Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mutator-derived lncRNAs drive genomic instability and represent a kind of prognostic marker for glioma

Shenglun Li
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Yujia Chen
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Yuduo Guo
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Jiacheng Xu
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Xiang Wang
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Weihai Ning
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Lixin Ma
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Yanming Qu Qu
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Mingshan Zhang Zhang
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University
Hongwei Zhang (zhanghongwei@ccmu.edu.cn)
Department of Neurosurgery,Sanbo Brain Hospital, Capital Medical University

Research Article

Keywords: genomic instability, glioma, long non-coding RNAs, prognostic model, cancer biomaker

Posted Date: July 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1728794/v1

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

Abstract

Glioma is the most common malignant tumor in brain and it has abundant long non-coding RNAs (lncRNAs), However, the specific role of lncRNAs and its mechanism in glioma remain unclear. LncRNA is a valuable cancer biomarker and they have been reported to be involved in genomic instability. Moreover, lncRNAs related to genomic instability has not been reported in glioma. In this study, we performed a prognostic model constructed with genomic instability-associated lncRNAs and further validated it with mutation correlation analysis, model comparison, independent prognostic value analysis, clinical stratification, examination of the external data set and cell line validation. This model provides us with a new kit for evaluating the prognosis of glioma.

Introduction

Genetic abnormality and mutation are crucial factors in cancer, resulting in the loss of balance in the nucleotide chain of most human tumors[1]. Almost all tumor cells were associated with genomic instability, for example, more than two-thirds of human tumors gained extra or lose whole chromosomes during cell division as reported[2]. For instance, abnormal expression of CDK4, Ras downstream prokaryotic gene, and BRAF gene could lead to gene instability[3,4]. As a marker of cancer evolution, genomic instability (caused mainly by mutations in DNA repair gene) promoted cancer progression and had been identified as a key prognostic factor[5,6,7]. In addition, in contrast to different numbers of genetic mutations, various cancer types exhibited different patterns of somatic cell mutation, indicating specific carcinogenic mechanisms in tissues and cells[8]. Therefore, it is important to identify the potential molecular characteristics associated with cancer genomic instability and explore their clinical significance.

Glioma, commonly known as glioma cerebri (GC), approximately accounts 81% of malignant primary brain tumors[9]. Glioblastoma (GBM), the most malignant among them, with an average survival time of just 16 months for patients[10]. GBM is characterized as uncontrolled cell proliferation, diffuse infiltration, necrosis, intense angiogenesis, strong resistance to apoptosis and rampant gene instability[11]. Although glioma's etiology is still unclear, but exposure to high doses of ionizing radiation; and genetic mutations associated with a high penetrance of rare syndromes had been identified as two related risk factors in glioma tumorigenesis. It showed that it was critical to identify biomarkers associated with genomic instability to accurately evaluate the clinical prognosis of glioma patients. .

Long non-coding RNAs-a kind of RNA with over 200 nucleotides and incapable of coding proteins[15]. Previous research has demonstrated that lncRNAs can play a role in various life events[16,17] and their abnormal expression could affect cell proliferation, tumor progression, or metastasis[18]. Ling et al.[19] reported a novel lncRNA CCAT2 containing rs6983267 SNP had high overexpression in microsatellite-stabilized colorectal cancer and could promote tumor growth metastasis, and cause chromosome instability. It indicated that lncRNAs involved in the biological process of gene modification contribute to

genomic instability and tumor progression, but the study about the genomic instability lncRNAs was still absent.

In this study, we identified a set of lncRNAs associated with genomic instability, constracted a genomic instability-related lncRNA signature (GIrLncSig) and validated its prognostic significance in glioma patients.The results showed that this signature had a great prediction role in glioma patients prognosis.

Materials And Methods

Data collection

We collected GBM and low-grade glioma (LGG) patients clinical characteristics, transcription group data and somatic cell mutation data from cancer genome atlas database (https://portal.gdc.cancer.gov/). These data were matched by samples' name. Patients with no information about survival or less than 30 days duration were excluded to eliminate interference from non-cancer causes. We differentiated mRNA and lncRNA using human genome profiles, and 629 samples retaining paired lncRNA and mRNA expression profiles, survival information, somatic mutation information and common clinicopathological features were obtained from HUGO Gene Nomenclature Committee (HGNC2) database (https://www.genenames.org/) for further study. All glioma patients were randomly divided into two groups, namedas training sets and test sets. A total of 316 patients in training sets were used to identify the prognostic features of lncRNA and establish a risk model for the outcome. The test set included 313 patients was used to independently validate the performance of prognostic risk model.We downloaded GSE43378 from Gene Expression Matrix data set (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi? acc = GSE43378) asexternal validation to test the model.

Technical route

The process of this study was displayed in Fig. 1. After collecting data, we analyzed data from somatic cell mutations and transcription groups to obtain genomic instability-related lncRNAs (GIrlncRNAs). The relationship between GIrlncRNAs and mRNA was analyzed by co-expression analysis. Next, we randomly divided patient cohorts into training and testing sets. Furthermore, Cox regression and lasso regression analysis of GIrlncRNAs were conducted to construct prognostic signature. The signature was evaluated by mutation correlation analysis, model comparison, independent prognostic value analysis, clinical stratification, examination of the external data set and cell line validation.

Identification of GIrlncRNAs

To identify GIrlncRNAs, a method derived from Mutator Hypothesis was applied[20]. The patients with the highest cumulative mutation count and the lowest 25% were assigned to genome unsteady-like (GU) and genome steady-like (GS) groups. The average expression of lncRNAs between the two groups were compared by Wilcoxon rank-sum test in limma package of R software. The cut-off thresholds were intended to be |fold change|>2.0 and false discovery rate (FDR) < 0.05.

Construction of GIrLncSig

The “survival” software package of R software was used to conduct univariate Cox regression analysis on the training set to assess the relationship between the expression level of GIrlncRNAs and the overall survival time of patients. The least absolute shrinkage and selection operator (LASSO) regression algorithm was used to further screen candidate GIrlncRNAs to constructed the GIrlncRNAs prognostic signature (GIrLncSig). The following formula based on a combination of the Cox coefficient and gene expression was used to calculate the signature risk score.

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

GIrLncSig was the prognostic risk score of glioma patients. Ei represented the expression level of lncRNAi in patients and coefi represented the coefficient of lncRNAi. The median GIrLncSig score was used as the risk cut-off point to divide glioma patients into low-risk and high-risk groups. The survival curves of the two groups were plotted by Kaplan-Meier method using "Survminer" and "Survival" package in R language, and the log-rank sum test obtained $p < 0.05$ was considered significant..

Real Time-PCR validation of cell lines

Cell lines U87, U251, LN229, U343 cell lines of human glioblastoma and immortalized cell line SVGp12 were used for cellular level validation of lncRNA in the model.

Cell line culture conditions

DMEM medium with 10% fetal bovine serum, 100 U/ml penicillin and 100 U/ml streptomycin, placed at 37 ℃ in a 5% CO2 incubator.

The collected cells were added with appropriate amount of 1 ml of Trizol (Invitrogen, Waltham, Massachusetts) to extract total cellular RNA, and the absorbance value of RNA at 260nm was measured using a Nanodrop 2000 UV spectrophotometer. 1ug of RNA was taken according to the concentration to synthesize cDNA by reverse transcription (New England Biolabs, Ipswich, MA), and SYBR Green (Applied Biosystems, Foster City, CA) method and CFX96 real-time PCR system (Bio-Rad, Hercules, California) were used for real-time polymerase chain reaction (RT-PCR), and Actin was used as an internal control. Amplification was set at 95°C/120 s followed by 39 cycles of 95°C/5 s and 60°C/30 s. The relative expression of RNA was calculated by the $2^{-\Delta\Delta Ct}$ method. Primers were generated by Sangon Biotechnology (Shanghai,China). primers for lncRNA in GIrLncSig were as follows (Table 1).

Table 1
qPCR primers designed to amplify mRNA of lncRNAs in GIrlncSig 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'

Result

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

To identify genomic instability related lncRNAs, we sorted glioma patients with the number of somatic mutation sitesdefined the top 25% (n = 227) as the genome unsteady (GU) group and the last 25% (n = 110) as the genome steady (GS) group. Next, we compared the expression profiles of lncRNAs in these two groups to identify lncRNAs that were significantly different. After screening, we identified 91 differentially expressed lncRNAs (Supplement Table 1), we chose the top fortieth significantly different lncRNAs between GS and GU groups for heatmap plotting (Fig. 2A). These 91 differentially expressed lncRNAs were significantly associated with genomic instability thus these lncRNAs were defined as genomic instability-related lncRNAs (GIrlncRNAs). Consensus cluster analysis was then performed on 896 samples from the cancer genome atlas (TCGA) collection (including GBM 390 and LGG 506), and all samples were divided into two groups based on these differential expression of GIrlncRNAs and the median number of accumulated somatic mutations (Fig. 2B). The group with higher number of accumulated somatic mutations was defined as GU-like group and the group with lower number of accumulated somatic mutations was defined as GS-like group. The two clustered groups had significantly different somatic mutation patterns (Fig. 2C). Considering lncRNAs cannot perform direct biological functions but can regulate mRNAs we analyzed the expression correlation between lncRNAs and their target mRNAs, selected the top ninestrongly correlated mRNAs as target genes,constructed a co-expression network and shown it in Fig. 2D.

Construction of GIrlncRNAs signature in Training set.

629 glioma patients from TCGA project were randomly divided into training dataset (n = 316) and test dataset (n = 313). The clinicopathological characteristics of these patients were shown in Table 2. A total of 80 prognosticrelated lncRNAs were identified based on univariate Cox proportional risk regression in the training set (Supplement Table 2, Supplement Fig. 2). Given that lncRNAs have distinct biological functions and that many lncRNAs interact with each other, and that there are numerous data and inaccurate model construction, Lasso regression analysis and stepwise multifactor Cox proportional risk regression analysis were performed and 17 lncRNAs were screened out as independent prognostic factors to build the prognostic model (Fig. 3A-B, Table 3) A prediction model was finally obtained: genomic instability-related lncRNA signature (GIrLncSig) score = (0.0441217761138621×LINC01579)+ (-0.197814767909076×AL022344.1)+ (0.0387926268692573×AC025171.5)+ (0.00072629147373604 9×LINC01116)+(0.140438109893274×MIR155HG)+ (0.205877895933553×AC131097.3)+(-0.036976991848185 4×LINC00906)+ (0.0588040388194378×CYTOR)+(-0.033630796494406 9×AC015540.1)+ (-0.108512558356276×SLC25A21.AS1)+(0.006132943324606 68×H19)+ (0.0819381463964954×AL133415.1)+(0.004950654935799 31×SNHG18)+ (0.0180933234213193×FOXD3.AS1)+(-0.026311303886406 1×LINC02593)+ (0.0103411968758757×AL354919.2)+(0.046081456238200 4×CRNDE).A positive coefficient for lncRNA suggested that high expression was associated withlong survival timeas a protective factor in glioma. On the contrary, a negative coefficient for lncRNA mean it was a risk factor in glioma. The risk score of each patient in the training set was obtained using GIrLncSig, and then these patients were divided into high-risk and low-risk groups by the median risk score (0.0103411968758757)The survival analysis revealed that low-risk group had significantly better survival than high-risk group (Fig. 3C).The 5-year survival rate was 4.43% in high -risk group and 14.56% in low-risk group. The ROC curve showed that the AUC was 0.934, 0.898, 0.904at 1 year, 3 years, 5 years (Fig. 3D). We sorted patients in training set by their risk score and showed the different expression level of GIrlncRNAs in two groups (Fig. 3E). Patients with high-risk scores exhibited increased expression of risk genes and decreased expression of protective genes, whereas patients with low-risk scores exhibited the opposite. Also, we found a significant difference in somatic mutation patterns between patients in high-risk and low-risk groups, implying the model could well reflect the somatic mutation situation in glioma (Fig. 3F).

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.					

Table 3
The 17 prognositic-related GIrlncRNAsobtained by LASSO analysis.

Gene	Coef
LINC01579	0.044121776
AL022344.1	-0.197814768
AC025171.5	0.038792627
LINC01116	0.000726291
MIR155HG	0.14043811
AC131097.3	0.205877896
LINC00906	-0.036976992
CYTOR	0.058804039
AC015540.1	-0.033630796
SLC25A21.AS1	-0.108512558
H19	0.006132943
AL133415.1	0.081938146
SNHG18	0.004950655
FOXD3.AS1	0.018093323
LINC02593	-0.026311304
AL354919.2	0.010341197
CRNDE	0.046081456

Independent validation of GIrLncSig in the glioma cerebri data set of transcription data and external validation in GEO data set.

To examine the performance of GIrLncSig, it was subsequently tested in an independent TCGA test set of 313 patients. When the test set used the same GIrLncSig and risk thresholds as the training set, 313 patients in the test set were divided into high-risk (n = 173) and low-risk groups (n = 140), with a significant difference in overall survival (Fig. 4A). The overall survival rate was significantly lower in high-risk group than in low-risk group, consistent with the results of the training group. The 5-year survival rate was 7.51% in the high-risk group, lower than the 16.4% in the low-risk group (Fig. 4A). The ROC curve analysis showed the AUC of GIrLncSig was 0.875 for 1 year and 0.773 for 3 years (Fig. 4B). The expression levels of GIrLncSig, patient death distribution counts, and model lncRNA expression were presented in Fig. 4C. A significant difference was observed in the somatic mutation pattern between

patients in high-risk and low-risk groups (Fig. 4D), and this result was similar in the training group. To further validate the prognostic significance of GIrLncSig, cross-platform was performed in other independent datasets from different platforms.GSE43378, a dataset from the Gene Expression Matrix data set, was downloaded to further analyse because of the large sample size and complete clinicopathological features. Therefore, 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 covered by GSE43378. The expression of Lnc01116 and CRNDE had positive correlation with tumor grade (Fig. 4E-F), and the survivaltime was significant difference between high and low expression subgroup of Linc01116 and CRNDE. Figure 4G-H).These results were consistent with those observed in the training set and the test set.

Evaluation Of Independent Prognostic Significance Of Girlncsig And Clinical Stratification Analysis

In TCGA dataset, the prognosis of GIrLncSig was analyzed by adjusting for clinical stratification, including age(> 65 and ≤ 65), gender(male and female), tumor classification(WHO grade , and)and other clinical factors. In all clinical subgroups, the survival rate in low-risk group was higher than that in high-risk group(Fig. 5). This demonstrated that GIrLncSig exhibited significantly independent prognostic prediction valueon the overall survival of glioma patients under different clinical stratification conditions.

GIrLncSig was associated with IDH1 mutation status and comparison between GIrLncSig prediction and other lncRNAs model prediction.

Next, we observed the difference of IDH1 status in thecohort of glioma patients. IDH1 was a security mutant in the nervous region as reported, so we evaluated the connection between GIrLncSig and IDH1mutation status. We compared the differences between high-risk and low-risk groups on the training setand test set, and the results revealed that the proportion of IDH1 mutation was significantly higher in low-risk group than in high-risk group ,no matter in all glioma cerebri (GC) patients (Fig. 6A) or in low-grade glioma patients (Fig. 6C), implying that patients with IDH1 mutations were at lower risk. Following that, we divided the patients in to four subgroups, IDH1 mutation/GS-like, IDH1 mutation/GU-like, IDH1 wild-type/GS-likeand IDH1 wild-type/GU-like to consider both the GIrLncSig and IDH mutation status. As displayed in Figs. 6B and 6D, there were significant difference in survival time among these four groups. It suggested that IDH1 mutation status might be related to the genomic instability and combining them together was better for predicting prognosis in glioma patiens.

We then compared the predictability between GIrLncSig and two recently reported lncRNA signatures for survival prediction using the same TCGA glioma patients. Compared to 10-lncRNA prognostic signature reported by PAN[21] (PanlncSig) and 4-lncRNA prediction signature reported by Li[22] (LilncSig), as displayed in Fig. 6E, our GIrLncSig had an AUC value of 0.889 in 1-year OS, which was more effective than

PanLncSig (AUC value = 0.852) and LilncSig (AUC value = 0.835) in predicting patient survival. In short, the above results confirmed the reliability and effectiveness of GIrLncSig in predicting GC patients.

Cellular Validation Of Girlncsig

To further validate the prognostic significance of GIrLncSig in glioma, we verified eight lncRNAs in GIrLncSig in four glioblastoma cell lines (U87, U251, LN229, and U343) and one normal cell line (SVGp12). The expression levels of these eight lncRNAs were significantly higher in glioblastoma cell lines than those in normal cells (Fig. 7A-H), further validating the prognostic significance of GIrLncSig.

Discussion

Genomic instability is caused by chromosomal segregation errors during mitosis that result in aneuploidy mutations of whole chromosomes in daughter cells or by DNA damage that causes chromosome structure changes, resulting in gene translocations, deletions, inversions, and breaks[23,24]. Genomic changes occur at different levels, from mutations in single or few nucleotides to the gain or loss of entire chromosomes, which may trigger abnormal divisions, multinucleation, and tripolar mitosis[25,26]. Maintaining genetic integrity is critical for cell viability and is accomplished through a complex repair process. When the repair link is defective, genomic instability occurs, resulting in accumulating chromosomal mutations that can cause cancer susceptibility[27]. Genomic instability plays a fundamental and major role in cancer progression and recurrence, implying that the pattern and degree of genomic instability have significant diagnostic and prognostic implications[28,29]. It has been recently demonstrated that lncRNA, a promising tumor biomarker with abnormal expression in tumors associated with disease progression, may serve as a prognostic marker for patients[30,31,32]. Furthermore, recent advances in understanding the functional mechanisms underlying lncRNAs have recognized that lncRNAs are essential for genomic stability[33,34].

Glioma is the most common primary intracranial tumor, and no studies examining lncRNA signature of glioma genomic instability have been published. This study identified a group of GIrlncRNAs in GC and determined their significance in predicting patient survival. We identified 91 GIrlncRNAs and compared their expression levels in different mutation counts. Systematic clustering analysis and subsequent differential analysis of mutation counts confirmed the association of these lncRNAs with genomic instability. Based on co-expression with 91 GIrlncRNAs, we further investigated whether GIrlncRNAs could predict glioma patients clinical outcomes by constructing a GIrLncSig consisting of 17 GIrlncRNAs (LINC01579, AL022344.1, AC025171.5. LINC01116, MIR155HG, AC131097.3, CYTOR, AC015540.1, SLC25A21.AS1, H19, AL133415.1, SNHG18, FOXD3.AS1, LINC02593, AL354919.2, and CRNDE). GIrLncSig classified patients into two risk groups with a significant difference in survival on the Training set, which was validated on the independent Test set. In addition, similar results were found in the external GEO dataset GSE43378. According to these validation results across multiple datasets and technology platforms, GIrLncSig can be an indicator of genomic instability in cancer patients. In addition,

we validated 11 GIrlncRNAs expression level which had known sequences in four glioma cell lines, and the results revealed significant differences in the expression level of eight lncRNAs as risk factorsbetween glioma cell lines and ordinary cells.

Although our study showed there were some links betweengenomic instability and prognosis in patients with glioma, several limitations remain. We lack the validation of clinical samples of our GIrLncSig e to ensure its accuracy and reproducibility. Therefore, further studies are necessary, and additionally its mechanism of action in glioma development and progression remain to be further investigated.

Conclusion

In summary, our study constructed a risk prediction signal consisting of 17 lncRNAs associated with genomic instability. It can not only predict the prognostic of glioma patients and reveal the genomic instability, but also has great significance for the clinical hierarchical management and individualized treatment of glioma patients. And it might be an important tool to further investigate the role of lncRNAs in genomic instability.

Declarations

Ethics approval and consent to participate: This investigation was carried out according to the Declaration of Helsinki.All methods were performed in accordance with the relevant guidelines and regulations.
Consent for publication: Not applicable.Availability of data and materials: All data generated or analysed during this study are included in this published article

Competing interests: The authors declare that they have no competing interests

Funding: Not applicable

Authors' contributions: Li SL and Chen YJ are the co-first authors,have the same contribution.Zhang HW is the corresponding author.All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the design, analysis, writing, or revision of the manuscript.

Acknowledgement: Not applicable

References

1. Hanahan Douglas., Weinberg Robert A.(2011). Hallmarks of cancer: the next generation. Cell, 144(5), 646–74. doi:10.1016/j.cell.2011.02.013
2. Carvalho Claudia M B., Lupski James R.(2016). Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet, 17(4), 224–38. doi:10.1038/nrg.2015.25

3. Kamata Tamihiro., Hussain Jahan., Giblett Susan., Hayward Robert., Marais Richard., Pritchard Catrin.(2010). BRAF inactivation drives aneuploidy by deregulating CRAF. Cancer Res, 70(21), 8475-86. doi:10.1158/0008-5472.CAN-10-0603
4. Cui Yongping., Borysova Meghan K., Johnson Joseph O., Guadagno Thomas M.(2010). Oncogenic B-Raf(V600E) induces spindle abnormalities, supernumerary centrosomes, and aneuploidy in human melanocytic cells. Cancer Res, 70(2), 675–84. doi:10.1158/0008-5472.CAN-09-1491
5. Suzuki, K., Ohnami, S., Tanabe, C., Sasaki, H., Y asuda, J., Katai, H., et al. (2003).The genomic damage estimated by arbitrarily primed PCR DNA fingerprinting is useful for the prognosis of gastric cancer. Gastroenterology 125, 1330–1340. doi: 10.1016/j.gastro.2003.07.006
6. Ottini L., Falchetti M., Lupi R., Rizzolo P., Agnese V., Colucci G., Bazan V., Russo A.(2006). Patterns of genomic instability in gastric cancer: clinical implications and perspectives. Ann Oncol, null(undefined), vii97-102. doi:10.1093/annonc/mdl960
7. Negrini Simona., Gorgoulis Vassilis G., Halazonetis Thanos D.(2010). Genomic instability–an evolving hallmark of cancer. Nat Rev Mol Cell Biol, 11(3), 220–8. doi:10.1038/nrm2858
8. Anandakrishnan Ramu., Varghese Robin T., Kinney Nicholas A., Garner Harold R.(2019). Estimating the number of genetic mutations (hits) required for carcinogenesis based on the distribution of somatic mutations. PLoS Comput Biol, 15(3), e1006881. doi:10.1371/journal.pcbi.1006881
9. Jahangir Moini., PirouzPiran.(2020). Chapter 1-Histophysiology.Functional and Clinical Neuroanatomy,1–49. doi: 10.1016/B978-0-12-817424-1.00001-X.
10. Stupp Roger., Mason Warren P., van den Bent Martin J., Weller Michael., Fisher Barbara., Taphoorn Martin J B., Belanger Karl., Brandes Alba A., Marosi Christine., Bogdahn Ulrich., Curschmann Jürgen., Janzer Robert C., Ludwin Samuel K., Gorlia Thierry., Allgeier Anouk., Lacombe Denis., Cairncross J Gregory., Eisenhauer Elizabeth., Mirimanoff René O., European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups., National Cancer Institute of Canada Clinical Trials Group.(2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med, 352(10), 987 – 96. doi:10.1056/NEJMoa043330
11. Furnari Frank B., Fenton Tim., Bachoo Robert M., Mukasa Akitake., Stommel Jayne M., Stegh Alexander., Hahn William C., Ligon Keith L., Louis David N., Brennan Cameron., Chin Lynda., DePinho Ronald A., Cavenee Webster K.(2007). Malignant astrocytic glioma: genetics, biology, and paths to treatment. Genes Dev, 21(21), 2683 – 710. doi:10.1101/gad.1596707
12. Negrini Simona., Gorgoulis Vassilis G., Halazonetis Thanos D.(2010). Genomic instability–an evolving hallmark of cancer. Nat Rev Mol Cell Biol, 11(3), 220–8. doi:10.1038/nrm2858
13. Ottini L., Falchetti M., Lupi R., Rizzolo P., Agnese V., Colucci G., Bazan V., Russo A.(2006). Patterns of genomic instability in gastric cancer: clinical implications and perspectives. Ann Oncol, null(undefined), vii97-102. doi:10.1093/annonc/mdl960
14. Suzuki Koichi., Ohnami Sumiko., Tanabe Chikako., Sasaki Hiroki., Yasuda Jun., Katai Hitoshi., Yoshimura Kimio., Terada Masaaki., Perucho Manuel., Yoshida Teruhiko.(2003). The genomic

damage estimated by arbitrarily primed PCR DNA fingerprinting is useful for the prognosis of gastric cancer. Gastroenterology, 125(5), 1330-40. doi:10.1016/j.gastro.2003.07.006

15. Mattick John S., Rinn John L.(2015). Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol, 22(1), 5–7. doi:10.1038/nsmb.2942
16. Koziol Magdalena J., Rinn John L.(2010). RNA traffic control of chromatin complexes. Curr Opin Genet Dev, 20(2), 142–8. doi:10.1016/j.gde.2010.03.003
17. Mercer Tim R., Mattick John S.(2013). Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol, 20(3), 300–7. doi:10.1038/nsmb.2480
18. Sanchez Calle Anna., Kawamura Yumi., Yamamoto Yusuke., Takeshita Fumitaka., Ochiya Takahiro. (2018). Emerging roles of long non-coding RNA in cancer. Cancer Sci, 109(7), 2093–2100. doi:10.1111/cas.13642
19. Ling Hui., Spizzo Riccardo., Atlasi Yaser., Nicoloso Milena., Shimizu Masayoshi., Redis Roxana S., Nishida Naohiro., Gafà Roberta., Song Jian., Guo Zhiyi., Ivan Cristina., Barbarotto Elisa., De Vries Ingrid., Zhang Xinna., Ferracin Manuela., Churchman Mike., van Galen Janneke F., Beverloo Berna H., Shariati Maryam., Haderk Franziska., Estecio Marcos R., Garcia-Manero Guillermo., Patijn Gijs A., Gotley David C., Bhardwaj Vikas., Shureiqi Imad., Sen Subrata., Multani Asha S., Welsh James., Yamamoto Ken., Taniguchi Itsuki., Song Min-Ae., Gallinger Steven., Casey Graham., Thibodeau Stephen N., Le Marchand Loïc., Tiirikainen Maarit., Mani Sendurai A., Zhang Wei., Davuluri Ramana V., Mimori Koshi., Mori Masaki., Sieuwerts Anieta M., Martens John W M., Tomlinson Ian., Negrini Massimo., Berindan-Neagoe Ioana., Foekens John A., Hamilton Stanley R., Lanza Giovanni., Kopetz Scott., Fodde Riccardo., Calin George A.(2013). CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. Genome Res, 23(9), 1446-61. doi:10.1101/gr.152942.112
20. Bao Siqi., Zhao Hengqiang., Yuan Jian., Fan Dandan., Zhang Zicheng., Su Jianzhong., Zhou Meng. (2020). 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, 21(5), 1742–1755. doi:10.1093/bib/bbz118
21. Pan Yuan-Bo., Zhu Yiming., Zhang Qing-Wei., Zhang Chi-Hao., Shao Anwen., Zhang Jianmin.(2020). Prognostic and Predictive Value of a Long Non-coding RNA Signature in Glioma: A lncRNA Expression Analysis. Front Oncol, 10(undefined), 1057. doi:10.3389/fonc.2020.01057
22. Li Depei., Lu Jie., Li Hong., Qi Songtao., Yu Lei.(2019). Identification of a long noncoding RNA signature to predict outcomes of glioblastoma. Mol Med Rep, 19(6), 5406–5416. doi:10.3892/mmr.2019.10184
23. Carvalho Claudia M B., Lupski James R.(2016). Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet, 17(4), 224–38. doi:10.1038/nrg.2015.25
24. Wickramasinghe Vihandha O., Venkitaraman Ashok R.(2016). RNA Processing and Genome Stability: Cause and Consequence. Mol Cell, 61(4), 496–505. doi:10.1016/j.molcel.2016.02.001

25. Mackay Hannah L., Moore David., Hall Callum., Birkbak Nicolai J., Jamal-Hanjani Mariam., Karim Saadia A., Phatak Vinaya M., Piñon Lucia., Morton Jennifer P., Swanton Charles., Le Quesne John., Muller Patricia A J.(2018). Genomic instability in mutant p53 cancer cells upon entotic engulfment. Nat Commun, 9(1), 3070. doi:10.1038/s41467-018-05368-1
26. Zhang ShiQi., Pan XiaoYong., Zeng Tao., Guo Wei., Gan Zijun., Zhang Yu-Hang., Chen Lei., Zhang YunHua., Huang Tao., Cai Yu-Dong.(2019). Copy Number Variation Pattern for Discriminating MACROD2 States of Colorectal Cancer Subtypes. Front Bioeng Biotechnol, 7(undefined), 407. doi:10.3389/fbioe.2019.00407
27. Rancoule Chloé., Vallard Alexis., Guy Jean-Baptiste., Espenel Sophie., Sauvaigo Sylvie., Rodriguez-Lafrasse Claire., Magné Nicolas.(2017). [Impairment of DNA damage response and cancer]. Bull Cancer, 104(11), 962–970. doi:10.1016/j.bulcan.2017.09.006
28. Kronenwett Ulrike., Ploner Alexander., Zetterberg Anders., Bergh Jonas., Hall Per., Auer Gert., Pawitan Yudi.(2006). Genomic instability and prognosis in breast carcinomas. Cancer Epidemiol Biomarkers Prev, 15(9), 1630-5. doi:10.1158/1055-9965.EPI-06-0080
29. Mettu Rama K R., Wan Ying-Wooi., Habermann Jens K., Ried Thomas., Guo Nancy Lan.(2010). A 12-gene genomic instability signature predicts clinical outcomes in multiple cancer types. Int J Biol Markers, 25(4), 219–28. doi:10.5301/jbm.2010.6079
30. Gupta Rajnish A., Shah Nilay., Wang Kevin C., Kim Jeewon., Horlings Hugo M., Wong David J., Tsai Miao-Chih., Hung Tiffany., Argani Pedram., Rinn John L., Wang Yulei., Brzoska Pius., Kong Benjamin., Li Rui., West Robert B., van de Vijver Marc J., Sukumar Saraswati., Chang Howard Y.(2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature, 464(7291), 1071-6. doi:10.1038/nature08975
31. Huarte Maite., Rinn John L.(2010). Large non-coding RNAs: missing links in cancer? Hum Mol Genet, 19(null), R152-61. doi:10.1093/hmg/ddq353
32. Prensner John R., Iyer Matthew K., Balbin O Alejandro., Dhanasekaran Saravana M., Cao Qi., Brenner J Chad., Laxman Bharathi., Asangani Irfan A., Grasso Catherine S., Kominsky Hal D., Cao Xuhong., Jing Xiaojun., Wang Xiaoju., Siddiqui Javed., Wei John T., Robinson Daniel., Iyer Hari K., Palanisamy Nallasivam., Maher Christopher A., Chinnaiyan Arul M.(2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol, 29(8), 742-9. doi:10.1038/nbt.1914
33. Munschauer Mathias., Nguyen Celina T., Sirokman Klara., Hartigan Christina R., Hogstrom Larson., Engreitz Jesse M., Ulirsch Jacob C., Fulco Charles P., Subramanian Vidya., Chen Jenny., Schenone Monica., Guttman Mitchell., Carr Steven A., Lander Eric S.(2018). The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. Nature, 561(7721), 132–136. doi:10.1038/s41586-018-0453-z
34. Hu Wang Lai., Jin Lei., Xu An., Wang Yu Fang., Thorne Rick F., Zhang Xu Dong., Wu Mian.(2018). GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. Nat Cell Biol, 20(4), 492–502. doi:10.1038/s41556-018-0066-7

Figures

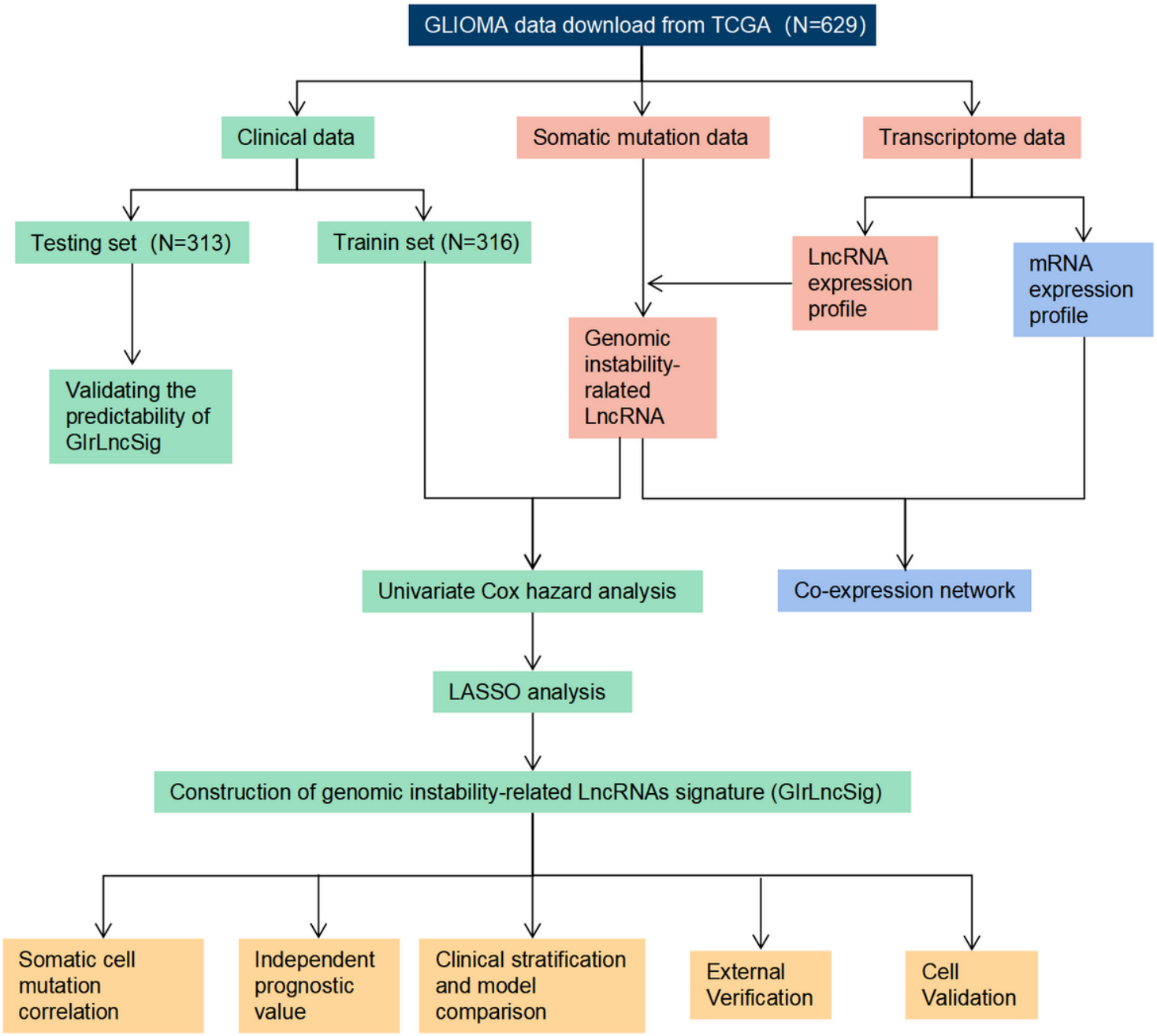

Figure 1

Technology road map of the research.

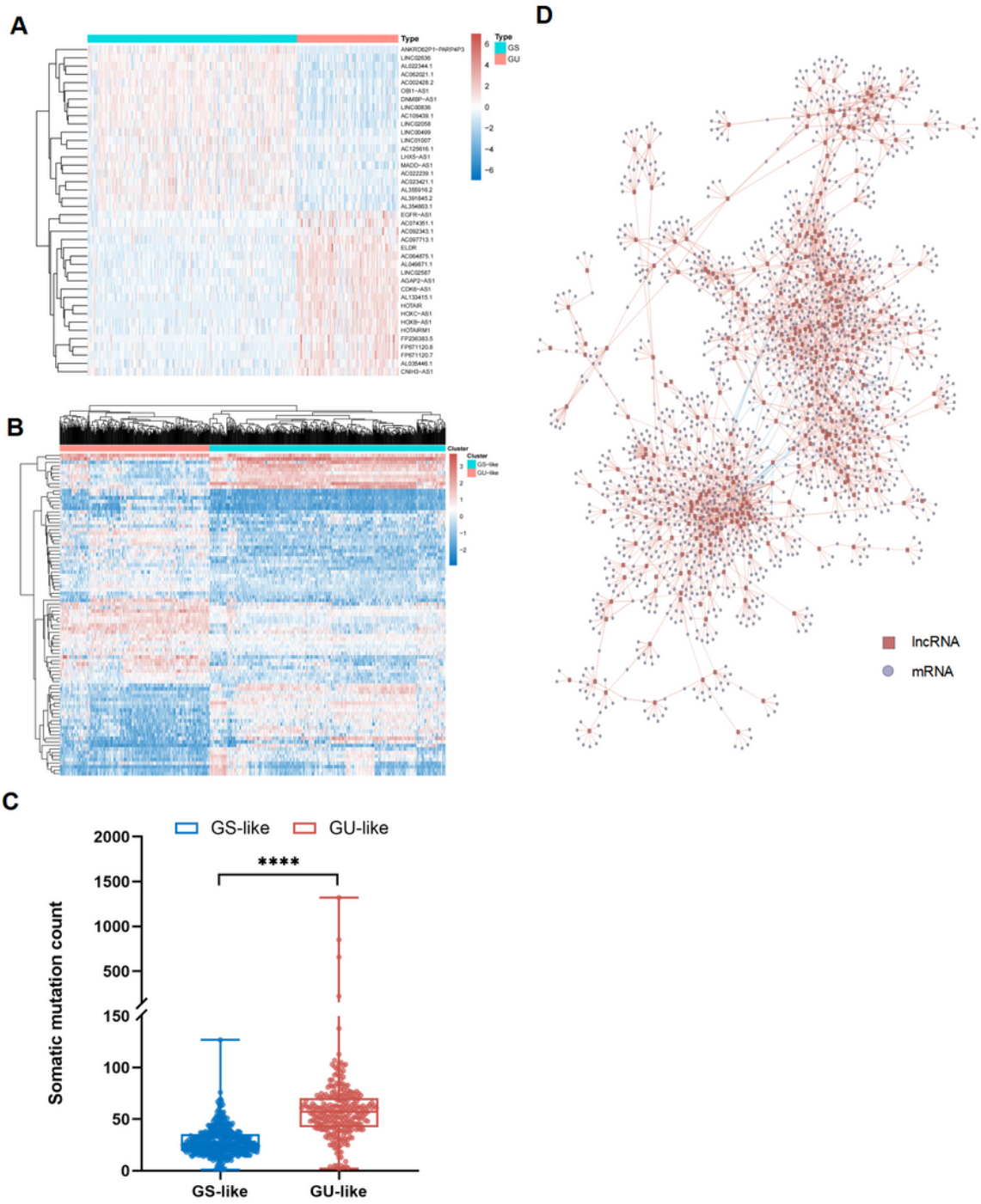

Figure 2

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.

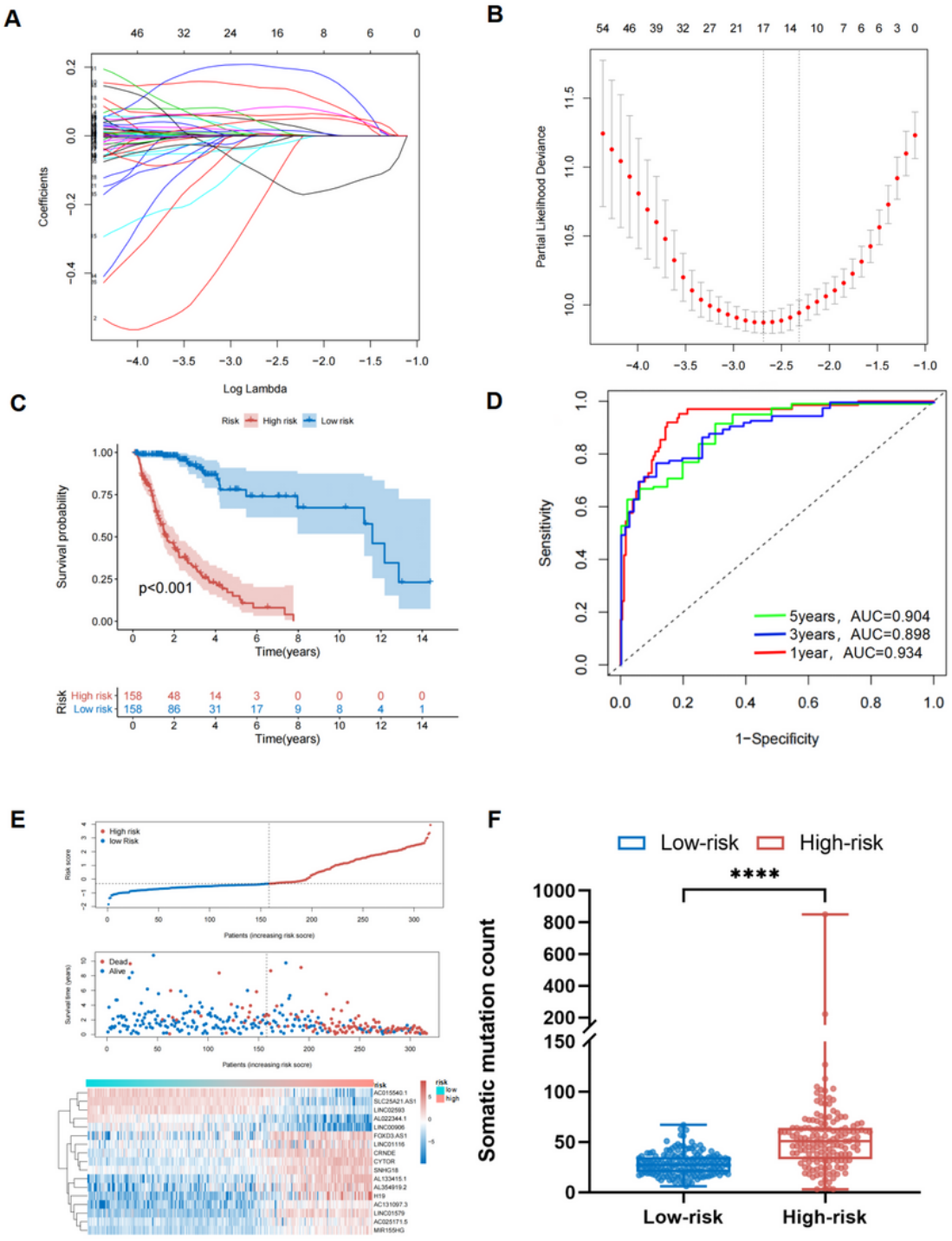

Figure 3

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)** Kaplan-Meier survival curves for patients in high-risk and low-risk groups classified by GIrLncSig score in the training set. Patients in the low-risk group had prolonged survival compared with the high-risk group (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, $P<0.01$).

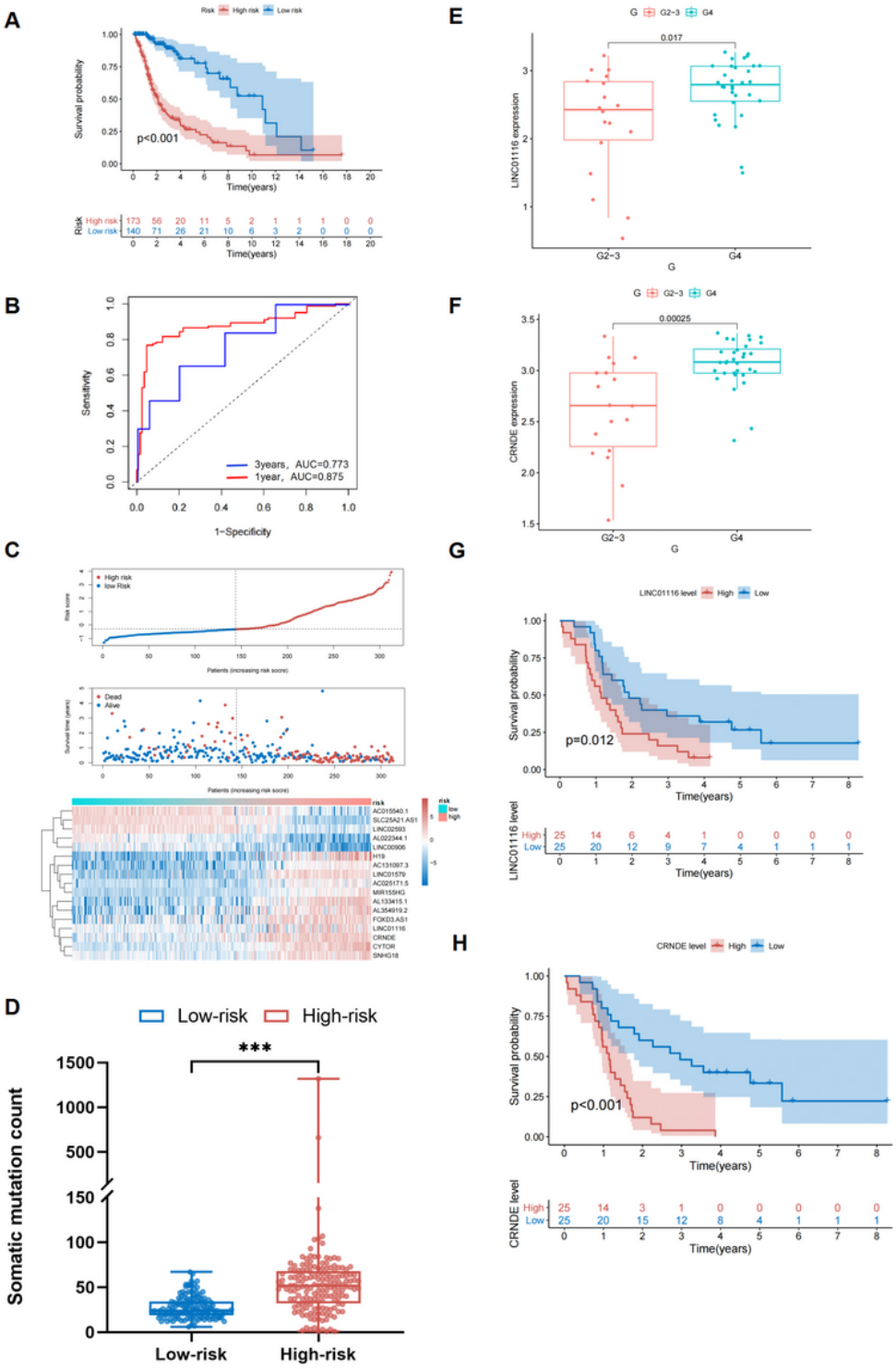

Figure 4

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 the 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).

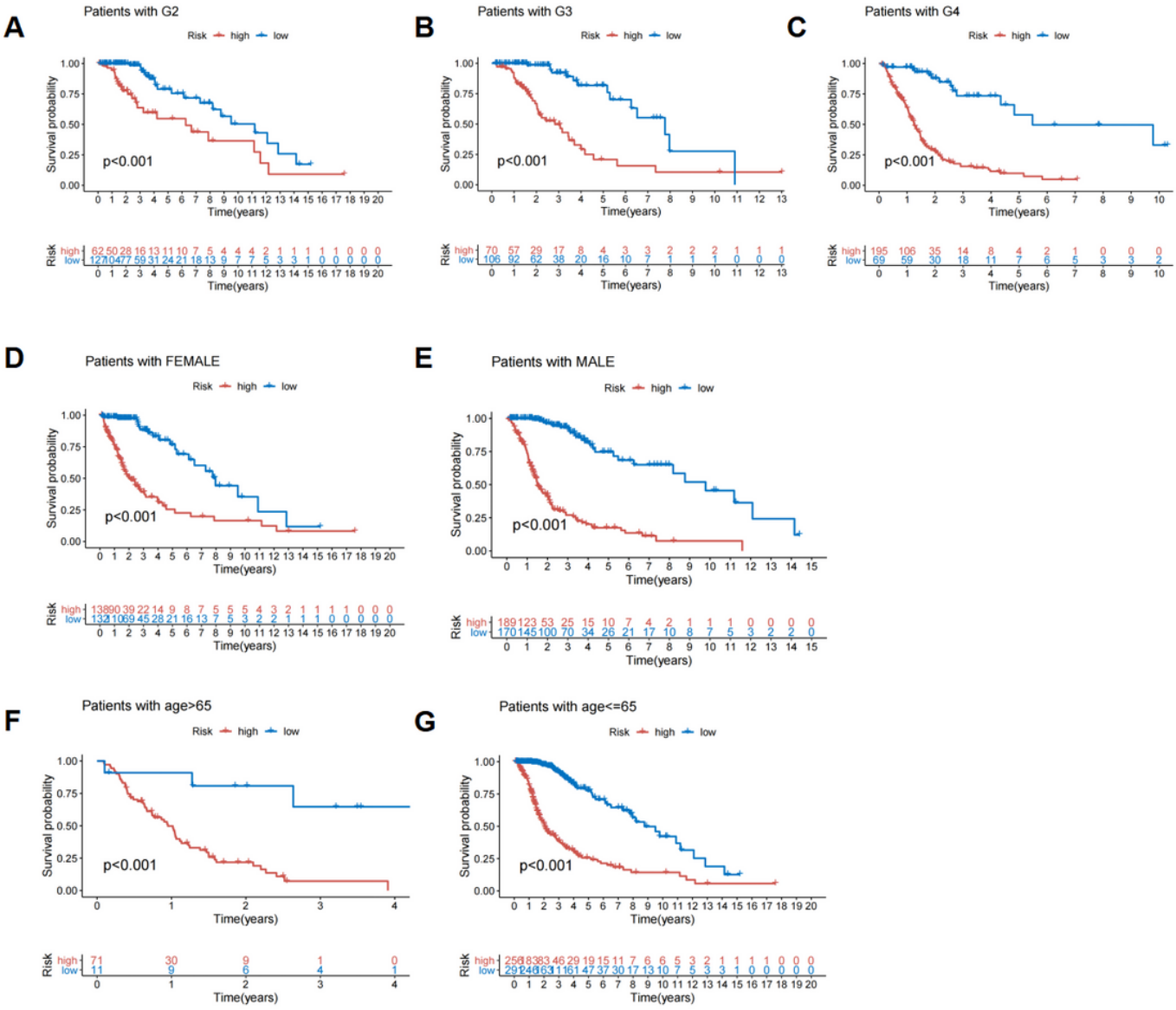

Figure 5

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).

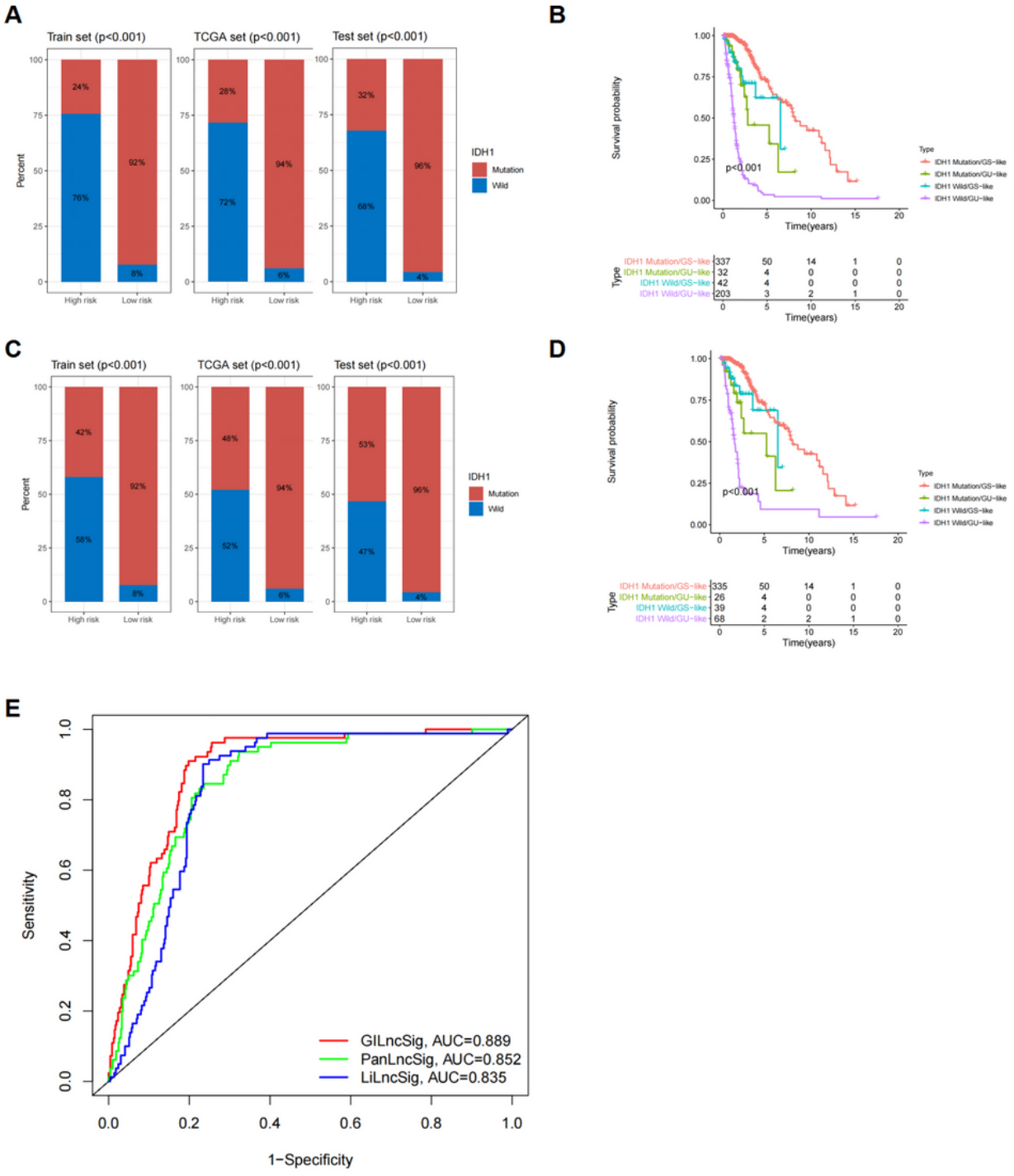

Figure 6

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.

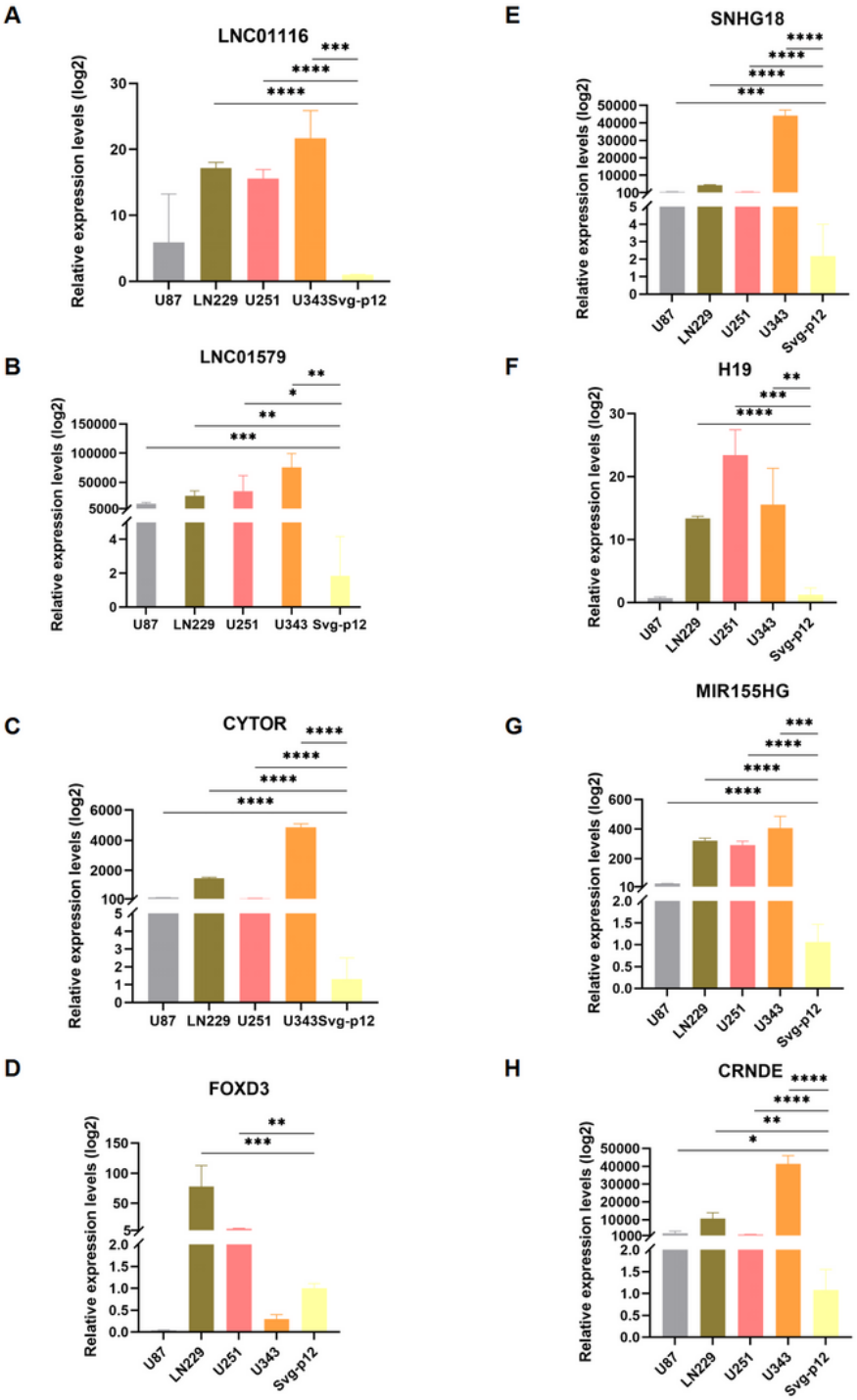

Figure 7

RT-PCR validation of cell lines. (A-H**)** The expression of 8 risk factors lncRNAs in cellular validation differed between tumor and normal cells (unpaired t-test, P<0.05).

Supplementary Files

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

- Supplementary.zip