genetic mutations (hits) required for carcinogenesis based on the

380   distribution   of   somatic   mutations.   PLoS   Comput   Biol,   15(3),   e1006881.

381   doi:10.1371/journal.pcbi.1006881

382   11.   Jahangir   Moini.,   PirouzPiran.(2020).Chapter   1-Histophysiology.Functional   and   Clinical

383   Neuroanatomy,1-49. doi: 10.1016/B978-0-12-817424-1.00001-X.

384   12.   Stupp Roger., Mason Warren P., van den Bent Martin J., Weller Michael., Fisher Barbara.,

385   Taphoorn Martin J B., Belanger Karl., Brandes Alba A., Marosi Christine., Bogdahn Ulrich.,

386   Curschmann Jürgen., Janzer Robert C., Ludwin Samuel K., Gorlia Thierry., Allgeier Anouk.,

387   Lacombe Denis., Cairncross J Gregory., Eisenhauer Elizabeth., Mirimanoff René O.,

388   European   Organisation   for   Research   and   Treatment   of   Cancer   Brain   Tumor   and

389   Radiotherapy Groups.,   National Cancer Institute of Canada Clinical Trials Group.(2005).

390   Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med,

391   352(10), 987-96. doi:10.1056/NEJMoa043330

392   13.   Negrini Simona., Gorgoulis Vassilis G., Halazonetis Thanos D.(2010). Genomic instability--

393   an evolving hallmark of cancer. Nat Rev Mol Cell Biol, 11(3), 220-8. doi:10.1038/nrm2858

394   14.   Mattick John S., Rinn John L.(2015). Discovery and annotation of long noncoding

395   RNAs. Nat Struct Mol Biol, 22(1), 5-7. doi:10.1038/nsmb.2942

396   15. Mercer Tim R., Mattick John S.(2013). Structure and function of long noncoding

397   RNAs in epigenetic regulation. Nat Struct Mol Biol, 20(3), 300-7.

398   doi:10.1038/nsmb.2480

399   16. Shi Jia., Dong Bo., Cao Jiachao., Mao Yumin., Guan Wei., Peng Ya., Wang Suinuan.

400   (2017). Long non-coding RNA in glioma: signaling pathways. Oncotarget, 8(16),

401   27582-27592. doi:10.18632/oncotarget.15175

402   17. Sanchez Calle Anna., Kawamura Yumi., Yamamoto Yusuke., Takeshita Fumitaka.,

403   Ochiya Takahiro.(2018). Emerging roles of long non-coding RNA in cancer. Cancer

404   Sci, 109(7), 2093-2100. doi:10.1111/cas.13642

405   18. Wang Qixue., Zhang Junxia., Liu Yanwei., Zhang Wei., Zhou Junhu., Duan Ran., Pu

406   Peiyu., Kang Chunsheng., Han Lei.(2016). A novel cell cycle-associated lncRNA,

407   HOXA11-AS, is transcribed from the 5-prime end of the HOXA transcript and is a

408   biomarker of progression in glioma. Cancer Lett, 373(2), 251-9.

409   doi:10.1016/j.canlet.2016.01.039

410   19. Jing S-Y., Lu Y-Y., Yang J-K., Deng W-Y., Zhou Q., Jiao B-H.(2016). Expression of

411   long non-coding RNA CRNDE in glioma and its correlation with tumor progression

412   and patient survival. Eur Rev Med Pharmacol Sci, 20(19), 3992-3996.

413   20. Wang Yunliang., Wang Yutong., Li Jinfeng., Zhang Yuzhen., Yin Honglei., Han

414   Bing.(2015). CRNDE, a long-noncoding RNA, promotes glioma cell growth and

415   invasion through mTOR signaling. Cancer Lett, 367(2), 122-8.

416   doi:10.1016/j.canlet.2015.03.027

417   21. Ling Hui., Spizzo Riccardo., Atlasi Yaser., Nicoloso Milena., Shimizu Masayoshi.,

418   Redis Roxana S., Nishida Naohiro., Gafà Roberta., Song Jian., Guo Zhiyi., Ivan

419   Cristina., Barbarotto Elisa., De Vries Ingrid., Zhang Xinna., Ferracin Manuela.,

420   Churchman Mike., van Galen Janneke F., Beverloo Berna H., Shariati Maryam.,

421   Haderk Franziska., Estecio Marcos R., Garcia-Manero Guillermo., Patijn Gijs A.,

422   Gotley David C., Bhardwaj Vikas., Shureiqi Imad., Sen Subrata., Multani Asha S.,

423   Welsh James., Yamamoto Ken., Taniguchi Itsuki., Song Min-Ae., Gallinger Steven.,

424   Casey Graham., Thibodeau Stephen N., Le Marchand Loïc., Tiirikainen Maarit.,

425   Mani Sendurai A., Zhang Wei., Davuluri Ramana V., Mimori Koshi., Mori Masaki.,

426   Sieuwerts Anieta M., Martens John W M., Tomlinson Ian., Negrini Massimo.,

427   Berindan-Neagoe Ioana., Foekens John A., Hamilton Stanley R., Lanza Giovanni.,

428   Kopetz Scott., Fodde Riccardo., Calin George A.(2013). CCAT2, a novel noncoding

429   RNA mapping to 8q24, underlies metastatic progression and chromosomal

430   instability in colon cancer. Genome Res, 23(9), 1446-61. doi:10.1101/gr.152942.112

431 22. Bao Siqi., Zhao Hengqiang., Yuan Jian., Fan Dandan., Zhang Zicheng., Su

432   Jianzhong., Zhou Meng.(2020). Computational identification of mutator-derived

433   lncRNA signatures of genome instability for improving the clinical outcome of

434   cancers: a case study in breast cancer. Brief Bioinform, 21(5), 1742-1755.

435   doi:10.1093/bib/bbz118

436 23. Pan Yuan-Bo., Zhu Yiming., Zhang Qing-Wei., Zhang Chi-Hao., Shao Anwen.,

437   Zhang Jianmin.(2020). Prognostic and Predictive Value of a Long Non-coding RNA

438   Signature in Glioma: A lncRNA Expression Analysis. Front Oncol, 10(undefined),

439   1057. doi:10.3389/fonc.2020.01057

440   24. Li Depei., Lu Jie., Li Hong., Qi Songtao., Yu Lei.(2019). Identification of a long

441   noncoding RNA signature to predict outcomes of glioblastoma. Mol Med Rep, 19(6),

442   5406-5416. doi:10.3892/mmr.2019.10184

443   25. Li Xiaomeng., Sun Li., Wang Xue., Wang Nan., Xu Kanghong., Jiang Xinquan., Xu

444   Shuo.(2021). A Five Immune-Related lncRNA Signature as a Prognostic Target for

445   Glioblastoma.   Front   Mol   Biosci,   8(undefined),   632837.

446   doi:10.3389/fmolb.2021.632837

447   26. Luan Fangkun., Chen Wenjie., Chen Miao., Yan Jun., Chen Hao., Yu Haiyue., Liu

448   Tieqi., Mo Ligen.(2019). An autophagy-related long non-coding RNA signature for

449   glioma. FEBS Open Bio, 9(4), 653-667. doi:10.1002/2211-5463.12601

450   27. Tao Chuming., Luo Haitao., Chen Luyue., Li Jingying., Zhu Xingen., Huang Kai.

451   (2021). Identification of an epithelial-mesenchymal transition related long non-

452   coding RNA (LncRNA) signature in Glioma. Bioengineered, 12(1), 4016-4031.

453   doi:10.1080/21655979.2021.1951927

454   28. Carvalho Claudia M B., Lupski James R.(2016). Mechanisms underlying structural

455   variant formation in genomic disorders. Nat Rev Genet, 17(4), 224-38. doi:10.1038/

456   nrg.2015.25

457   29. Wickramasinghe Vihandha O., Venkitaraman Ashok R.(2016). RNA Processing and

458   Genome   Stability: Cause   and   Consequence.   Mol   Cell,   61(4),   496-505.

459   doi:10.1016/j.molcel.2016.02.001

460   30. Mackay Hannah L., Moore David., Hall Callum., Birkbak Nicolai J., Jamal-Hanjani

461   Mariam., Karim Saadia A., Phatak Vinaya M., Piñon Lucia., Morton Jennifer P.,

462   Swanton Charles., Le Quesne John., Muller Patricia A J.(2018). Genomic instability

463   in mutant p53 cancer cells upon entotic engulfment. Nat Commun, 9(1), 3070.

464   doi:10.1038/s41467-018-05368-1

465   31. Rajendran Barani Kumar., Deng Chu-Xia.(2017). Characterization of potential

466   driver mutations involved in human breast cancer by computational approaches.

467   Oncotarget, 8(30), 50252-50272. doi:10.18632/oncotarget.17225

468   32. Zhang ShiQi., Pan XiaoYong., Zeng Tao., Guo Wei., Gan Zijun., Zhang Yu-Hang.,

469   Chen Lei., Zhang YunHua., Huang Tao., Cai Yu-Dong.(2019). Copy Number

470   Variation Pattern for Discriminating MACROD2 States of Colorectal Cancer

471   Subtypes.   Front   Bioeng   Biotechnol,   7(undefined),   407.

472   doi:10.3389/fbioe.2019.00407

473   33. Rancoule Chloé., Vallard Alexis., Guy Jean-Baptiste., Espenel Sophie., Sauvaigo

474   Sylvie., Rodriguez-Lafrasse Claire., Magné Nicolas.(2017). [Impairment of DNA

475   damage   response   and   cancer].   Bull   Cancer,   104(11),   962-970.

476   doi:10.1016/j.bulcan.2017.09.006

477   34. Kronenwett Ulrike., Ploner Alexander., Zetterberg Anders., Bergh Jonas., Hall Per.,

478   Auer Gert., Pawitan Yudi.(2006). Genomic instability and prognosis in breast

479   carcinomas. Cancer Epidemiol Biomarkers Prev, 15(9), 1630-5. doi:10.1158/1055-

480   9965.EPI-06-0080

481   35. Mettu Rama K R., Wan Ying-Wooi., Habermann Jens K., Ried Thomas., Guo Nancy

482   Lan.(2010). A 12-gene genomic instability signature predicts clinical outcomes in

483   multiple cancer types. Int J Biol Markers, 25(4), 219-28. doi:10.5301/jbm.2010.6079

484   36. Gupta Rajnish A., Shah Nilay., Wang Kevin C., Kim Jeewon., Horlings Hugo M.,

485   Wong David J., Tsai Miao-Chih., Hung Tiffany., Argani Pedram., Rinn John L.,

486   Wang Yulei., Brzoska Pius., Kong Benjamin., Li Rui., West Robert B., van de Vijver

487   Marc J., Sukumar Saraswati., Chang Howard Y.(2010). Long non-coding RNA

488   HOTAIR reprograms chromatin state to promote cancer metastasis. Nature,

489   464(7291), 1071-6. doi:10.1038/nature08975

490   37. Huarte Maite., Rinn John L.(2010). Large non-coding RNAs: missing links in

491   cancer? Hum Mol Genet, 19(null), R152-61. doi:10.1093/hmg/ddq353

492   38. Prensner John R., Iyer Matthew K., Balbin O Alejandro., Dhanasekaran Saravana

493   M., Cao Qi., Brenner J Chad., Laxman Bharathi., Asangani Irfan A., Grasso

494   Catherine S., Kominsky Hal D., Cao Xuhong., Jing Xiaojun., Wang Xiaoju.,

495   Siddiqui Javed., Wei John T., Robinson Daniel., Iyer Hari K., Palanisamy

496   Nallasivam., Maher Christopher A., Chinnaiyan Arul M.(2011). Transcriptome

497   sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated

498   lincRNA implicated in disease progression. Nat Biotechnol, 29(8), 742-9.

499   doi:10.1038/nbt.1914

500   39. Kiran M, Chatrath A, Tang X, Keenan DM, Dutta A.(2019). A Prognostic Signature for

501   Lower Grade Gliomas Based on Expression of Long Non-Coding RNAs. Molecular

502   Neurobiology volume 56, pages 4786–4798 doi:10.1007/s12035-018-1416-y

503   40. Zheng J, Zhou Z, Qiu Y, Wang M, Yu H, Wu Z, Wang X, Jiang X.(2021). A Prognostic

504   Ferroptosis-Related lncRNAs Signature Associated With Immune Landscape and

505   Radiotherapy Response in Glioma. Front. Cell Dev. Biol. doi:10.3389/fcell.2021.675555

506   41. Munschauer Mathias., Nguyen Celina T., Sirokman Klara., Hartigan Christina R.,

507   Hogstrom Larson., Engreitz Jesse M., Ulirsch Jacob C., Fulco Charles P.,

508   Subramanian Vidya., Chen Jenny., Schenone Monica., Guttman Mitchell., Carr

509   Steven A., Lander Eric S.(2018). The NORAD lncRNA assembles a topoisomerase

510   complex   critical   for   genome   stability.   Nature,   561(7721),   132-136.

511   doi:10.1038/s41586-018-0453-z

512   42. Hu Wang Lai., Jin Lei., Xu An., Wang Yu Fang., Thorne Rick F., Zhang Xu Dong.,

513   Wu Mian.(2018). GUARDIN is a p53-responsive long non-coding RNA that is

514   essential for genomic stability. Nat Cell Biol, 20(4), 492-502. doi:10.1038/s41556-

515   018-0066-7

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

Figure 1

Technology road map of the research

Figure 2

Selection of lncRNAs associated with genomic instability in GC patients and the demonstration of their target genes.

(A) Heat map of the expression of 40 of the most signiûcantly diûerent (20 each of up- and down-regulated expression) lncRNAs in GU and GS groups. (B) Unsupervised clustering of 896 GC patients based on 91 GIrlncRNAs expression patterns. The red cluster on the left is a GS-like cluster, and the blue one on the right is a GU-like cluster. (C) Comparative box plots of cumulative mutation counts in somatic cells. The number of mutations in GU-like group was significantly higher than that in the GS-like group (P<0.001, Mann-Whitney U test). (D) Co-expression network of lncRNAs and mRNAs associated with genomic instability based on Pearson correlation coefficients. Red circles represent lncRNAs, and blue circles represent mRNAs.

Figure 3

LASSO analysis to evaluate and validate the predictive performance of genomic instability-related lncRNA signature (GIrLncSig) on the overall survival of GC patients in the training set.

552 (A) The distribution of Lasso coefficients is plotted. When log Lambda equals -2.7, 17

553 variables are retained. (B) Distribution of partial likelihood deviation of Lasso coefficients.

554 Seventeen variables were retained when bias likelihood deviation was minimized (log

555 Lambda = -2.7). (C) Log-rank test, P<0.05. (D) ROC curves of GIrLncSig for predicting 1-

556 year, 3-year, and 5- year survival in the training set. (E) A set of risk maps, including risk

557 score maps, survival distribution maps, and lncRNAs expression heatmaps, were used for the

558 training set. As the GIrLncSig score increased, the expression of lncRNAs and patient death

559 rate also changed.(F) Box plots comparing somatic mutation counts between high and low-

560 risk groups in the training set (Mann-Whitney U test, P<0.01).

Figure 4

562 Performance evaluation of GIrLncSig in TCGA set and testing set of GC patients and

563 GEO dataset was subjected to external validation. Kaplan-Meier survival curves for

564 patients in high-risk and low-risk groups classified by GIrLncSig score in testing set.

565 (A) Patients in the low-risk group had prolonged survival compared with the high-risk group

566 (log-rank test, P<0.05). ROC curves of GIrLncSig predicting 1-year and 3-year survival in

567 the testing set. (B) Regarding risk score plots, survival distribution plots, and lncRNAs

568 expression heatmaps for the testing set, (C) the expression of lncRNAs and patient mortality

569 changed with increasing GIrLncSig scores. The box plot compares the number of somatic

570 mutations in the

571 high-risk and low-risk groups in the testing set (D) (Mann-Whitney U test,

572 P<0.01). LINC01116 expression in high-grade and low-grade gliomas in GEO dataset (E)

573 and expression of CRNDE (F). Survival analysis was performed according to grouping of

574 high expression of LINC01116 (G) and CRNDE (H), and there was a difference in survival

575 between the two groups (log-rank test, P = 0.012, P < 0.001).

Figure 5

577 Patients in high-risk and low-risk groups of GC were clinically stratified by age,

578 gender, and tumor grade for survival differences.

579   Kaplan-Meier survival curves of patients in high-risk and low-risk groups were analyzed for

580   seven clinically stratified subgroups, including tumor grade II (A), tumor grade III (B), tumor

581   grade IV (C), female (D), male (E), high-risk group (F), and low-risk group (G). In all

582   clinically stratified subgroups, patients in the low-risk group had better survival outcomes

583   than those in high-risk group (log-ranch test, P<0.001).

Figure 6

585   Correlation analysis of GInLncSig and IDH1 mutation status.

586   Box plots of the proportion of IDH1 mutations in all glioma patients (A) and patients with

587   low-grade gliomas (C) in high-risk and low-risk groups (Chi-square test, P<0.05). Kaplan-

588   Meier survival curves for IDH1 mutation and GIrLncRNAs grouping of all glioma patients

589   (B) and low-grade glioma patients (D) revealed statistically significant differences in overall

590   survival between the four groups (log-ranking test, P<0.001). (E) ROC curves for 1-year

591   survival prediction for GIrLncSig and two other existing signatures.

Figure 7

593   RT-PCR validation of cell lines and clinical samples.

594   (A) The expression of eight risk factors lncRNAs in cellular validation differed between

595   tumor and normal cells (unpaired T-test, P<0.05). (B) The expression of eight risk factor

596   lncRNAs in clinical samples (Mann-Whitney test).

Figure 8

598   Correlation of eight risk factors lncRNAs and cell phenotypes of patients with glioma

599   in the TCGA database.

600   (A) The correlation between eight risk factor lncRNAs and cell proliferation marker mki67.

601   (B) Correlation with cell proliferation marker PCNA. (C) Correlation with the mesenchymal

602   transformation gene CDH2 and (D) correlation with the mesenchymal transformation gene

603   Vimentin (Pearson's p < 0.001).

604 **Table 1**

605 qPCR primers designed to amplify mRNA of lncRNAs in GIrlncSig as risk factor.

606 **Table 2**

607 Clinicopathological features of glioma patients in each set.

608 **Table 3**

609 The 17 prognostic-related GIrlncRNAs obtained by LASSO analysis.

610

Figure 1

Technology road map of the research.

Technology road map of the research.

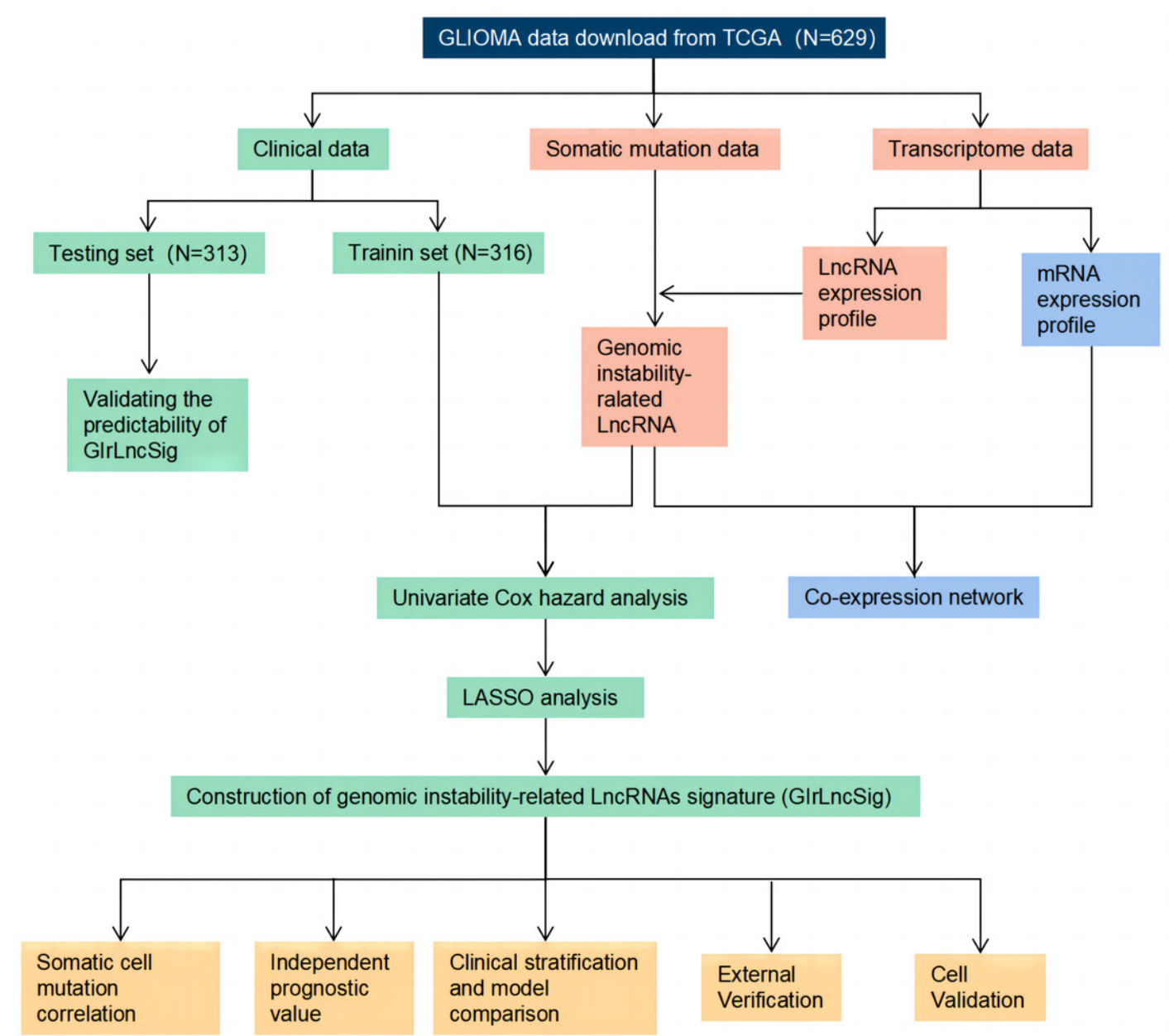



Figure 2

Selection of lncRNAs associated with genomic instability in GC patients and demonstration of their target genes.

(A) Heat map of the expression of 40 of the most significantly different (20 each of up- and down-regulated expression) lncRNAs in GU and GS groups. **(B)** Unsupervised clustering of 896 GC patients based on 91 GIrlncRNAs expression patterns. The red cluster on the left is a GS-like cluster, and the blue one on the right is a GU-like cluster. **(C)** Comparative box plots of cumulative mutation counts in somatic cells. The number of mutations in GU-like group was significantly higher than that in the GS-like group (P<0.001, Mann-Whitney U test). **(D)** Co-expression network of lncRNAs and mRNAs associated with genomic instability based on Pearson correlation coefficients. Red circles represent lncRNAs, and blue circles represent mRNAs.

A

B

C

D

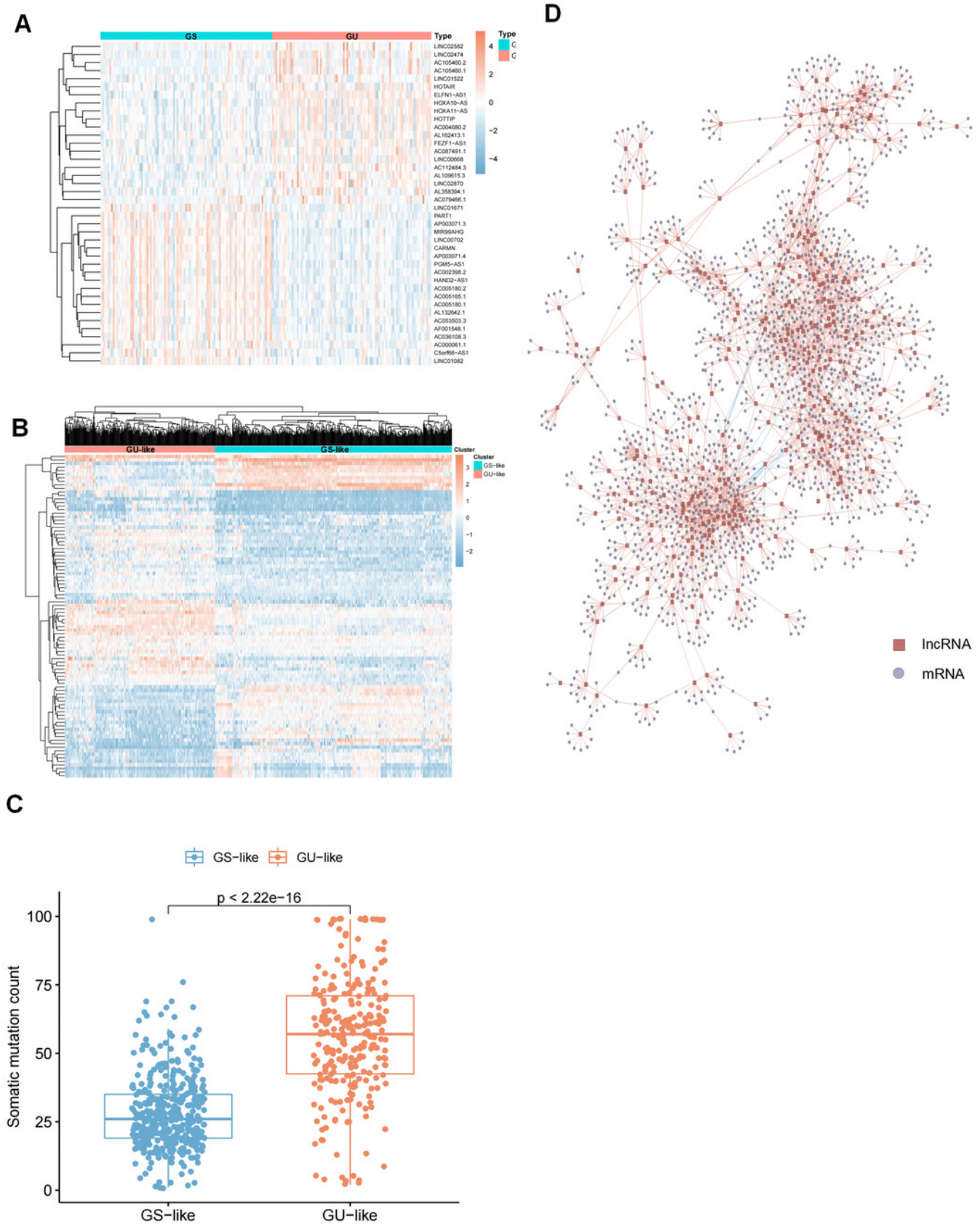



Figure 3

LASSO analysis and to evaluate and validate the predictive performance of genomic instability-related lncRNA signature (GIrLncSig) on the overall survival of GC patients in the training set.

(A) Distribution of lasso coefficients is plotted. When Log Lambda equals -2.7, 17 variables are retained.(B) Distribution of partial likelihood deviation of lasso coefficients. Seventeen variables were retained when bias likelihood deviation was minimized (Log Lambda = -2.7). (C) (log-rank test, P<0.05). (D) ROC curves of GIrLncSig for predicting 1-year, 3-year, and 5-year survival in the training set. (E) A set of risk maps, including risk score maps, survival distribution maps, and lncRNAs expression heatmaps, were used for the training set. As GIrLncSig score increased, the expression of lncRNAs and patient death rate also changed.(F) Box plots comparing somatic mutation counts between high and low-risk groups in the training set(Mann-Whitney U test, P<0.01).

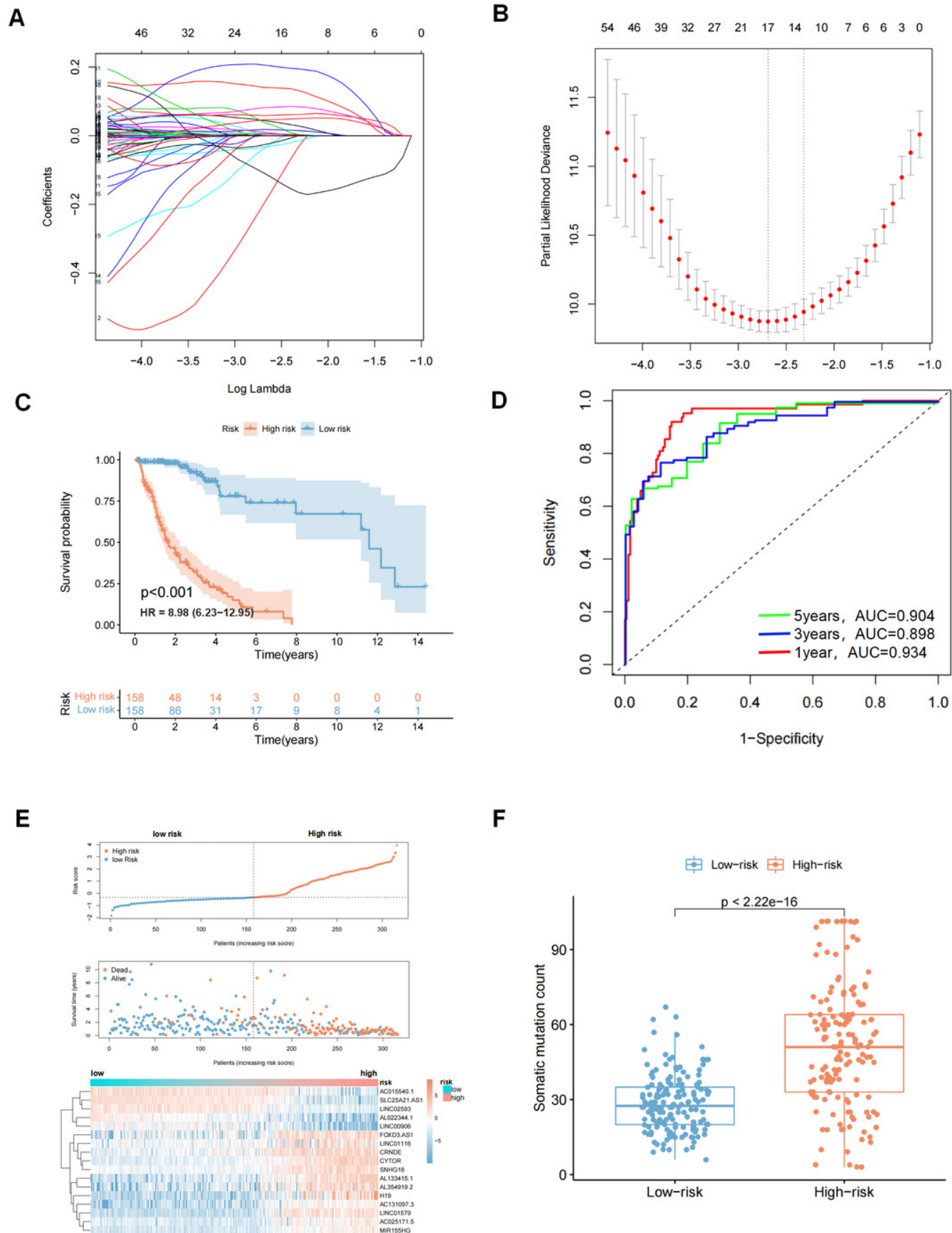



Figure 4

Performance evaluation of GIrLncSig in TCGA set and testing set of GC patients and GEO dataset was subjected to external validation. Kaplan-Meier survival curves for patients in high-risk and low-risk groups classified by GIrLncSig score in testing set.

(A) . Patients in the low-risk group had prolonged survival compared with the high-risk group (log-rank test, P<0.05). ROC curves of GIrLncSig predicting 1-year and 3-year survival in the testing set (B). Regarding risk score plots, survival distribution plots, and lncRNAs expression heatmaps for the testing set (C) , the expression of lncRNAs and patient mortality changed with increasing GIrLncSig scores. The box plot compares the number of somatic mutations in high-risk and low-risk groups in the testing set (D). (Mann-Whitney U test, P<0.01).LINC01116 expression in high-grade and low-grade gliomas in GEO dataset (E) and expression of CRNDE (F). Survival analysis was performed according to grouping of high expression of LINC01116 (G) and CRNDE (H), and there was a difference in survival between the two groups (log-rank test, P = 0.012, P < 0.001).

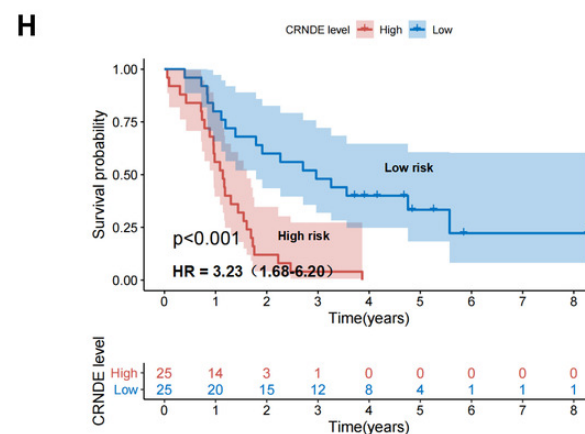
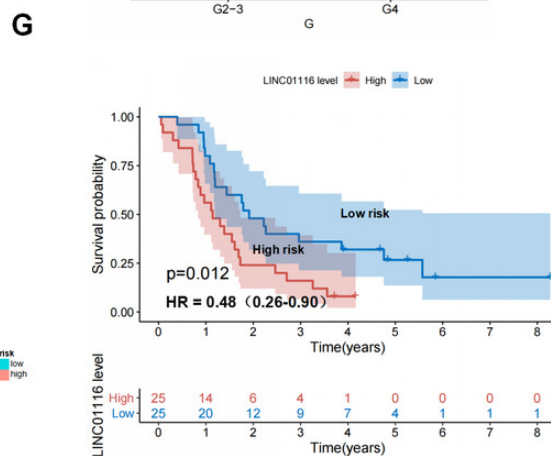
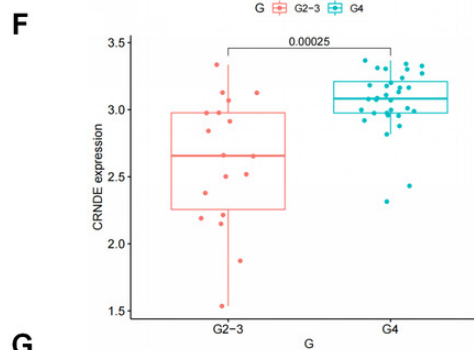
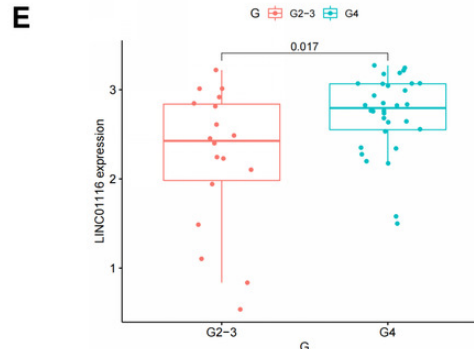
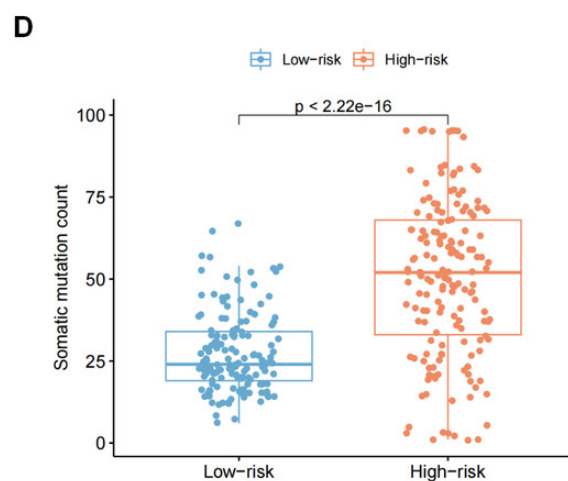
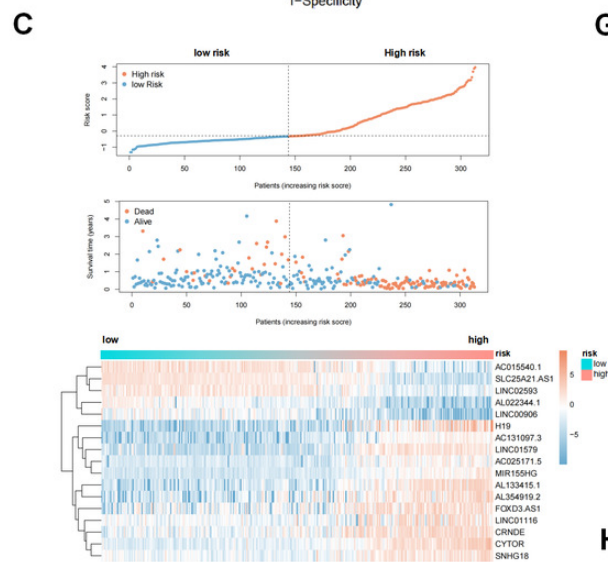
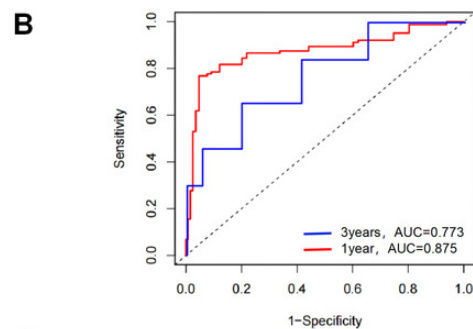
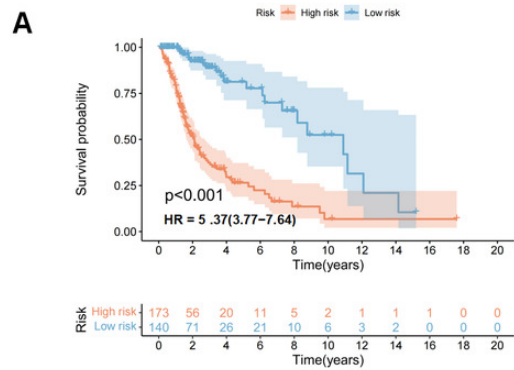



Figure 5

Patients in high-risk and low-risk groups of GC were clinically stratified by age, gender, and tumor grade for survival differences.

Kaplan-Meier survival curves of patients in high-risk and low-risk groups were analyzed for seven clinically stratified subgroups, including tumor grade II (A), tumor grade III (B), tumor grade IV (C), female (D), male (E), high-risk group (F), and low-risk group (G). In all clinically stratified subgroups, patients in low-risk group had better survival outcomes than those in high-risk group (log-ranch test, P<0.001).

Figure 6

Correlation analysis of GInLncSig and IDH1 mutation status.

Box plots of the proportion of IDH1 mutations in all glioma patients (A) and patients with low-grade gliomas (C) in high-risk and low-risk groups (Chi-square test, P<0.05). Kaplan-Meier survival curves for IDH1 mutation and GIrLncRNAs grouping of all glioma patients (B) and low-grade glioma patients (D) revealed statistically significant differences in overall survival between the four groups (log-ranking test, P<0.001). (E) ROC curves for 1-year survival prediction for GIrLncSig and two other existing signatures.

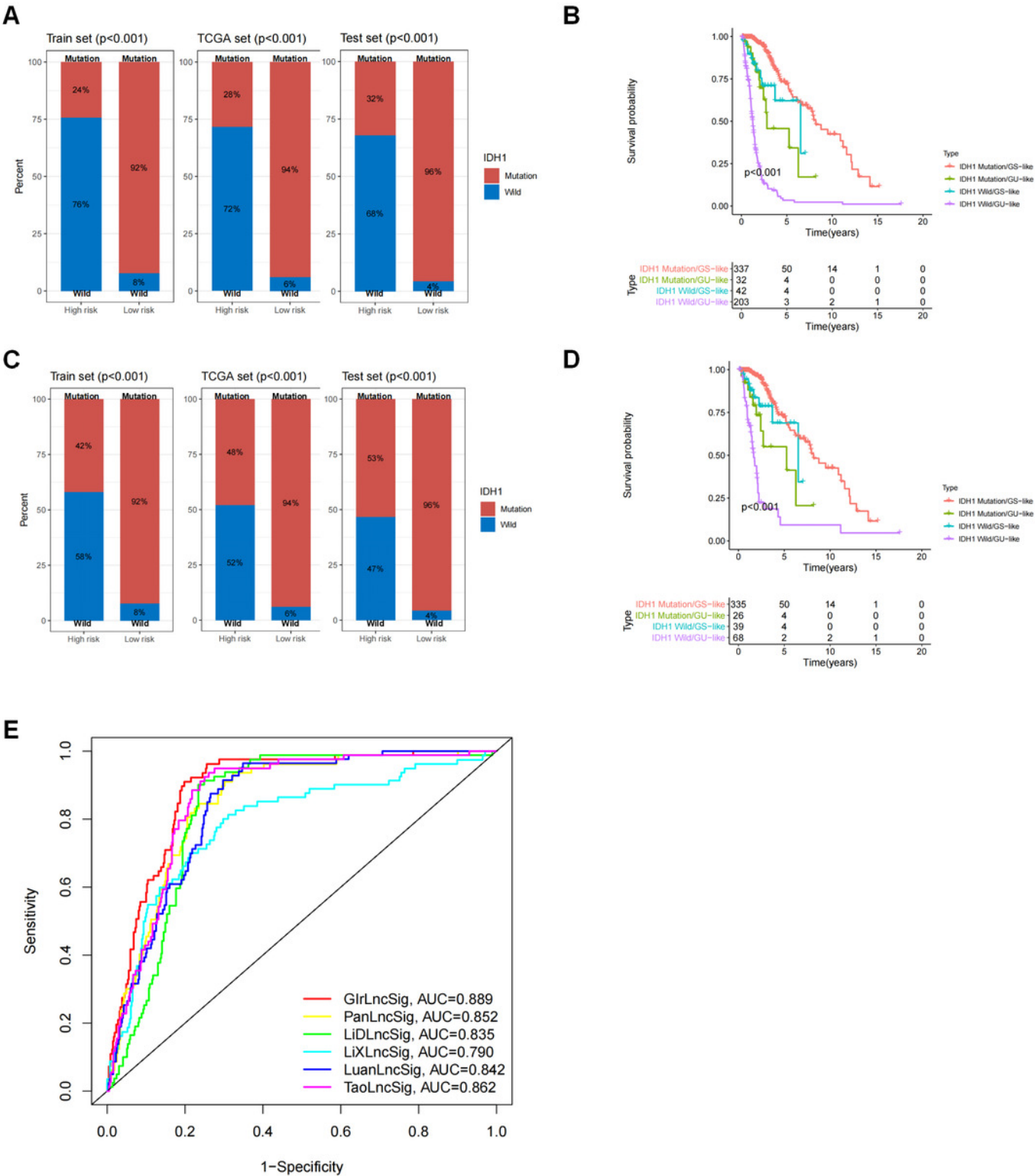
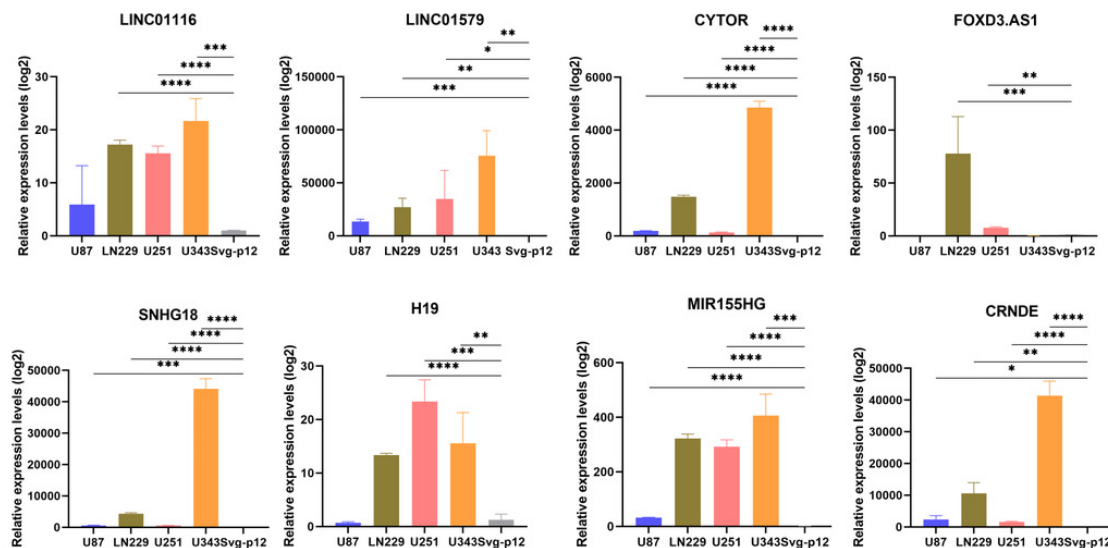



Figure 7

RT-PCR validation of cell lines and clinical samples.

(A) The expression of 8 risk factors lncRNAs in cellular validation differed between tumor and normal cells (unpaired T-test, P<0.05). (B) The expression of 8 risk factors lncRNAs in clinical samples (Mann Whitney test).

A

B

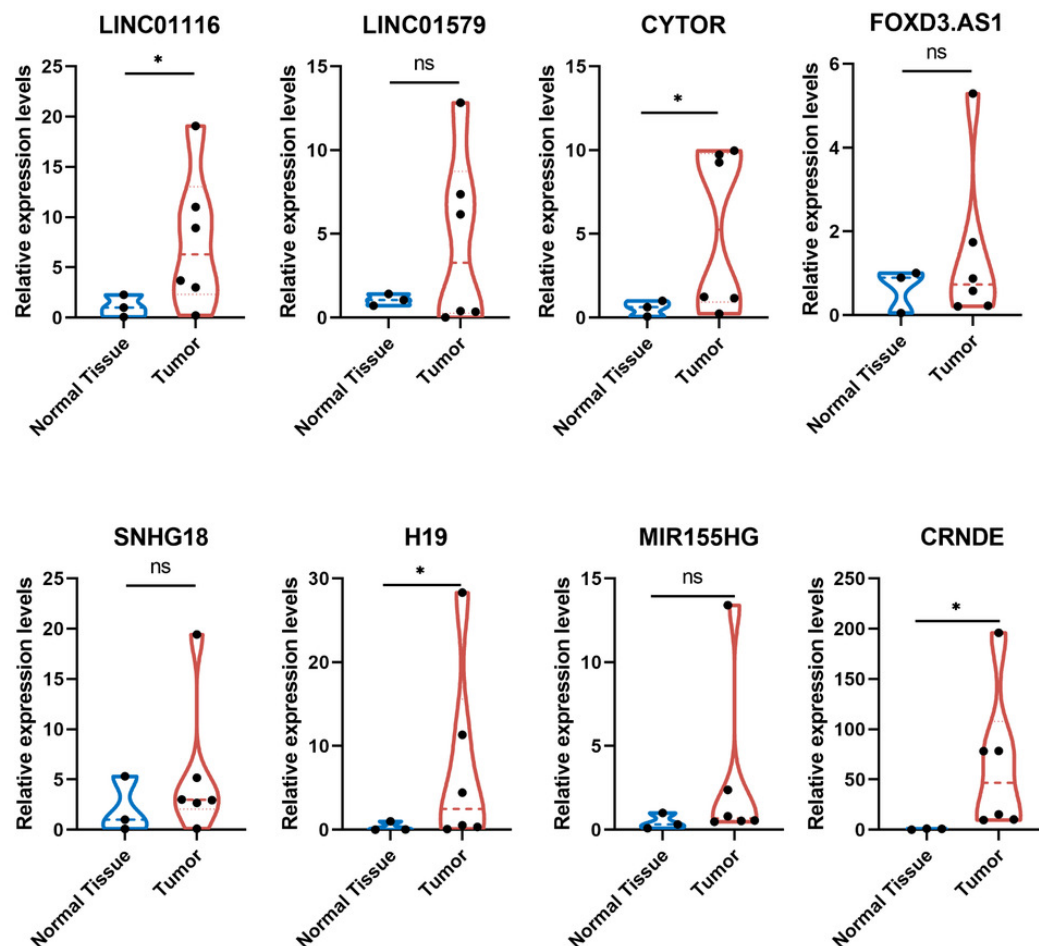



Figure 8

Correlation of eight risk factor lncRNAs and cell phenotypes of patients with glioma in the TCGA database

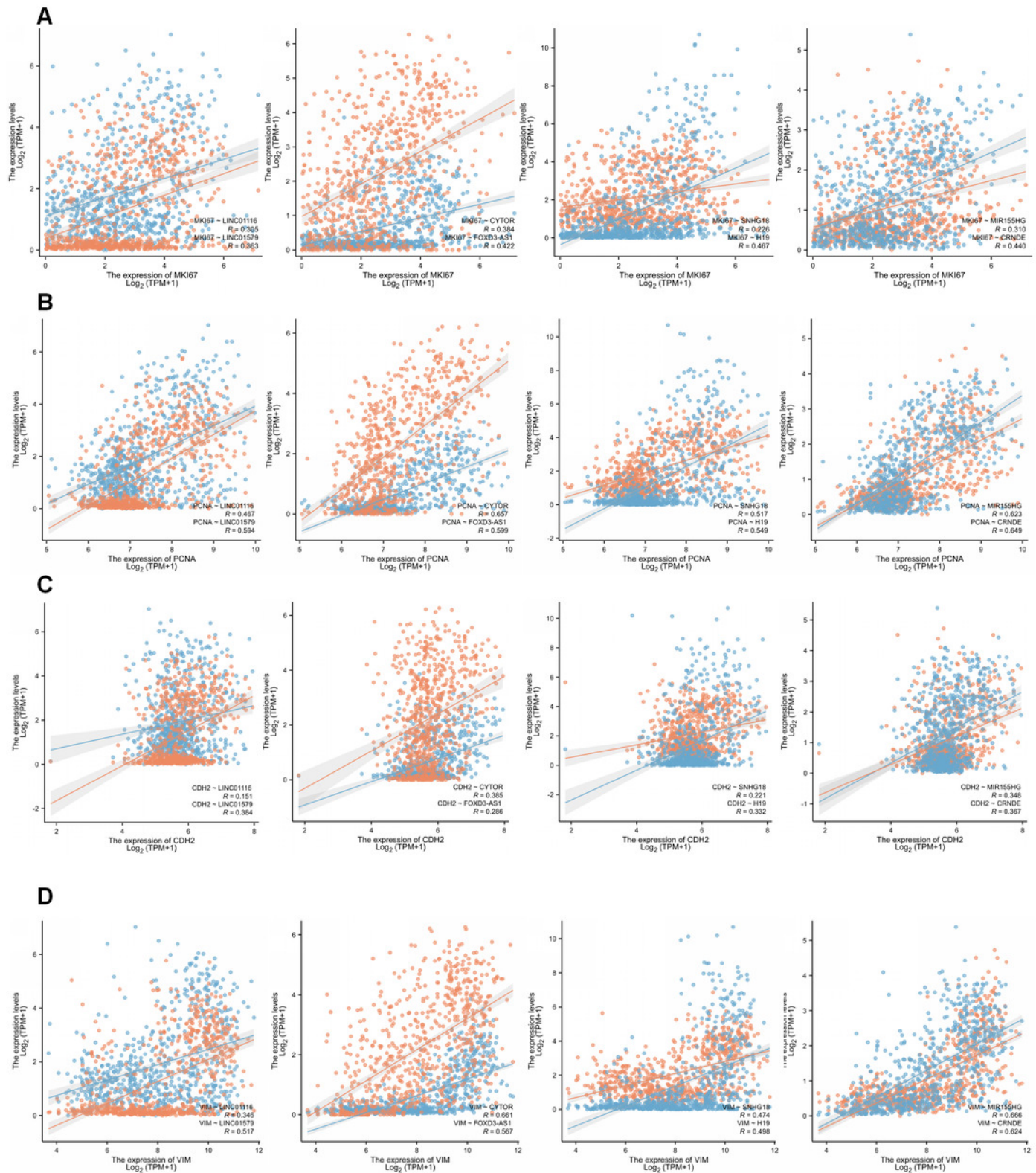
(A) The correlation between 8 risk factor lncRNAs and cell proliferation marker mki67. (B) Correlation with cell proliferation marker PCNA. (C) Correlation with the mesenchymal transformation gene CDH2 and (D) correlation with the mesenchymal transformation gene Vimentin (Pearson p < 0.001).

qPCR primers designed to amplify mRNA of lncRNAs in GIrlncSig as risk factor.

Table 1 qPCR primers designed to amplify mRNA of lncRNAs in GlrlncSig as risk factor.

LncRNA	Forward	Reverse
LINC01579	5'-TCCCAGTGAAGAGAGAGCGA-3'	5'-CTAAGTTCCACGTCACGGCT-3'
LINC01116	5'-GAATGGCAAAGCACTTGGGG-3'	5'-AGCTCTCCTTGCAGGTAGGT-3'
MIR155HG	5'-AGGGGTTTTTGCCTCCAACT-3'	5'-TCTTTGTCATCCTCCCACGG-3'
CYTOR	5'-TTCCAACCTCCGTCTGCATC-3'	5'-AATGGGAAACCGACCAGACC-3'
H19	5'-GACATCTGGAGTCTGGCAGG-3'	5'-CTGCCACGTCCTGTAACCAA-3'
SNHG18	5'-TGCACTTTGCCACTGCTACA-3'	5'-GGGGAATGTGGTTCTCCCTT-3'
FOXD3.AS1	5'-AAGAGTAAGAGCAGCGCACC-3'	5'-ACCTGAGTGGTTTGGTTGGG-3'
CRNDE	5'-ATTCAGCCGTTGGTCTTTGA-3'	5'-CTTCTGCGTGACAACTGAGGA-3'

Clinicopathological features of glioma patients in each set

1 **Table 2.** Clinicopathological features of glioma patients in each set.

Covariates	Type	Total (n=629)	Test (N=313)	Train (N=316)	Pvalue
Age	<=65	547(86.96%)	276(88.18%)	271(85.76%)	0.4338
	>65	82(13.04%)	37(11.82%)	45(14.24%)	
Gender	Female	270(42.93%)	136(43.45%)	134(42.41%)	0.8538
	Male	359(57.07%)	177(56.55%)	182(57.59%)	
Tumor Grade	G2	189(30.05%)	101(32.27%)	88(27.85%)	0.5808
	G3	176(27.98%)	88(28.12%)	88(27.85%)	
	G4	264(41.97%)	124(39.62%)	140(44.3%)	

Chi-squared test, P < 0.05 means significantly different.

The 17 prognositic-related GIrlncRNAsobtained by LASSO analysis.

Gene	Description	Coef
LINC01579	Long intergenic non-protein coding RNA 1579	0.0441
AL022344.1	Homo sapiens chromosome 10 clone XX-Y214H10, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces	-0.1978
AC025171.5	Homo sapiens chromosome 5 clone CTD-2035E11, WORKING DRAFT SEQUENCE, 16 unordered pieces	0.0388
LINC01116	Long intergenic non-protein coding RNA 1116	0.0007
MIR155HG	MIR155 host gene	0.1404
AC131097.3	Homo sapiens BAC clone RP11-789L24 from 2, complete sequence	0.2059
LINC00906	Long intergenic non-protein coding RNA 906	-0.0370
CYTOR	Cytoskeleton regulator RNA(C2orf59; LINC00152)	0.0588
AC015540.1	Homo sapiens clone RP11-385G16, *** SEQUENCING IN PROGRESS ***, 89 unordered pieces	-0.0336
SLC25A21.AS1	SLC25A21 antisense RNA 1	-0.1085
H19	H19 imprinted maternally expressed transcript	0.0061
AL133415.1	Homo sapiens chromosome 10 clone RP11-124N14, *** SEQUENCING IN PROGRESS ***.	0.0819
SNHG18	Small nucleolar RNA host gene 18	0.0050
FOXD3.AS1	FOXD3 antisense RNA 1	0.0181
LINC02593	Long intergenic non-protein coding RNA 2593	-0.0263
AL354919.2	Long intergenic non-protein coding RNA AL354919.2	0.0103
CRNDE	Colorectal neoplasia differentially expressed gene	0.0461