A novel 8-genome instability-associated lncRNAs signature predicting prognosis and drug sensitivity in gastric cancer

Changhong Yi[1,#], Xiulan Zhang[2,#], Xia Chen[3,#], Birun Huang[4], Jing Song[5], Minghui Ma[6] ⓘ, Xiaolu Yuan[6] and Chaohao Zhang[6]

Abstract

Background: Genome instability lncRNA (GILnc) is prevalently related with gastric cancer (GC) pathophysiology. However, the study on the relationship GILnc and prognosis and drug sensitivity of GC remains scarce.

Method: We extracted expression data of 375 GC patients from TCGA cohort and 205 GC patients from GSE26942 cohort. Then, lncRNA was separated from expression data, and systematically characterized the 8 marker lncRNAs using the LASSO method. Next, we constructed a GILnc model (GILnc score) to quantify the GILnc index of each GC patient. Finally, we analyzed the relationship between GILnc score and clinical traits including survival outcomes, TP53, and drug sensitivity of GC.

Results: Based on a computational frame, 205 GILncs in GC has been identified. Then, a 8 GILncs was successfully established to predict overall survival in GC patients based on LASSO analysis, divided GC samples into high GILnc score and low GILnc score groups with significantly different outcome and was validated in multiple independent patient cohorts. Furthermore, GILnc model is better than the prediction performance of two recently published lncRNA signatures, and the high GILnc score group was more sensitive to mitomycin. Besides, the GILnc score has greater prognostic significance than TP53 mutation status alone and is capable of identifying intermediate subtype group existing with partial TP53 functionality in TP53 wild-type patients. Finally, GILnc signature as verified in GSE26942.

Conclusion: We applied bioinformatics approaches to suggest that a 8 GILnc signature could serve as prognostic biomarkers, and provide a novel direction to explore the pathogenesis of GC.

Keywords

Gastric cancer, TCGA, Genome instability, LncRNA, drug sensitivity

[1]Department of Interventional, Cancer Hospital of Shantou University Medical College, Shantou, China
[2]Department of Nuclear Medicine, The First People's Hospital of Jingzhou, The First Affiliated Hospital of Yangtze University, Jingzhou, Hubei, China
[3]Department of Oncology, Jingzhou Central Hospital, The Second Clinical Medical College, Yangtze University, China
[4]Department of Vascular Surgery, The First People's Hospital of Jingzhou, The First Affiliated Hospital of Yangtze University, Jingzhou, China
[5]Department of Nursing, Hubei College of Chinese Medicine, Jingzhou, People's Republic of China
[6]Department of Gastrointestinal Surgery, Maoming People's Hospital, Maoming, China

#Changhong Yi, Xiulan Zhang and Xia Chen contributed equally to this work

Corresponding author:
Birun Huang, The First People's Hospital of Jingzhou, The First Affiliated Hospital of Yangtze University, No. 8 Hangkong Road, Shashi, Jingzhou 434000, China.
Email: dr__huangbirun@163.com

Introduction

Gastric cancer (GC) is one of the leading contributors to the global cancer disease burden, which has brought heavy burden to the society and family.[1,2] Despite the many treatments available to treat GC, it still ranks fourth in tumor-related mortality.[3,4] The genesis and development of GC are reported to be due to the accumulation of gene mutations and epigenetic changes, which in turn lead to transcriptional or translational dysregulation.[5,6] Due to gene mutations that have occurred in the early stages of the malignant tumor, molecular mechanisms of gene mutations regulation have to be further explored, and the analysis of its relations with GC metastasis and drug resistance will help to find new early GC markers, and will help to improve and enrich the early diagnosis of GC, in order to achieve targeted drugs which provide theoretical basis for the development of GC, and will help to improve the prognosis of patients with GC.

Molecular biology studies have shown that there is genomic instability or genetic instability in GC.[7] Loss of heterozygosity (LOH) and microsatellite instability (MSI) caused by DNA mismatch repair (MMR) gene repair errors are considered to be the two main phenotypic features of genome instability.[8] Genomic instability can be a hallmark of both human genetic disease and cancer.[9] Tumor cells have biological characteristics such as unrestricted proliferation, vascular regeneration, and metastasis, and the instability of the genome is the most essential characteristic of tumor cells. The relative stability of the genome is the basic prerequisite for cells to be faithfully passaged. When there are genetic defects or exposure to adverse environmental factors such as biological, physical, and chemical tests, it will lead to genomic instability.[10] Studies have constructed 10-miRNA signatures related to DNA damage response, and have shown that 10-miRNA signatures are associated with poor prognosis in ovarian cancer.[11] More and more evidence indicates that lncRNA plays an important role in tumors, and abnormal lncRNA expression may affect tumor cell proliferation, tumor progression or metastasis.[12-14] Therefore, constructing lncRNAs related to genomic instability may be a prognostic factor for GC.

In this study, we extracted the transcription data of 375 GC patients from TCGA cohort and 205 GC patients from GSE26942 cohort. Then, lncRNA was separated from expression data, and systematically characterized 8 marker genome instability lncRNAs (GILnc) using the LASSO method. We constructed a GILnc score to quantify the GILnc index of each GC patient. We revealed that the distribution of GILnc score was validated in the GSE26942 cohort, suggesting that this scoring system based on GILnc plays an important role in predicting the prognosis of GC.

Materials and methods

Data source and preprocessing

This study was retrospective in nature. All data in this study was obtained from Public DataBase. We systematically searched GC-related array datasets from the public databases and selected the GC microarray data since 2010. The GC microarray datasets were recruited from Gene Expression Ominibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) with the following criteria: (1) only from Affymetrix platform; (2) gastric cancer; (3) the number of patients ≥50; (4) with more than 12,000 protein coding genes. Finally, GSE26942 (n = 205) microarrays dataset were retrieved.[15] One dataset was from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/repository): TCGA-STAD (n = 375). Then lncRNA was separated from expression data. Next, we divided all TCGA-STAD samples into a training set and a test set. The training set included 168 samples for the creation of a clinical outcome lncRNA risk model. The test set included 169 patients, used to validate the predictive ability of the prognostic risk model. Meanwhile, we calculated the tumor mutation burden (TMB) in the samples and estimate the average number of mutations in the tumor genome. Information on the data obtained is summarized in Table 1.

Selection of GILnc signature in GC

To identify the GILncSigs in GC, the top 25% somatic mutations per patient (n = 90) and the last 25% somatic mutations per patient (n = 97) of the patients were assigned to GU-like (genomic unstable) group and GS-like (genomic stable) group. Use the "Limma" software package in the R statistical software to extract and analyze the downloaded data, and screen out the differentially GILncs between GU-like group and GS-like group. We set the adjusted p-value < 0.05 and $\log_2 FC > 1$ as a significance threshold.[16]

Construction of a risk model with prognostic value in GC

Next, univariate Cox regression analysis was used for differentially expressed GILncs, and the differentially expressed GILncs related to the patient's overall survival (OS) were screened out based on $p<0.05$. Then, LASSO was used to identify the most powerful prognostic GILncs. A risk score for each patient was calculated according to the following formula

$$GlncSig\,Riskcore = \sum_{i=1}^{n} coef(lncRNAi)^* expr(lncRNAi)$$

Table 1. Clinical characteristics of GC patient datasets in this study.

Characteristic		TCGA dataset (N = 337)	Discovery dataset (N = 168)	Validation dataset (N = 169)	p-value
Age (years), n (%)	≤65	153 (45.4%)	74 (44.1%)	79 (46.8%)	0.735
	>65	181 (53.7%)	92 (54.8%)	89 (52.7%)	
	Unknown	3 (0.9%)	2 (1.2%)	1 (0.6%)	
Gender, n (%)	Female	119 (35.3%)	66 (39.3%)	53 (31.4%)	0.159
	Male	218 (64.7%)	102 (60.7%)	116 (68.6%)	
Grade, n (%)	G1-2	129 (38.3%)	65 (38.7%)	64 (37.9%)	0.968
	G3-4	199 (59.1%)	102 (60.7%)	97 (57.4%)	
	Unknown	9 (2.67%)	1 (0.6%)	8 (4.7%)	
Stage, n (%)	Stage I/II	152 (45.1%)	73 (43.5%)	79 (46.8%)	0.614
	Stage III/IV	171 (50.7%)	88 (52.4%)	83 (49.1%)	
	Unknown	14 (4.2%)	7 (4.2%)	7 (4.1%)	
T, n (%)	T1-2	89 (26.4%)	45 (26.8%)	44 (26.0%)	0.974
	T3-4	244 (72.4%)	121 (72.0%)	123 (72.8%)	
	Unknown	4 (1.2%)	2 (1.2%)	2 (1.2%)	
M, n (%)	M0	303 (89.9%)	157 (93.5%)	146 (86.4%)	0.238
	M1	22 (6.5%)	8 (4.8%)	14 (8.3%)	
	Unknown	12 (3.6%)	3 (1.8%)	9 (5.3%)	
N, n (%)	N0	99 (29.4%)	46 (27.4%)	53 (31.4%)	0.385
	N1-3	227 (67.4%)	119 (70.8%)	108 (63.9%)	
	Unknown	11 (3.3%)	3 (1.8%)	8 (4.7%)	
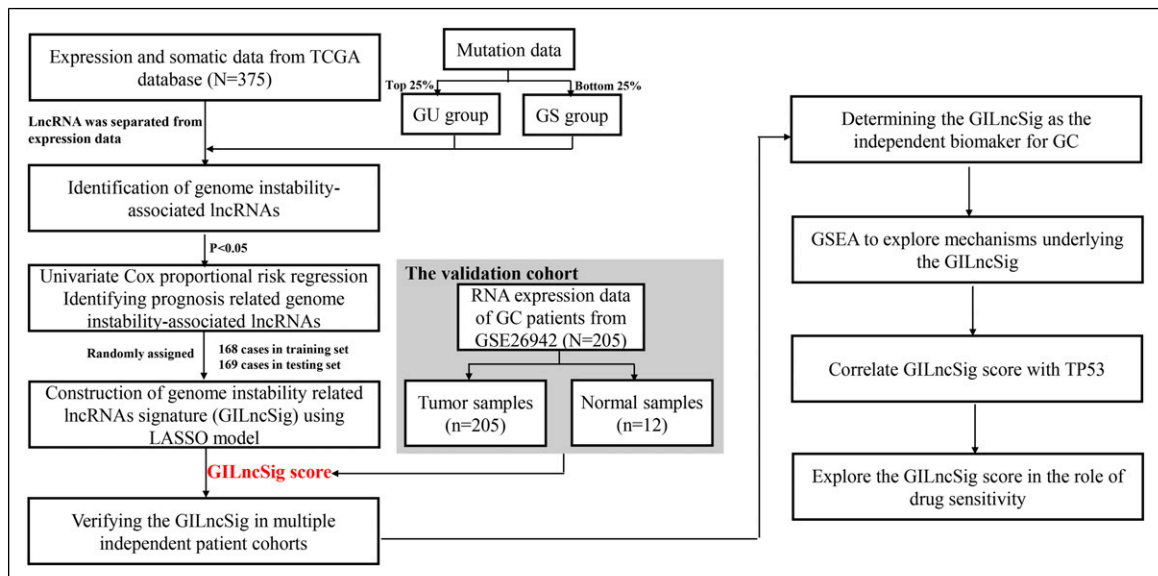
Figure 1. Flow chart of the steps in the performed analyses.

where i means the GILnc feature genes.[17,18] Next, we divided GC samples into high-risk and low-risk groups base on the median of GILnc score.

Gene set enrichment analysis (GSEA)

We performed GSEA to identify differences in the enrichment of pathways and biological processes between high-risk and low-risk groups. GSVA was conducted using the "GSEA" package in R. We downloaded the gene sets of "c2. cp.kegg.v7.3. symbols" and "h.all.v7.3. symbols" from the MSigDB database for GSEA (http://www.gsea-msigdb.org/gsea/downloads.jsp). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the "clusterProfiler" package. A significance level of 0.05 (FDR) was considered to indicate statistical significance.[19]

Immune cell infiltration

We performed the ssGSEA, CIBERSORT, and xCell algorithm methods to calculate the composition of immune cells between the clusters

Statistical analysis

The normality of the variables was evaluated using the Shapiro–Wilk normality test. Continuous variables between two groups were compared using the unpaired Student t-test and Mann–Whitney U test for parametric data and non-parametric data, respectively. For comparison between more than two groups, we used parametric one-way ANOVA or non-parametric Kruskal–Wallis test. Kaplan–Meier analysis was used to generate survival curves using the "survival" and "survminer" packages, and the cut-off values were determined through the "surv_cutpoint" function in the

packages. The receiver operating characteristic (ROC) curve was generated using the "pROC" package. To calculate the hazard ratios and identify the independent prognostic factors, univariate and multivariate Cox regression analyses were performed using the "survival" package. All statistical analyses were two-sided and considered $p < 0.05$ as the threshold for statistical significance. The statistical results were all analyzed by R (version3.6.2).

Result

Identification of GILncs signature in GC patients

The workflow of our study was presented in Figure 1. As showed in Supplemental Table 1, GILncs were significantly different expression between GU-like group and GS-like (Figure 2(a)). Using 205 differentially expressed GILncs, all 375 patients were arranged into two clusters,
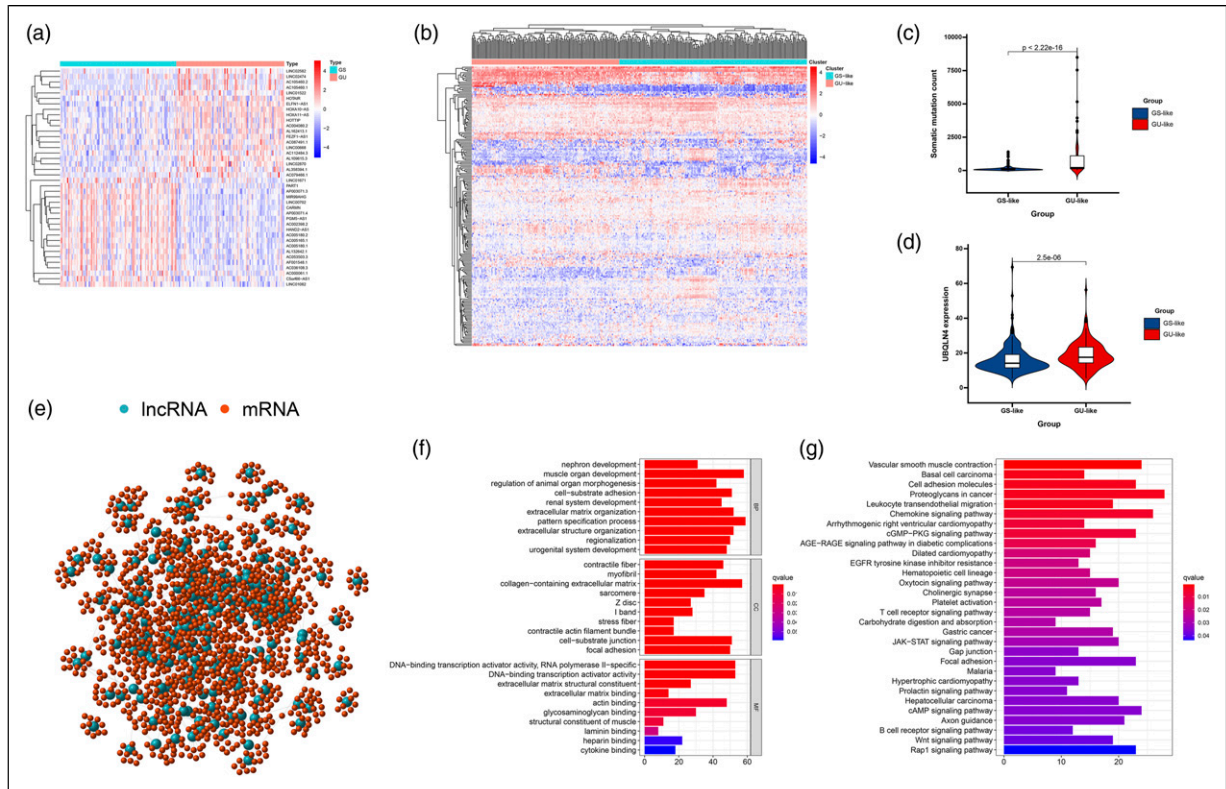

Figure 2. Identification and functional annotations of genomic instability-related lncRNAs in patients with gastric cancer. (A) Heatmap of the top 20 genome instability-associated lncRNAs expressing the most upregulation and downregulation. (B) Unsupervised clustering of 375 GC patients based on the expression pattern of 205 candidate genomic instability-related lncRNAs. The left blue cluster is GS-like group, and the right red cluster is GU-like group. (C) Boxplots of somatic mutations in the GU-like group and GS-like group. Somatic cumulative mutations in the GU-like group are significantly higher than those in the GS-like group. (D) Boxplots of UBQLN4 expression level in the GU-like group and GS-like group. The expression level of UBQLN4 in the GU-like group is significantly higher than that in the GS-like group. Horizontal lines: median values. Statistical analysis was performed using the Mann–Whitney U test. (E) Coexpression network of genomic instability-related lncRNAs and mRNAs based on the Pearson correlation coefficient. The red circles represent lncRNAs, and the blue circles represent mRNAs. (F) Functional enrichment analysis of GO for mRNAs co-expressed lncRNAs. (G) Functional enrichment analysis of KEGG for mRNAs co-expressed lncRNAs.

the GS-like cluster and the GU-like cluster (Figure 2(b)). Next, we further analyzed the difference between the cumulative somatic mutation value between the GU group and the GS group, and the results showed that the cumulative somatic mutation value of the GU group was significantly higher than that of the GS group (Figure 2(c)). Researchers have discovered UBQLN4 is a newly identified driver of genomic instability, and harmful UBQLN4 mutations in families with autosomal recessive syndromes. Loss of UBQLN4 can lead to increasing sensitivity to genotypic stress and delayed DNA double-strand break repair.[20] Therefore, we compared the expression level of UBQLN4 in the GU group and the GS group, and the results showed that the expression level of UBQLN4 in the GU group was significantly higher than that in the GS group (Figure 2(d)).

To better understand the functions of these GILncs, we constructed a lncRNA-mRNA coexpression network, where the nodes are lncRNA and mRNA, if they are related to each other, lncRNA and mRNA will be linked together (Figure 1(e)). In addition, functional enrichment analysis of go and kegg for lncRNA-correlated mRNA, go and kegg showed that lncRNA-correlated mRNA are mainly cancer-specific pathways (Figure 2(f) and (g)).

Development of a GILncs signature for prognosis of patients with GC in the training set

Next, 375 GC patients were divided into the training set and the testing set according to best batches. Table 1 shows the clinical characteristic of training set, the testing set and TCGA dataset. In order to predict the clinical outcomes of
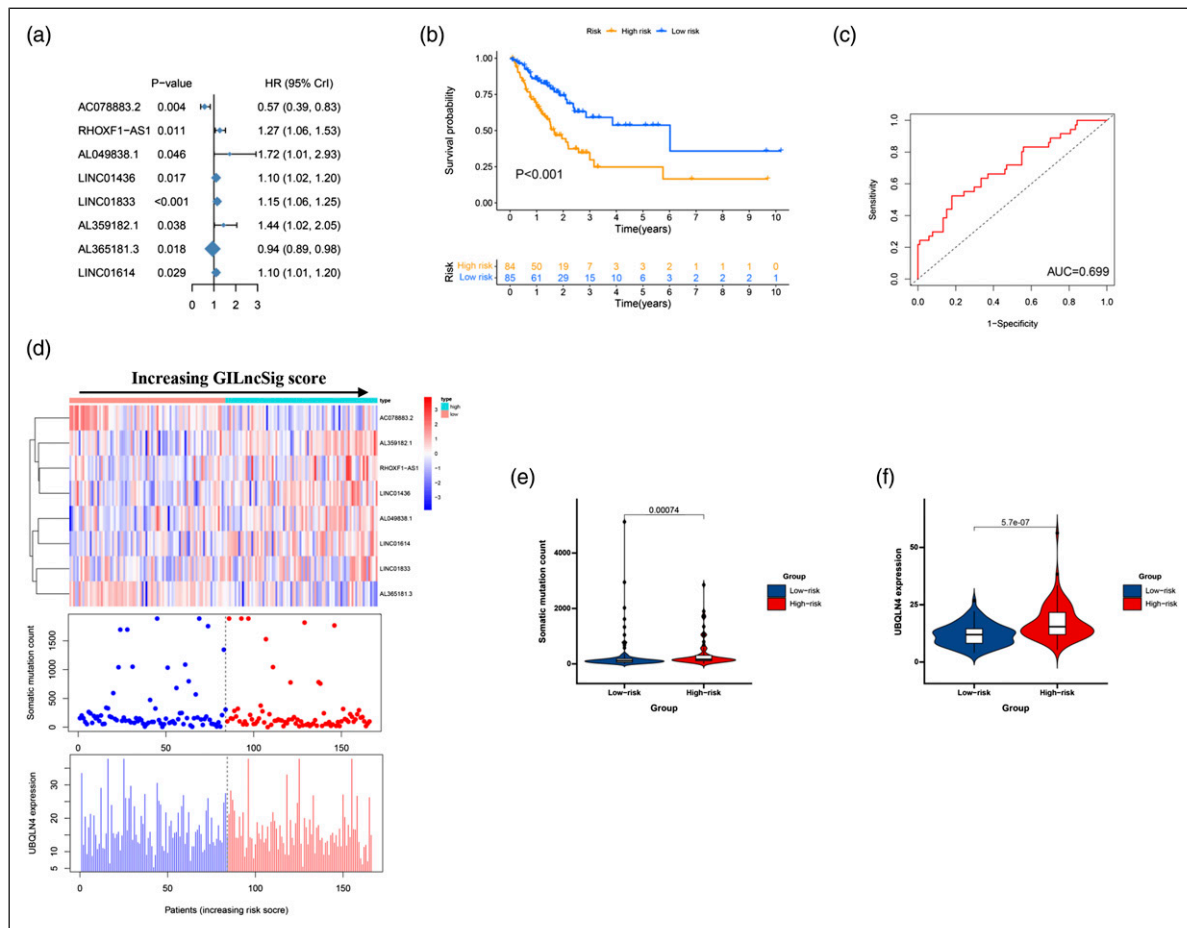

Figure 3. Identification of the genomic instability-derived lncRNA signature (GILncSig) for outcome prediction in the training set. (A) Forest plots showing the results of the multivariate Cox regression between GILncSig expression and OS. (B Kaplan–Meier estimates of overall survival of patients with low or high risk predicted by the GILncSig in the training set. Statistical analysis was performed using the log-rank test and univariate Cox analysis. (C) Time-dependent ROC curves analysis of the GILncSig at 1-year. (D) LncRNA expression patterns and the distribution of somatic mutation and UBQLN4 expression with increasing GILncSig score. The distribution of somatic cumulative mutations (E) and UBQLN4 expression in the (F) in the high-risk and low-risk groups for GC patients. The red represents the high-risk group, and the blue represents the low-risk group. Horizontal lines: median values. Statistical analysis was performed using the Mann–Whitney U test.

Table 2. Multivariate Cox regression analysis of genome instability-related lncRNAs associated with overall survival in GC

| Ensembl ID	Gene symbol	Genomic location	Coefficient	HR	95% CI	*p*-value
ENSG00000232788	AC078883.2	Chromosome 2: 172,464,262–172,466,022	−0.57	0.57	0.39–0.83	0.004
ENSG00000258545	RHOXF1-AS1	Chromosome X: 120,036,236–120,146,854	0.24	1.27	1.06–1.53	0.011
ENSG00000259039.3	AL049838.1	Chromosome 14: 57,578,409–57,600,404	0.54	1.72	1.01–2.93	0.046
ENSG00000231106	LINC01436	Chromosome 21: 36,005,338–36,007,838	0.10	1.10	1.02–1.20	0.017
ENSG00000259439	LINC01833	Chromosome 2: 44,921,077–44,939,199	0.14	1.15	1.06–1.25	<0.001
ENSG00000083622	AL359182.1	Chromosome 7: 117,604,791–117,647,415	0.37	1.45	1.02–2.05	0.038
ENSG00000272405	AL365181.3	Chromosome 1: 156,641,666–156,644,887	−0.06	0.94	0.89–0.98	0.018
ENSG00000230838	LINC01614	Chromosome 2: 215,718,043–215,719,424	0.09	1.10	1.01–1.20	0.029
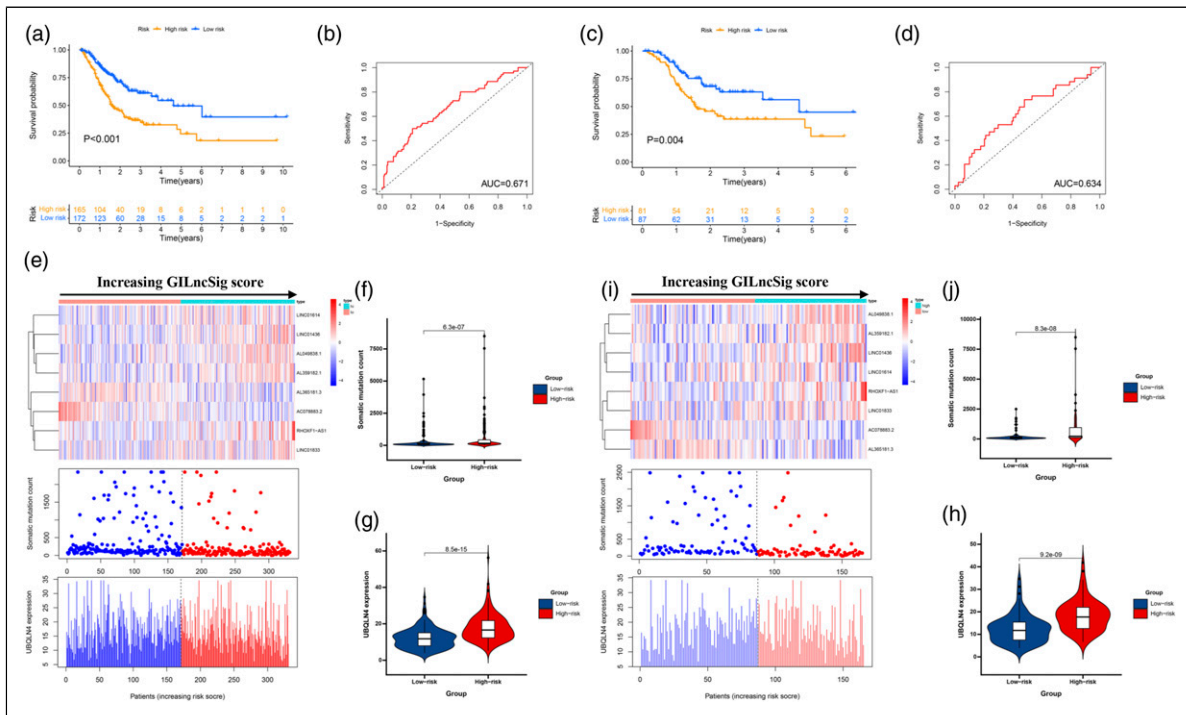
Figure 4. Performance evaluation of the GILncSig in the TCGA set and testing set. Kaplan–Meier estimates of overall survival of patients with low or high risk predicted by the GILncSig in the TCGA set (A) and testing set (C). Statistical analysis was performed using the log-rank test and univariate Cox analysis. time-dependent ROC curves analysis of the GILncSig at 1-year in the training set (B) and testing set (D). LncRNA expression patterns and the distribution of somatic mutation count distribution and UBQLN4 expression for patients in high- and low-risk groups in the training set (E) and testing set (H). The distribution of somatic mutation in patients of high- and low-risk groups in the training set (F) and testing set (I). The distribution of UBQLN4 expression in patients of high- and low-risk groups in the training set (G) and testing set (J). Horizontal lines: median values. Statistical analysis was performed using the Mann–Whitney U test.

GC with GILncs signature, we applied LASSO Cox regression algorithm to the 205 GILncs in the training set. Eight GILncSigs were selected to build the risk signature based on the minimum criteria (Supplemental figure 1(a) and (b)). Next, to inspect whether eight GILncSigs were related to prognosis in GC, multivariate Cox regression was employed to analyze the hazard ratio (HR) of 8 GILncSigs in GC. Forest plot showed that high expression levels of 6 genes including RHOXF1-AS1, AL049838.1, LINC01436, LINC01833, AL359182.1, and LINC01614

were significantly related to poor OS of GC patients. While, high expression levels of AC078883.2 and AL365181.3 were closely related to relatively better OS in GC patients (Figure 3(a), Table 2). Furthermore, Kaplan–Meier log-rank analysis revealed that high GILnc score group was significantly correlated with poor overall survival (OS) (Figure 3(b)). To test the efficiency of GILncSig, ROC curve was constructed. The risk score's AUC was 0.699, indicating that its efficiency to predict prognosis was accurate (Figure 3(c)). Simultaneously, as increasing GILnc

score, the chang of expression in the training set (Figure 3(d)). We further analyzed the difference between the cumulative somatic mutation value between the two groups. The result showed the cumulative somatic mutation value in the high GILnc score group was higher than that in the other group (Figure 3(e)). We also compared the expression level of UBQLN4 in two groups. The result showed the UBQLN4 in the high GILnc score group was higher than that in the other group (Figure 3(f)).

Validation of GILnc score for outcome prediction in the testing set and TCGA set

To examine the robustness of the GILnc score in the testing set and TCGA set. The results showed that patients in the high GILnc score group had a poor survival as compared with low GILnc score group in the testing set and TCGA set (Figure 4(a) and (c)). The risk score's AUC was 0.671 in TCGA set and 0.634 in testing set, indicating that its efficiency to predict prognosis was accurate (Figure 4(d) and (d)).

Simultaneously, as increasing GILnc score, the chang of expression in the testing set and TCGA set (Figure 4(e) and (i)). We further analyzed the difference between the cumulative somatic mutation value between the two groups. The result showed the cumulative somatic mutation value in the high GILnc score was higher than that in the testing set and TCGA set (Figure 4(f) and (j)). We also compared the expression level of UBQLN4 in two groups in the testing set and TCGA set, and the expression level of UBQLN4 in the high GILnc score group was significantly higher than that in the low GILnc score group (Figure 4(g) and (k)). Functional enrichment analysis revealed that the high GILnc score group was enriched in cancer-related pathways, and low GILnc score group was enriched in metabolism-related pathways (Figure 5(a) and (b)).

Independent prognostic analysis of GILncs signature

Next, to investigate whether GILnc score could serve as an independent prognostic factor for GC, we performed
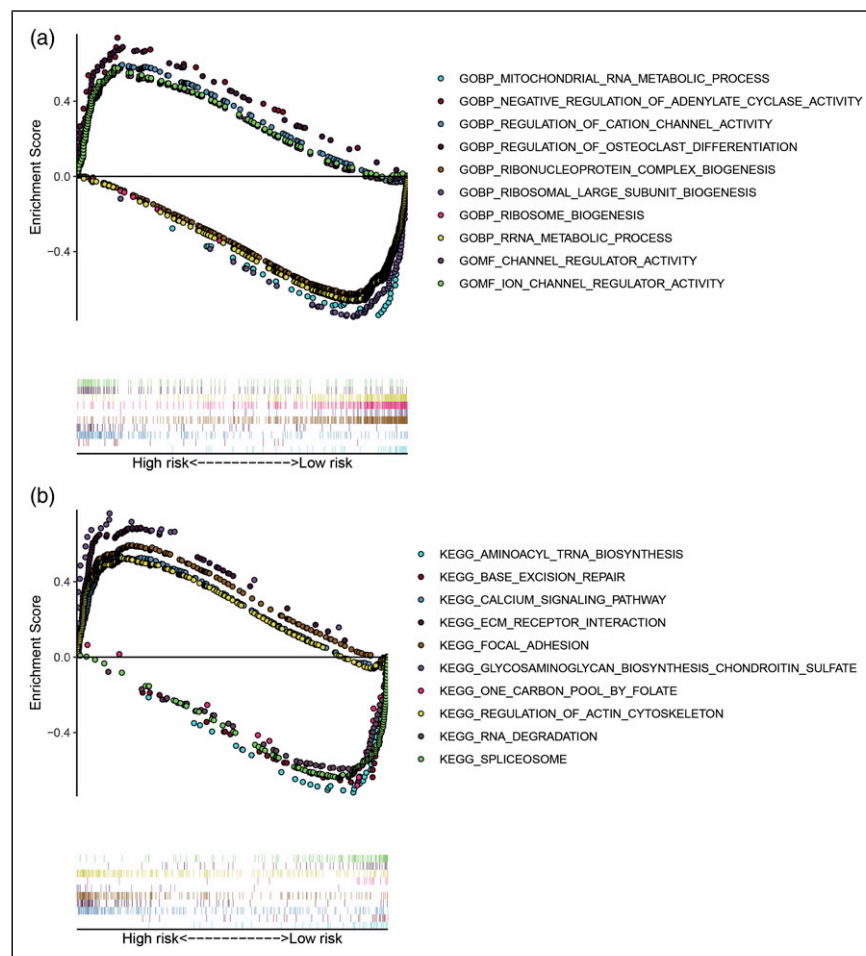

Figure 5. GSEA on the TCGA cohort to explore mechanisms underlying the 8-GILncSig. (A) GSEA GO identifies high and low risk related signaling pathway in GC. (B) GSEA KEGG identifies high and low risk related signaling pathway in GC.

Table 3. Univariate and multivariate Cox regression to identify independent prognosis predictor in both the TCGA cohort and the GSE26942 cohort.

Characteristics	TCGA cohort				GSE26942 cohort			
	Univariate mode		Multivariate model		Univariate mode		Multivariate model	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Age (≥65 vs <65)	1.06 (1.02–1.10)	<0.001	1.03 (1.00–1.05)	0.003	1.12 (0.75–1.75)	0.315		
Gender (male vs female)	1.39 (0.96–1.92)	0.082			1.07 (0.52–1.97)	0.819		
AFP (≥200 vs <200)	0.92 (0.62–1.36)	0.427						
TMB (TMB-H vs TMB-L)	1.34 (1.18–1.43)	<0.001	1.29 (1.14–1.43)	<0.001				
Tumor grade (G3/4 vs G1/2)	0.85 (0.49–1.51)	0.351						
Tumor stage (III/IV vs I/II)	1.42 (1.16–1.84)	<0.001	1.30 (0.83–1.93)	0.205	1.45 (0.62–3.60)	0.385		
Vascular invasion (yes vs no)	1.91 (0.92–3.96)	0.175						
Risk (high vs low)	1.15 (1.10–1.19)	<0.001	1.09 (1.03–1.16)	<0.001	1.71 (1.17–2.57)	0.010	1.53 (1.07–2.69)	0.037
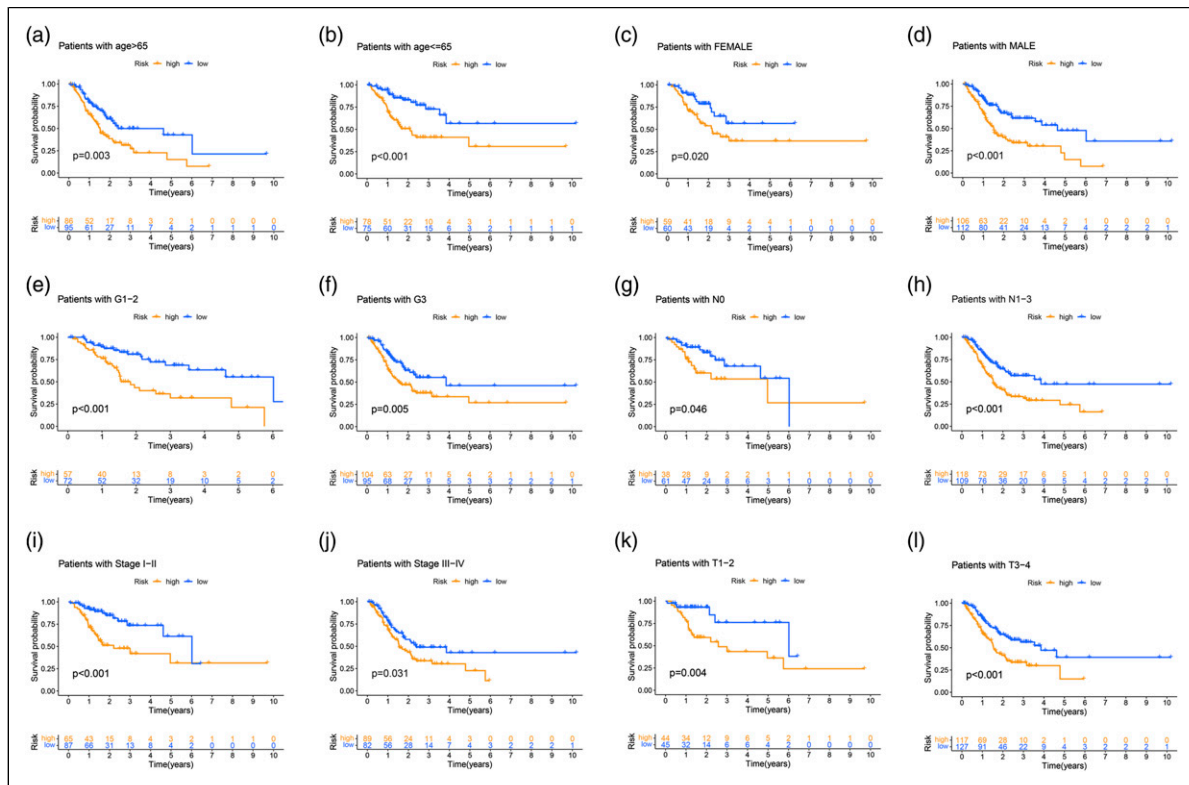
Figure 6. Kaplan–Meier curves were performed for patients stratified by clinicopathological features in the TCGA set. Impact of prognostic risk on overall survival for patients younger than 65 years old (A) and older than 65 years old (B); for female (C) and male patients (D); for patients in G1-2 (E) and G3 (F); for patients in N0 (G) and N1-3 (H); for patients in stage I-II (I) and stage III-IV (J); and for patients in stage T1-2 (K) and stage T3-4 (L).

univariate and multivariate Cox regression analyses. The results indicated that GILnc score is a robust independent prognostic factor in TCGA cohort (1.09 [1.03–1.16], $p < 0.001$) and GSE26942 cohort (1.53 [1.07–2.69], $p = 0.037$) (Table 3). Subsequently, Kaplan–Meier log-rank analysis revealed that high GILnc score group was significantly correlated with poor OS in the pathological stage (Figure 6(a)-6L). In addition, the AUCs for 1-year, 3-year, and 5-year OS were 0.671, 0.670, and 0.668 (Figure 7(a)). Furthermore, the results showed that the AUCs for risk, age, gender, and stage OS were 0.671, 0.563, 0.519, and 0.601 (Figure 7(b)). Based on the above results, GILnc score has independent research value in GC.

Association of GILnc score with TP53

TP53 gene is closely related to the occurrence and development of GC.[21] As shown in Figure 7(a), the proportion of patients with TP53 mutations in the high GILnc score group was significantly higher than that in the low GILnc score group ($p < 0.05$). Then, we divide all patients into TP53 mutation/GS, TP53 mutation/GU, TP53 wild/GS,

and TP53 wild/GU group. The results showed that patients in TP53 mutation/GU had a poor survival as compared with TP53 mutation/GS in TCGA set (Figure 7(b)). Furthermore, we compared the resulting GIlnc to the latest published signatures related to lncRNAs; the first signature is the 8-lncRNA signature (SunLncSig),[22] and the second signature is the 3-lncRNA signature (WangLncSig).[23] The results showed that the AUCs for SunLncSig, WangLncSig, and GILncSig OS were 0.586, 0.589, and 0.671 (Figure 7(c)). To gain further insight into the effects of the GILnc score on drug sensitivity, we evaluated the correlation between GILnc score and drug sensitivity, we found that the high GILnc score group showed significantly increased sensitivity to mitomycin drug (Figure 7(d)). In addition, the positive EBV status was significantly correlated with lower GILnc score (Figure 7(e)).

The GILncs signature in the role of PD-1/L1 immunotherapy

We first analyzed the immune infiltration between high-risk and low-risk group, and significantly higher immune infiltration in the high-risk group was observed relative to
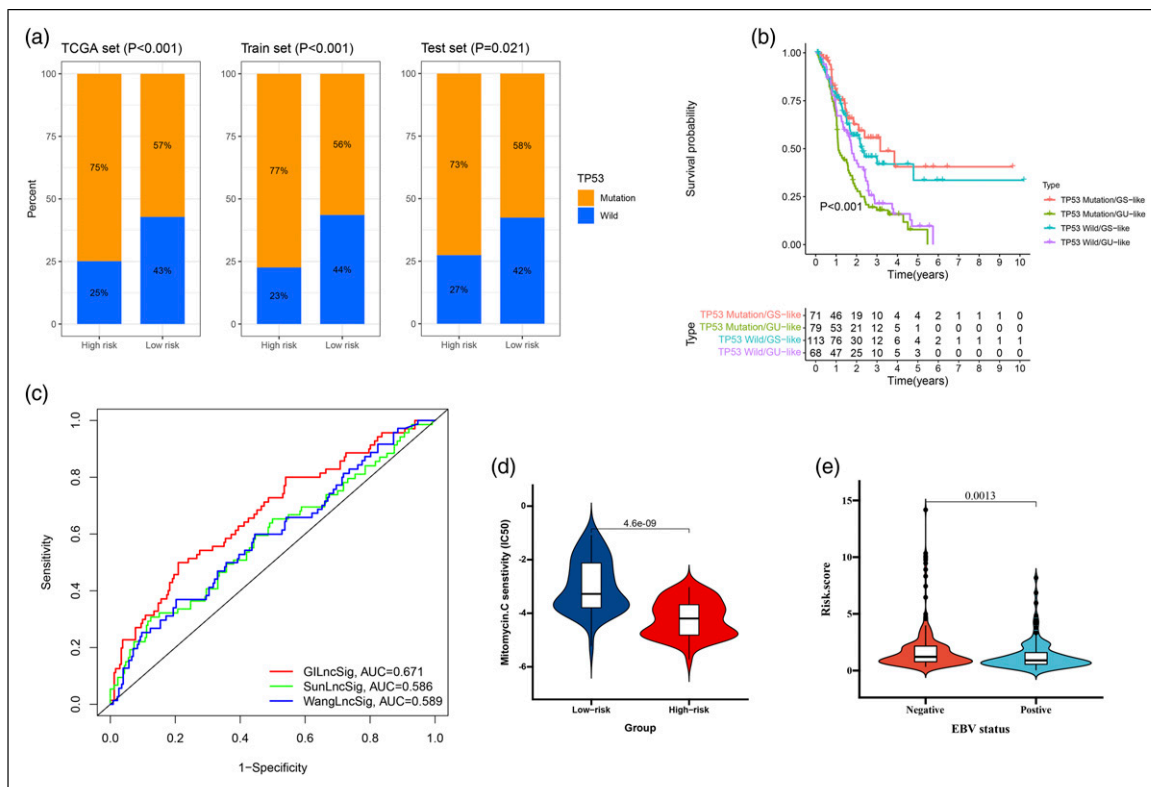
Figure 7. Relationship between the GILncSig and TP53 somatic mutation. (A) The proportion of TP53 mutation in high-risk and low-risk groups in the training set, testing set and the TCGA set. (B) Kaplan–Meier curve analysis of overall survival is shown for patients classified according to TP53 mutation status and the GILncSig. Statistical analysis was performed using the log-rank test. (C) The ROC analysis at 1-year of overall survival for the GILncSig, LilncSig and BailncSig. (D) The drug sensitivity in high-risk and low-risk groups. (E) The EBV status in high-risk and low-risk groups.

low-risk group (Figure 8(a)). Next, we download the immunotherapy data of TCGA-STAD patients, and the significant therapeutic advantages and clinical responses to PD-1/L1 immunotherapy in patients with high-risk group compared to those in low-risk group were confirmed (Figures 8(b) to (e)).

Construction and assessment of a predictive nomogram

A nomogram incorporating tumor stage and risk group was built to predict 1-year, 3-year, and 5-year OS (Figure 9(a)). Calibration plots showed that the nomogram was better at predicting short-term survival (1- and 3-year) rather than long-term survival (5-year), as indicated by agreement between the predicted survival and actual survival (Figure 9(b)). According to DCA curves, the nomogram also offered the highest net benefit among the three factors examined (Figure 9(c)).

Validation the expression and prognosis of GILncs signature in GSE26942

To further validate GILnc score in GC, GSE26942 cohort was used to measure the GILnc score, and the result showed that compared with normal group, the RHOXF1-AS1, AL049838.1, LINC01436, LINC01833, AL359182.1, and LINC01614 level were significantly higher in GC group, however, the AC078883.2 and AL365181.3 level were significantly lower in GC group (Figures 10(a) to (h)), and Kaplan–Meier analysis revealed that the prognosis of patients with high GILnc score group was significantly poor than that of patients with low GILnc score (Figure 10(i)).

Discussion

GC is characterized by genomic instability, researches show that genomic instability play significant roles in the pathological process of chronic atrophic gastritis, gastric
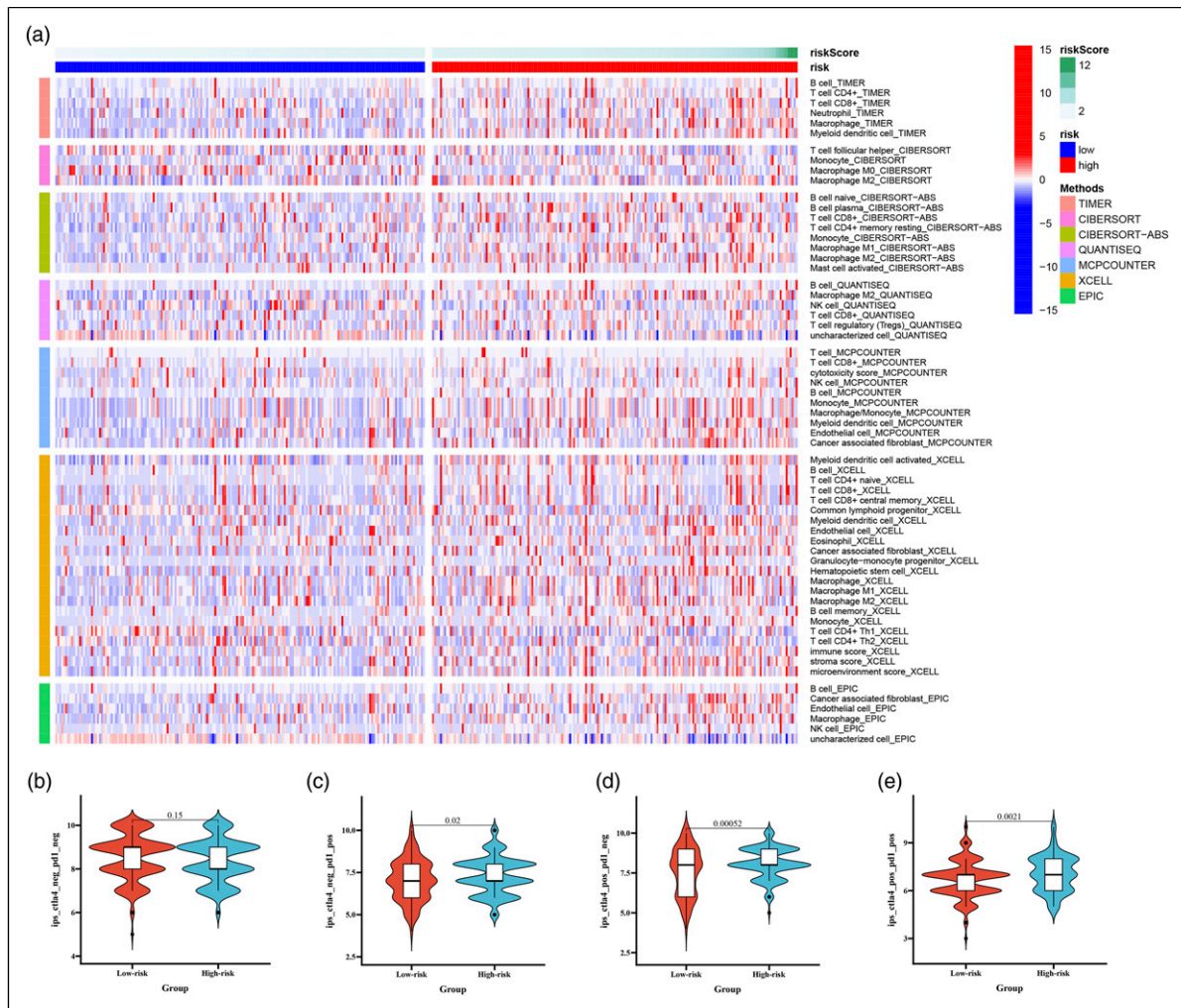
Figure 8. The GILncs signature in the role of PD-1/L1 immunotherapy. (A) Landscape of immune infiltration in high-risk and low-risk groups. The correlation of GILncs signature with clinical response to anti-PD-1 immunotherapy. (B) CTLA4_negative+ PD-1_negative, (C) CTLA4_negative + PD-1_positive, (D) CTLA4_positive + PD-1_negative, and (E) CTLA4_positive + PD-1_positive.

ulcer, atypical hyperplasia, intestinal metaplasia, and gastric carcinoma.[24] The occurrence of malignant tumor is a process of multi-gene participation and gradual evolution.[25] The progression from a normal cell to a malignant cell is actually a long process that involves various genetic mutations that lead to a precancerous lesion and then to a malignant tumor. This series of genomic evolution often requires DNA damage or replication abnormalities in chromosomal instability and even the emergence of a "hyperploidy" phenotype. As a result, most tumors often present a complex genetic map at the time of diagnosis, which is very different from that of normal controls, suggesting a high degree of genomic instability in the body at the time of tumor development.[26,27] Genomic instability is an important molecular feature of malignancy.[28] The relative stability of the genome is the basic prerequisite for faithful cell passage.[29] Detection of genomic instability is

now thought to be an early warning of tumorigenesis.[30] Current studies have also confirmed that genes that cause genomic instability are important clues to the causes of tumors.[31] Moreover, with the continuous improvement of modern molecular biology methods, more and more evidence shows that cancer patients can be timely treated by detecting their genomic instability.[32] Clinically, there have been some targeted drugs targeting genomic instability-related genes, which have brought good news to tumor patients.[33] It has been established that downregulation of these genomic instability-related genes has been clinically found to significantly enhance the sensitivity of cancer patients to platinum-based chemotherapy.[34,35] To sum up, it is of great significance to elucidate the relevant mechanisms of genomic instability in tumor cells and to conduct relevant assays to reduce the incidence of tumor, delay the progression and improve the disease condition.
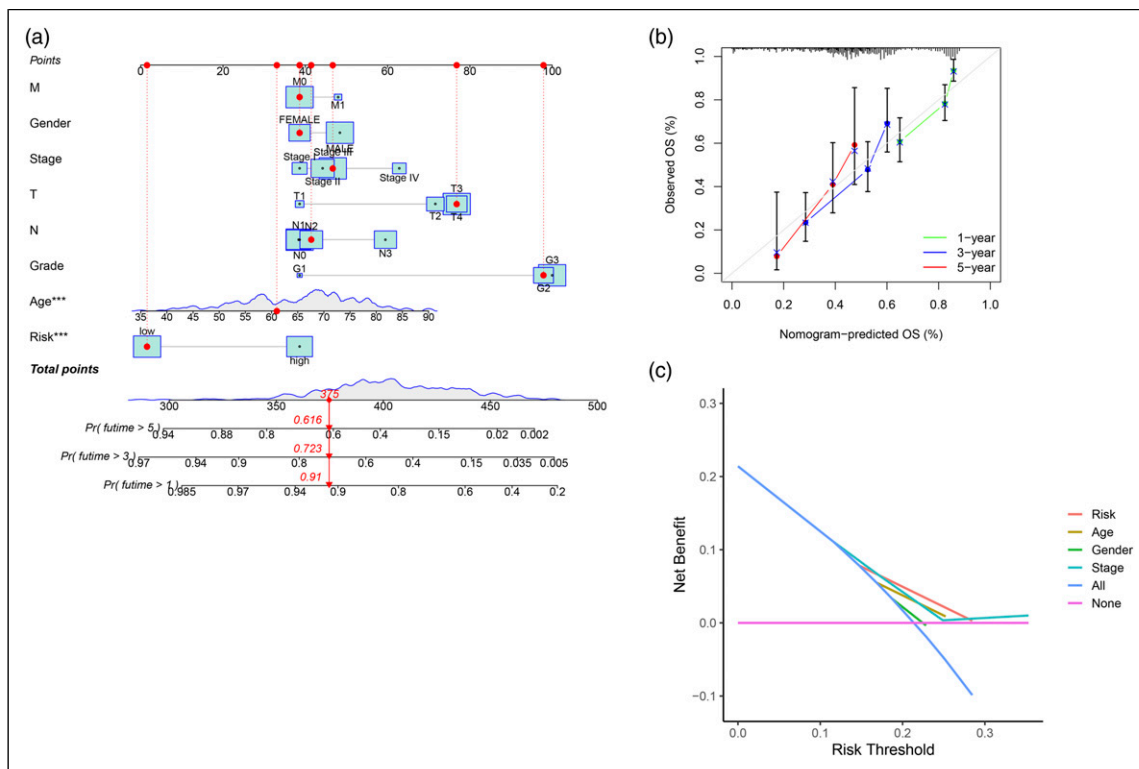
Figure 9. Construction and assessment of a predictive nomogram. (A) A nomogram predicting survival probability at 1-, 3-, and 5-year after surgery for GC patients; (B) Calibration curves for the nomogram; (C) DCA curves showing the comparison between the nomogram and tumor stage or risk group alone for predicting 1-, 3-, and 5-year overall survival in GC.

LncRNA is widely defined as a class of RNA molecules whose transcriptional length is greater than 200 nucleotides and lacks an open reading frame.[36] LncRNAs have a potential role in regulating the function of tumor cells.[37] LncRNA regulate gene expression at different levels including chromatin assembly, transcriptional, and posttranscriptional.[38,39] Recent studies have shown that NORAD and GUARDIN are essential for genomic stability.[40,41] Currently, a variety of tumor genomic instability detection technologies have sprung up rapidly, and the understanding of the role of genomic instability in the development of tumors is gradually deepening. However, the genom-wide identification of genomic instability-related lncRNAs and the systematic exploration of their clinical significance in cancer are still in their infancy. Therefore, it is of great significance to identify lncRNAs associated with genomic instability.

First, we downloaded the GC expression profile data and mutation data. We defined the 25% with the highest frequency as the high mutation group, and the 25% with the lowest frequency as the low mutation group. A total of 205 different lncRNAs were obtained compared with the low mutation group. These lncRNAs were characterized as genomic instability-related lncRNAs. Functional enrichment analysis of mRNAs co-expressed with 205 lncRNAs

indicated that these lncRNAs may play an important role in the pathogenesis of GC, which is consistent with other studies.[42] Abnormal repair of DNA damage is directly related to genomic stability. If the mechanism of repairing DNA damage is defective, it will directly lead to the persistence of DNA damage and the harmful changes of cells, until the tumor is triggered. DNA damage is mainly exogenous, such as chemical exposure, UV irradiation, biological hazards, and endogenous, such as in vivo spontaneous DNA damage events, cell cycle process and DNA replication process block. These damages, if not repaired in time, can induce genomic oxidation, alkylation, and even DNA crosslinking, dimer formation, and even DNA breakage.[43,44] Therefore, whether DNA damage can be repaired in time and correctly directly affects the maintenance of genome stability.

In the past, the pathological classification of gastric cancer commonly included Lauren classification, Borrmann classification, and WHO classification; however, these classifications have limited significance for the clinical treatment of gastric cancer.[45,46] Now, with the progress of genetic technology, the classification of gastric cancer at the molecular level is more meaningful for the selection of targeted drugs for gastric cancer patients. GILncSig divided patients into low GILnc score group and
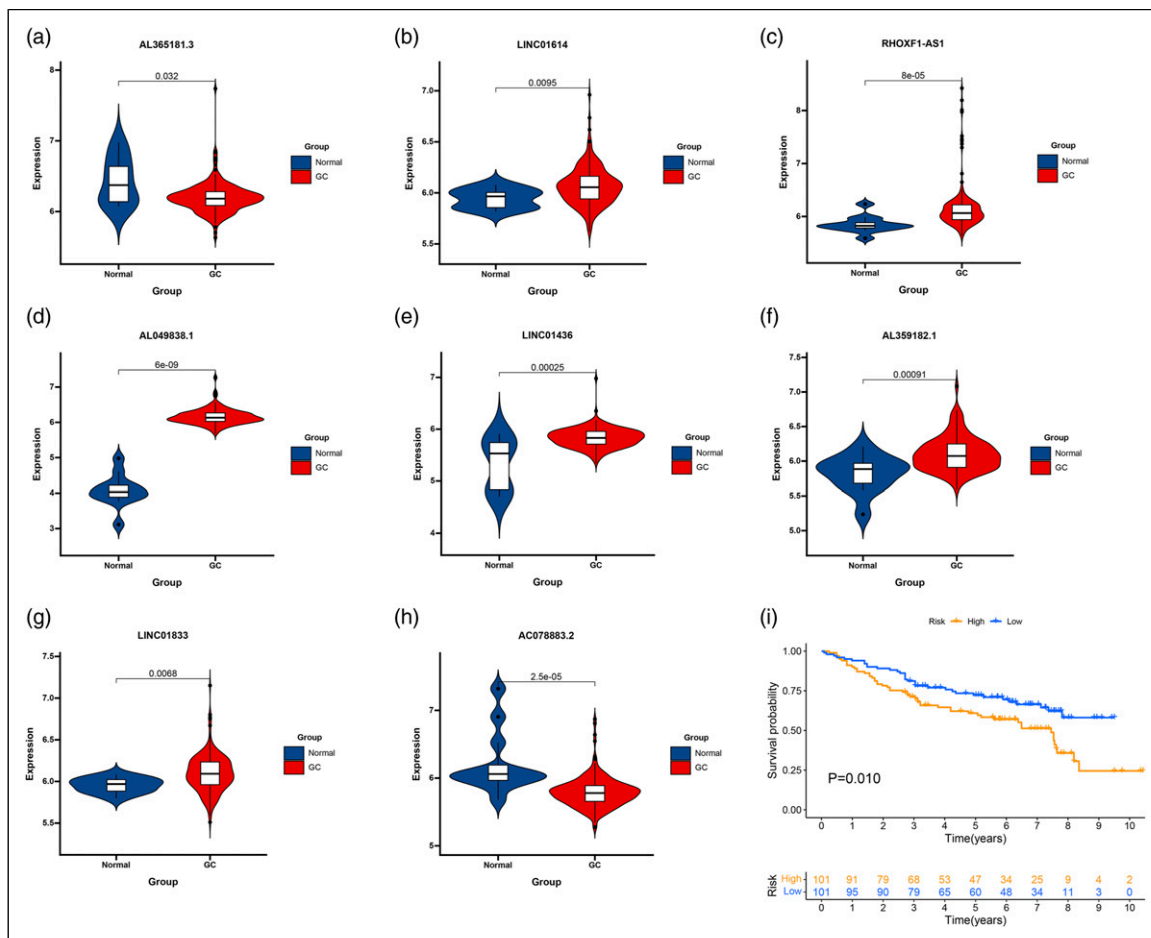
Figure 10. Validation of the prognostic performance of GIlncSig on GSE76427 datasets. (A) AL365181.3, (B) LINC01614, (C) RHOXF1-AS1, (D) AL049838.1, (E) LINC01436, (F) AL359182.1, (G) LINC01833, and (H) AC078883.2 expression levels are shown for the GC and normal tissue. (I) Kaplan–Meier curves illustrated that patients with high risk had worse overall survival than those with low risk (P value = 0.030).

high GILnc score group. Kaplan–Meier analysis revealed that the prognosis of patients with high GILnc score group was significantly poor than that of patients with low GILnc score group. In addition, the high GILnc score group was more sensitive to mitomycin. The above studies suggest that GILnc score classification has the potential to predict prognosis and guide medication.

Our study has some limitations. First, given the individual heterogeneity of GC, the results of our study should be further validated using more multicenter clinical data. Last, our findings have substantial implications for 8 GIlncs of GC, and the detailed molecular mechanisms require further research to explore deeper interactions.

Conclusion

In a conclusion, this study was focused on the analysis of GILnc in GC, and based on GILnc, we constructed the GILnc score model to explore the extensive regulation mechanisms by which genome instability affects tumorigenesis. The relationship between the scoring system and clinical outcomes in GC patients was demonstrated, and model validation was performed using external dataset from the GEO databases. This study provides novel insights into the role of genome instability in GC development.

ORCID iD

Minghui Ma 🄳 https://orcid.org/0000-0003-1593-8715

References

1. Siegel RL, Miller KD and Jemal A (2020) Cancer statistics, 2020. *CA A Cancer J Clin* 70: 7–30. DOI: 10.3322/caac.21590

2. Smyth EC, Nilsson M, Grabsch HI, et al. (2020) Gastric cancer. *Lancet* 396: 635–648. DOI: 10.1016/s0140-6736(20)31288-5

3. Blair VR, McLeod M, Carneiro F, et al. (2020) Hereditary diffuse gastric cancer: updated clinical practice guidelines. *Lancet Oncol* 21: e386–e397. DOI: 10.1016/s1470-2045(20)30219-9

4. Machlowska J, Baj J, Sitarz M, et al. (2020) Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int J Mol Sci* 21: 4012. DOI: 10.3390/ijms21114012

5. Liang Y, Zhang H, Song X, et al. (2020) Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets. *Seminars in Cancer Biology* 60: 14–27. DOI: 10.1016/j.semcancer.2019.08.012

6. Klein CA (2020) Cancer progression and the invisible phase of metastatic colonization. *Nature Reviews Cancer* 20: 681–694. DOI: 10.1038/s41568-020-00300-6

7. Ratti M, Lampis A, Hahne JC, et al. (2018) Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches. *Cellular and Molecular Life Sciences* 75: 4151–4162. DOI: 10.1007/s00018-018-2906-9

8. Tubbs A and Nussenzweig A (2017) Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* 168: 644–656. DOI: 10.1016/j.cell.2017.01.002

9. Andor N, Maley CC and Ji HP (2017) Genomic Instability in Cancer: Teetering on the Limit of Tolerance. *Cancer Research* 77: 2179–2185. DOI: 10.1158/0008-5472.Can-16-1553

10. Gerlach SU and Herranz H (2020) Genomic instability and cancer: lessons from Drosophila. *Open Biology* 10: 200060. DOI: 10.1098/rsob.200060

11. Wang T, Wang G, Zhang X, et al. (2017) The expression of miRNAs is associated with tumour genome instability and predicts the outcome of ovarian cancer patients treated with platinum agents. *Scientific reports* 7: 14736. DOI: 10.1038/s41598-017-12259-w

12. Sanchez Calle A, Kawamura Y, Yamamoto Y, et al. (2018) Emerging roles of long non-coding RNA in cancer. *Cancer Science* 109: 2093–2100. DOI: 10.1111/cas.13642

13. Kopp F and Mendell JT (2018) Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172: 393–407. DOI: 10.1016/j.cell.2018.01.011

14. Wei L, Sun J, Zhang N, et al. (2020) Noncoding RNAs in gastric cancer: implications for drug resistance. *Molecular Cancer* 19: 62. DOI: 10.1186/s12943-020-01185-7

15. Oh SC, Sohn BH, Cheong JH, et al. (2018) Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nature communications* 9: 1777. DOI: 10.1038/s41467-018-04179-8

16. Robinson MD and Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11: R25. DOI: 10.1186/gb-2010-11-3-r25

17. Simon N, Friedman J, Hastie T, et al. (2011) Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of Statistical Software* 39: 1–13. DOI: 10.18637/jss.v039.i05

18. Tibshirani R (1997) The lasso method for variable selection in the Cox model. *Statistics in Medicine* 16: 385–395. DOI: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2.-3

19. Carlevaro-Fita J, Liu L, Zhou Y, et al. (2019) LnCompare: gene set feature analysis for human long non-coding RNAs. *Nucleic Acids Research* 47: W523–W529. DOI: 10.1093/nar/gkz410

20. Jachimowicz RD, Beleggia F, Isensee J, et al. (2019) UBQLN4 represses homologous recombination and is overexpressed in aggressive tumors. *Cell* 176: 505–519. e22. DOI: 10.1016/j.cell.2018.11.024

21. Graziano F, Fischer NW, Bagaloni I, et al. (2020) TP53 mutation analysis in gastric cancer and clinical outcomes of patients with metastatic disease treated with ramucirumab/paclitaxel or standard chemotherapy. *Cancers* 12: 2049. DOI: 10.3390/cancers12082049

22. Sun D, Miao Y, Xu W, et al. (2020) Comprehensive analysis of competitive endogenous RNAs network reveals potential prognostic lncRNAs in gastric cancer. *Heliyon* 6: e03978. DOI: 10.1016/j.heliyon.2020.e03978

23. Wang Y, Zhang H and Wang J (2020) Discovery of a novel three-long non-coding RNA signature for predicting the prognosis of patients with gastric cancer. *Journal of Gastrointestinal Oncology* 11: 760–769. DOI: 10.21037/jgo-20-140

24. Puliga E, Corso S, Pietrantonio F, et al. (2021) Microsatellite instability in Gastric Cancer: Between lights and shadows. *Cancer Treatment Reviews* 95: 102175. DOI: 10.1016/j.ctrv.2021.102175

25. Hanahan D and Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646–674. DOI: 10.1016/j.cell.2011.02.013

26. Pikor L, Thu K, Vucic E and Lam W (2013) The detection and implication of genome instability in cancer. *Cancer and Metastasis Reviews* 32: 341–352. DOI: 10.1007/s10555-013-9429-5

27. Sieber OM, Heinimann K and Tomlinson IPM (2003) Genomic instability - the engine of tumorigenesis? *Nature Reviews Cancer* 3: 701–708. DOI: 10.1038/nrc1170

28. Abbas T, Keaton MA and Dutta A (2013) Genomic instability in cancer. *Cold Spring Harbor Perspectives in Biology* 5: a012914–a012914. DOI: 10.1101/cshperspect.a012914

29. Ovejero S, Bueno A and Sacristán MP (2020) Working on Genomic Stability: From the S-Phase to Mitosis. *Genes* 11: 225. DOI: 10.3390/genes11020225

30. Sieber O, Heinimann K and Tomlinson I (2005) Genomic stability and tumorigenesis. *Seminars in Cancer Biology* 15: 61–66. DOI: 10.1016/j.semcancer.2004.09.005

31. Petropoulos M, Champeris Tsaniras S, Taraviras S, et al. (2019) Replication Licensing Aberrations, Replication Stress, and Genomic Instability. *Trends in Biochemical Sciences* 44: 752–764. DOI: 10.1016/j.tibs.2019.03.011

32. Ben-David U, Beroukhim R and Golub TR (2019) Genomic evolution of cancer models: perils and opportunities. *Nature reviews. Cancer* 19: 97–109. DOI: 10.1038/s41568-018-0095-3

33. O'Shaughnessy J, Osborne C, Pippen JE, et al. (2011) Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *New England Journal of Medicine* 364: 205–214. DOI: 10.1056/NEJMoa1011418

34. Simmons AR, Clarke CH, Badgwell DB, et al. (2016) Validation of a Biomarker Panel and Longitudinal Biomarker Performance for Early Detection of Ovarian Cancer. *International Journal of Gynecologic Cancer* 26: 1070–1077. DOI: 10.1097/igc.0000000000000737

35. Ledermann J, Harter P, Gourley C, et al. (2012) Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *New England Journal of Medicine* 366: 1382–1392. DOI: 10.1056/NEJMoa1105535

36. Esteller M (2011) Non-coding RNAs in human disease. *Nature Reviews Genetics* 12: 861–874. DOI: 10.1038/nrg3074

37. Bhan A, Soleimani M and Mandal SS (2017) Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Research* 77: 3965–3981. DOI: 10.1158/0008-5472.Can-16-2634

38. Mercer TR, Dinger ME and Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* 10: 155–159. DOI: 10.1038/nrg2521

39. Schmitt AM and Chang HY (2016) Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* 29: 452–463. DOI: 10.1016/j.ccell.2016.03.010

40. Hu WL, Jin L, Xu A, et al. (2018) GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. *Nature Cell Biology* 20: 492–502. DOI: 10.1038/s41556-018-0066-7

41. Munschauer M, Nguyen CT, Sirokman K, et al. (2018) The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* 561: 132–136. DOI: 10.1038/s41586-018-0453-z

42. Li Z, Lü M, Zhou Y, et al. (2021) Role of Long Non-Coding RNAs in the Chemoresistance of Gastric Cancer: A Systematic Review. *OncoTargets and Therapy* 14: 503–518. DOI: 10.2147/ott.S294378

43. Pommier Y, Sun Y, Huang S-yN, et al. (2016) Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nature Reviews Molecular Cell Biology* 17: 703–721. DOI: 10.1038/nrm.2016.111

44. De Magis A, Manzo SG, Russo M, et al (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National Academy of Sciences* 116: 816–825. DOI: 10.1073/pnas.1810409116

45. Seeruttun SR, Yuan S, Qiu H, et al. (2017) A comprehensive analysis comparing the eighth AJCC gastric cancer pathological classification to the seventh, sixth, and fifth editions. *Cancer Medicine* 6: 2804–2813. DOI: 10.1002/cam4.1230

46. Jeong O, Jung MR, Kang JH, et al. (2020) Prognostic performance of preoperative staging: assessed by using multidetector computed tomography-between the new clinical classification and the pathological classification in the eighth american joint committee on cancer classification for gastric carcinoma. *Annals of Surgical Oncology* 27: 545–551. DOI: 10.1245/s10434-019-07845-3