Construction of a genome instability-derived lncRNA-based risk scoring system for the prognosis of hepatocellular carcinoma

Dan-Ping Huang[1,*], Mian-Mian Liao[1,2,*], Jing-Jing Tong[3,*], Wei-Qu Yuan[4], De-Ti Peng[1], Jian-Ping Lai[1], Yi-Hao Zeng[2], Yi-Jun Qiu[5,&], Guang-Dong Tong[1]

[1]Department of Hepatology, Shenzhen Traditional Chinese Medicine Hospital, The Fourth Clinical Medical College of Guangzhou University of Chinese Medicine, Shenzhen 518033, Guangdong Province, China

[2]College of Basic Medicine, Guangzhou University of Chinese Medicine, Guangzhou 510403, Guangdong, China

[3]The Affiliated Chencun Hospital of Shunde Hospital, Southern Medical University, Shunde 528300, Guangdong Province, China

[4]Department of Acupuncture, Shenzhen Traditional Chinese Medicine Hospital, The Fourth Clinical Medical College of Guangzhou University of Chinese Medicine, Shenzhen 518033, Guangdong Province, China

[5]The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510403, Guangdong Province, China

*Equal contribution

Correspondence to: Yi-Jun Qiu, Guang-Dong Tong; **email:** 576069949@qq.com, https://orcid.org/0000-0002-2402-5734; tgd755@163.com, https://orcid.org/0000-0002-8577-574X

ABSTRACT

Emerging evidence revealed the critical roles of long non-coding RNAs (lncRNAs) in maintaining genomic instability. However, genome instability-associated lncRNAs (GILncRNAs) and their performance in clinical prognostic significance in hepatocellular carcinoma (HCC) are rarely reported. Our study constructed a computational framework integrating somatic mutation information and lncRNA expression profiles of HCC genome and we identified 88 GILncRNAs of HCC. Function enrichment analysis revealed that GILncRNAs were involved in various metabolism processes and genome instability of cancer. A genome instability-derived lncRNA-based gene signature (GILncSig) was constructed using training set data. The performance of GILncSig for outcome prediction was validated in testing set and The Cancer Genome Atlas (TCGA) set. The multivariate cox regression analysis and stratification analysis demonstrated GILncSig could serve as an independent prognostic factor for the overall survival of HCC patients. The time-dependent Receiver Operating Characteristic (ROC) curve illustrated GILncSig outperformed two recently published lncRNA signatures for overall survival prediction. The combination of GILncSig and tumor protein p53 (TP53) mutation status exhibited better prognostic performance in survival evaluation compared to TP53 mutation status alone. AC145343.1 was further validated to be a risk factor for HCC *in vitro among GILncSig*. Overall, our study provided a novel approach for identification of genome instability-associated lncRNAs and established an independent risk score system for outcome prediction of HCC patients, which provided a new insight for exploring in-depth mechanism and potential therapy strategy.

INTRODUCTION

Liver cancer is the second leading cause of tumor-related death and disability-adjusted living-years causing more than 700,000 deaths each year [1]. There are about 840,000 people suffering from liver cancer, but the death toll is as high as 780,000 every year worldwide, most of which occur in developing countries according to the global cancer statistics in 2018 [2]. Although sensitivity and specificity of early diagnosis of liver cancer have improved to a certain extent, the 5-year survival rate is still only about 17%. Hepatocellular carcinoma (HCC) as the most important type of liver cancer accounting for 85% to 90% is critical for liver cancer research. Liver resection and liver trans-plantation are the main treatments for HCC, but only a few patients are eligible for surgery and approximately 70% of patients undergoing surgery will relapse within 5 years after surgery [3]. In addition, the symptoms of HCC commonly appear late leading to most HCC patients being diagnosed with advanced cancer accompanied by intrahepatic or distant metastasis with poor effect of surgical treatment and the prognosis of HCC is still very poor with a 5-year survival rate of less than 20% [4, 5]. To improve clinical outcomes for patients, serum biomarkers of HCC were being continuously developed. An abnormal increased level of plasma AFP in adults is considered to be a hallmark of pathological conditions of HCC [6]. AFP is over-expressed in more than 70% of clinical HCC patients. Serum AFP level has been considered the 'gold standard' biomarker for clinical liver cancer diagnosis over the last few decades [7]. A protein induced by vitamin K absence or antagonist-II (PIVKA-II) is also useful for the diagnosis of early HCC and has been used as a predictive marker of microvascular invasion [8]. In addition, 1 specific type of AFP—AFP-L3—binds to a lectin and displays serum levels that are in consistent with levels of AFP in human sera [9]. AFP-L3 can be used to differentiate an increase in AFP due to HCC or benign liver disease [10–12]. However, these biomarkers at the protein level are still limited by low sensitivity and specificity [13]. Genome instability, an increase in the tendency to acquire genomic changes ranging from base pair mutations to chromosomal aberrations, contributes to somatic cell heterogeneity and genetic diversity as a material for natural and artificial selection, while it also contributes to the progression of genetic related diseases including cancer [14, 15]. Genome instability is the basic feature of tumor cells and the core sign of tumorigenesis, and the evolution from early atypical hyperplasia to malignant and metastatic tumors is often accompanied by increasing genomic instability [16]. Moreover, genomic instability is closely related to tumor progression and affects prognosis and survival [17]. The causes of genome instability are very complicated which may be associated with replication dysfunction, DNA repair failure, abnormal transcription, various metabolism process and post-transcriptional regulation.

It is clear that long non-coding RNA (lncRNA) is becoming a potential regulator and quantitative measurement of genome instability [18, 19]. LncRNA mostly transcribed by RNA polymerase II without protein-coding function is defined as an RNA transcript with more than 200 nucleotides located in the nucleus or cytoplasm. LncRNA participates in cell cycle, differentiation, cell migration, invasion, proliferation and apoptosis, and functions as a cell microstructure original and small RNA precursor [20, 21]. Transcription and dysfunction of lncRNA is closely involved in tumorigenesis, including HCC [22]. Moreover, lncRNA with time and tissue specificity shows different expression levels in tumors and healthy states and different tumor stages, which suggests lncRNA has great potential to be a new prognostic biomarker [23]. For example, a controlled study involving 80 HCC patients and 50 healthy subjects finds that the expression of cancer susceptibility candidate 9 in HCC patients significantly increases with area under the curve (AUC) at 0.933 [24]. Study shows the volume of liver cancer tumors reduces by 82% in animal models of liver cancer when H19 (the first reported lncRNA) is knocked out, which fully confirms the tumorigenic effect of H19 [25]. Short-term recurrence after liver resection or liver transplantation in patients with HCC is related to the increasing expression of HOTAIR and the decreased expression of HOTAIR leads to the apoptosis of liver cancer cell lines [26]. The above researches show that lncRNAs are closely related to the progression of HCC, but the mechanism of regulating HCC is still elusive. Evidence shows that lncRNA participates in gene expression at the transcription and post-transcriptional levels, thus involves in regulating genomic instability [27]. However, genome instability-associated lncRNAs (GILncRNAs) and their clinical prognostic significance in HCC are rarely reported.

In this study, we constructed a computational framework integrating somatic mutation profiles and lncRNA expression profiles of HCC to recognize GILncSig for HCC, and confirmed the performance of GILncSig on HCC. We further validated the biology function of the most important lncRNA of the GILncSig-AC145343.1 with Cell Counting Kit-8 (CCK-8), colony formation, transwell and wound healing assay. Our study revealed a novel approach for identification of genome instability-associated lncRNAs and established an independent signature for outcome prediction of HCC.

RESULTS

Identification of GILncRNAs in HCC patients

The cumulative somatic mutations for each sample were first computed and ranked in the decreasing order. The top 25% samples and the last 25% ones in the ranking list were defined as genomic unstable (GU)-like group and genomic stable (GS)-like group, respectively (Figure 1A). Each group was consisted of 91 samples. Next, significant expressed lncRNAs were identified by comparing the lncRNA expression profiles between GU-like group and GS-like group. 88 lncRNAs were obtained based on the criteria of |fold change| > 1.5 and FDR adjusted $P<0.05$ and served as GILncRNAs. Of these, 32 lncRNAs were upregulated and 56 were downregulated in GU-like group (Supplementary Data 2). Then, we performed unsupervised hierarchical clustering analysis of all 364 samples using the filtered expression profile of differently expressed GILncRNAs. All samples were hence re-grouped based on the cluster result (Figure 1B). The group with higher cumulative somatic mutations was defined as GU-like group, and the other group was named as GS-like group. As shown in Figure 1B, the somatic mutation pattern was significantly different between the two groups. The median value of somatic cumulative mutations was 156.7 in the GU-like group while 124.0 in the GS-like group ($P<0.001$) (Figure 1C).

To validate the potential function of identified 88 lncRNAs and uncover their association with genomic instability, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were carried out using clusterProfiler software in R-version 3.5.2. Firstly, protein coding genes (PCGs) closely related to the expression of lncRNAs were screened out using Pearson Correlation Coefficients. The top 10 PCGs most correlated with each lncRNAs were retained for lncRNAs–mRNA co-expression network construction. In the co-expression network, the nodes represented lncRNAs and mRNAs, and the lines represented the relationships between lncRNAs and mRNAs (Figure 1D). Go analysis of the PCGs in the network indicated that genomic instability might impact various metabolism process including fatty acid / fatty acid derivative metabolic process, pentose metabolic process, cellular aldehyde metabolic process, small molecule catabolic process and so on (Figure 1E). In terms of KEGG pathway analysis, PCGs in the network were found to enriched in a variety of metabolism pathway, HIF-1 signal pathway, extracellular

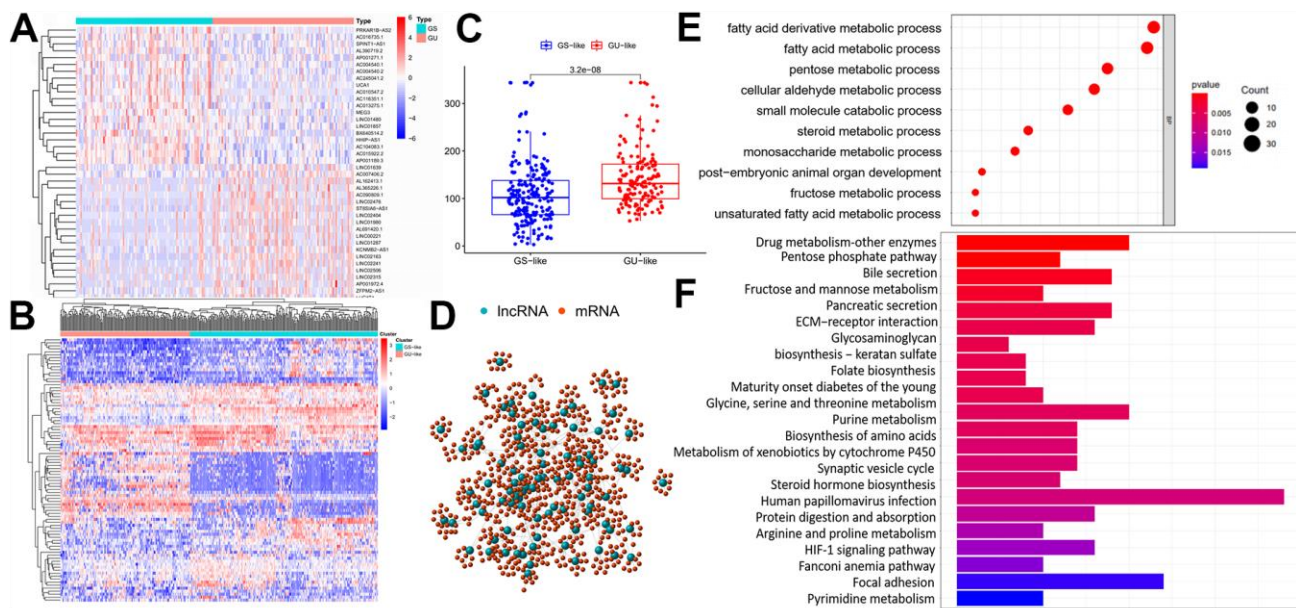**Figure 1. Identification and functional annotations of genomic instability-related lncRNAs.** (**A**) GU-like group and GS-like group identification according to the top 25% samples and the last 25% ones in the ranking list of cumulative somatic mutations. The left blue cluster is GS-like group, and the right red cluster is GU-like group. (**B**) Unsupervised clustering of 364 HCC patients based on the expression pattern of 88 candidate genomic instability-related lncRNAs. The left orange cluster is GU-like group, and the right blue cluster is GS-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**) Coexpression network of genomic instability-related lncRNAs and mRNAs. The red circles represent mRNAs, and the blue circles represent lncRNAs. (**E**, **F**) Functional enrichment analysis of GO and KEGG for mRNAs co-expressed lncRNAs.

matrix-receptor interaction, biosynthesis of amino acids and so on (Figure 1F). Enrichment results further reveal genomic instability-related lncRNAs were involved in various biological process of cancer. Expression of lncRNA could break the regulatory balance between lncRNA and PCGs, subsequently interfere with a variety of metabolism pathway, leading to gene damage repair process and exacerbation of genomic instability.

Construction of GILncSig for outcome prediction in the training set

To explore the role of GILncRNAs in the prognosis of HCC patients, 343 samples downloaded from TCGA were randomly divided into 2 groups named the training set (n=172) and testing set (n=171) respectively. As shown in Supplementary Table 1, there was no significant difference observed in the common clinical features characteristic between two groups ($P > 0.05$, Chis-square test). Then training set were used to establish GILncSig of HCC patients. We performed univariate Cox proportional hazard regression analysis to investigate the association of GILncRNAs and overall survival (OS) of HCC patients in the training set and found 9 GILncRNAs were closely related to the prognosis of HCC patients ($P<0.05$; Supplementary Table 2). Furthermore, multivariate Cox regression analysis was performed to evaluate the independent prognostic value of 9 GILncRNAs. 3 of 9 candidate lncRNAs including AC145343.1, AC004862.1 and ZFPM2-AS1 were obtained with prognostic significance in multivariate Cox analysis ($P<0.05$) (Table 1). Finally, GILncSig was established to predict outcome of HCC patients in the training set according to the equation mentioned above. The GILncSig was constructed as follow: GILncSig score = (0.3804 × expression level of AC145343.1) + (0.1253 × expression level of ZFPM2-AS1) + (−0.2344 × expression level of AC004862.1). A positive/negative regression coefficient demonstrated a positive/negative association between risk score and the expression level of lncRNA. Higher GILncSig score means higher risk of poor prognosis. Of the GILncSig, AC145343.1 and ZFPM2-AS1 tended to be risky factors while AC004862.1 was more likely to be a protective factor for the survival of HCC patients.

The GILncSig of each sample in the training set were calculated and then these patients were equally divided into high risk group and low risk group according to computed risk score with a decreasing order. Next we investigated the survival time of patients in the two groups using Kaplan–Meier analysis. The result showed that longer survival of patients in the low-risk group compared that in the high-risk.

Group ($P<0.001$, log rank test; Figure 2A). The time-dependent ROC curve was illustrated in Figure 2B and demonstrated an AUC of 0.781 for the GILncSig. We also plotted the expression levels of lncRNAs in the GILncSig and the count of somatic mutations with the increasing score in patients of the training set. As shown in Figure 2C, the risk lncRNA AC145343.1 and ZFPM2-AS1 showed up-regulated expression while the protective lncRNA AC004862.1 showed opposing expression pattern in the samples with high risk scores. Comparison analysis showed the number of somatic mutation between two groups has no significance ($P = 0.16$) but the average count of somatic mutation in patients of high-risk group was higher than that of low-risk group (Figure 2D).

Independent examination of GILncSig for HCC patients

The RNA-seq data of testing group including 171 HCC samples was analyzed to assess the stability, validity and prediction ability of the GILncSig. Patients in the testing group were also evenly separated into the high-risk group and low-risk group based on their GILncSig score. Kaplan–Meier survival analysis showed that patients in the high-risk group had a shorter survival time compared to those in the low-risk group ($P<0.05$) (Figure 3A; Left panel). The time-dependent ROC curves analysis of the GILncSig in the testing set yielded an AUC of 0.665 (Figure 3B; Left panel). Then we sorted the samples in the testing group according to GILncSig score with the increasing order to further demonstrate the alteration of the expression levels of the GILncSig (Figure 3C; Left panel). Concurring with the expectation, the risk lncRNAs AC145343.1 and ZFPM2-AS1 were more highly expressed and the protective lncRNA AC004862.1 showed lower expression pattern in the patients of testing set with higher risk scores. The number of the distribution of somatic mutation in patients with high scores was significantly higher than in patients with low scores ($P<0.05$, Figure 3D; Left panel).

Similar results have been obtained in the whole TCGA samples in terms of the prognostic value of the GILncSig. Kaplan–Meier plot analysis demonstrated that patients in the low-risk group had a more favorable survival time compared to those in the high-risk group ($P<0.001$) (Figure 3A; Right panel). As for ROC curve, the TCGA set yielded an AUC of 0.728 (Figure 3B; Right panel). The expression patterns of lncRNAs AC145343.1, ZFPM2-AS1 and AC004862.1 in TCGA set were consistent with that in the training and testing group (Figure 3C; Right panel). The high-risk group also showed increased somatic mutation counts compared to the low-risk group ($P<0.05$) (Figure 3D; Right panel).

Table 1. Multivariate Cox regression analysis.

Gene symbol	Coefficient	HR	95% CI	P-value
AC145343.1	0.380	1.463	1.134-1.888	0.003
AC004862.1	-0.234	0.791	0.657-0.952	0.013
ZFPM2-AS1	0.125	1.133	1.076-1.193	1.57E-06

Independence of the GILncSig from common clinical variables

To access the prognostic value of other clinical factors, we first conducted univariate Cox regression analyses on age, gender, tumor grade, tumor stage and GILncSig of each set. The result showed that only tumor stage and GILncSig were closely related to the survival time of HCC patients in each set (Table 2). Multivariate Cox regression was further analyzed to explore the independence performance of the GILncSig. Tumor stage and GILncSig also exhibited significant difference in each group based on the multivariate analysis result (Table 2). Then we continued to examine whether the prognosis performance of GILncSig was independent of the tumor stage. We excluded the patients with
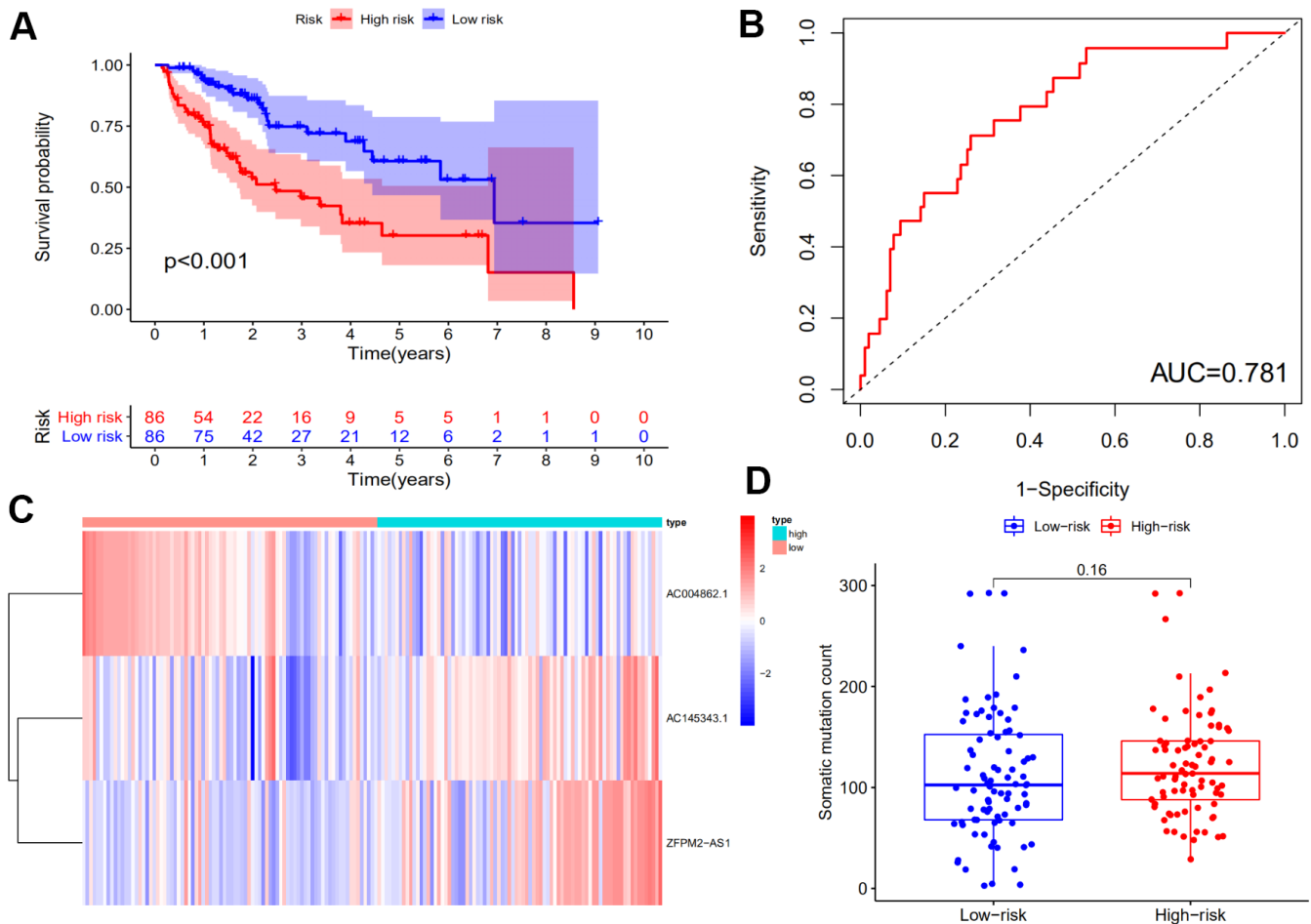

Figure 2. Identification of the genomic instability-derived lncRNA signature (GILncSig) using the training set. (**A**) Kaplan–Meier analysis of overall survival of patients with low or high risk according to the GILncSig score in the training set. Statistical analysis was performed using the log-rank test and univariate Cox analysis. (**B**) Time-dependent ROC curves analysis of the GILncSig. (**C**) LncRNA expression patterns with increasing GILncSig score. (**D**) Somatic mutations count in the high- and low-risk groups for the training set patients. The red represents the high-risk group, and the blue represents the low-risk group.
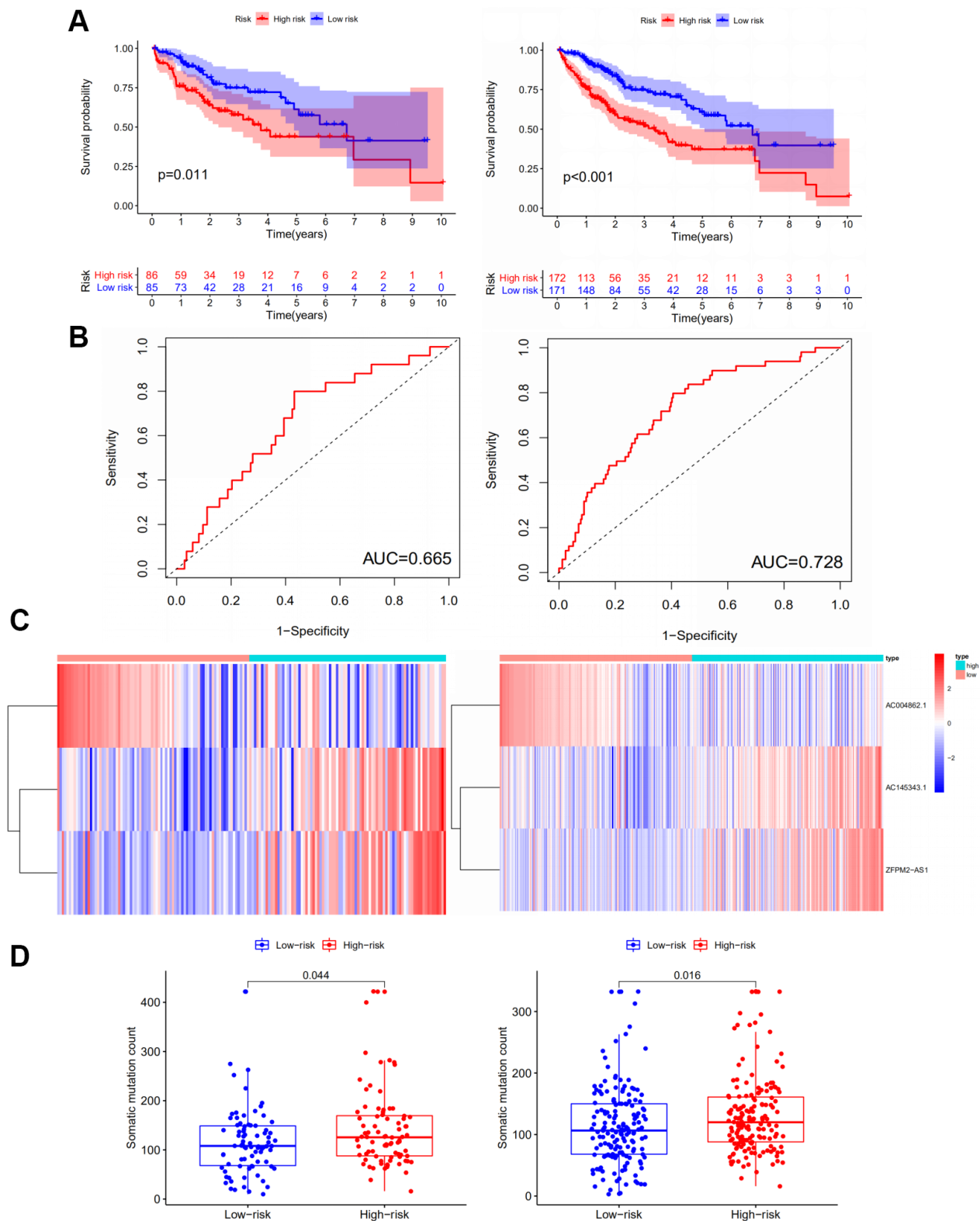
Figure 3. Performance examination of the GILncSig in the testing set and TCGA set. (**A**) Kaplan–Meier analysis of overall survival of patients with low or high risk according to the GILncSig score in the testing set (Left panel) and TCGA set (Right panel). Statistical analysis was performed using the log-rank test and univariate Cox analysis. (**B**) Time-dependent ROC curves analysis of the GILncSig in the testing set (Left panel) and TCGA set (Right panel). (**C**) LncRNA expression patterns with increasing GILncSig score in the testing set (Left panel) and TCGA set (Right panel). (**D**) Somatic mutations count in the high- and low-risk groups for the patients in the testing set (Left panel) and TCGA set (Right panel). The red represents the high-risk group, and the blue represents the low-risk group.

Table 2. Univariate and multivariate Cox regression analysis of the GILncSig and overall survival in different patient sets.

Variables		Univariable model			Multivariable model		
		HR	95% CI	P-value	HR	95% CI	P-value
Training set (n =182)							
GILncSig	High/Low	1.129	1.076	8.37E-07	1.153	1.096-1.213	2.96E-08
Age		1.004	0.984	1.025			
Gender		0.721	0.419	1.241			
Grade		1.300	0.932	0.122			
Stage		1.811	1.353	6.51E-05	1.931	1.428-2.611	1.90E-05
Testing set (n =182)							
GILncSig	High/Low	1.023	0.922-1.136	6.64E-03	1.102	1.042-1.145	2.23E-03
Age		1.007	0.986-1.029	0.497			
Gender		0.792	0.451-1.391	0.417			
Grade		0.887	0.593-1.327	0.560			
Stage		1.849	1.355-2.523	1.04E-04	1.849	1.355-2.523	1.04E-04
TCGA set (n =364)							
GILncSig	High/Low	1.091	1.042-1.141	1.68E-04	1.114	1.061-1.168	1.39E-08
Age		1.005	0.991-1.020	0.481			
Gender		0.758	0.513-1.118	0.162			
Grade		1.121	0..865-1.454	0.388			
Stage		1.808	1.463-2.234	4.31E-08	1.866	1.505-2.315	1.11E-05

unknown stage and stratified remaining samples into early-stage group (stage I-II, n = 238) and late-stage group (stage III-IV, n = 83) for stratification analysis. Samples in each group were further separated into high-risk and low-risk group. A significant difference in survival outcome between high-risk and low-risk group was observed in both early-stage group and a late-stage group ($P<0.01$) (Figure 4). It was clear that the GILncSig could serve as an independent prognostic factor for the overall survival of HCC patients.
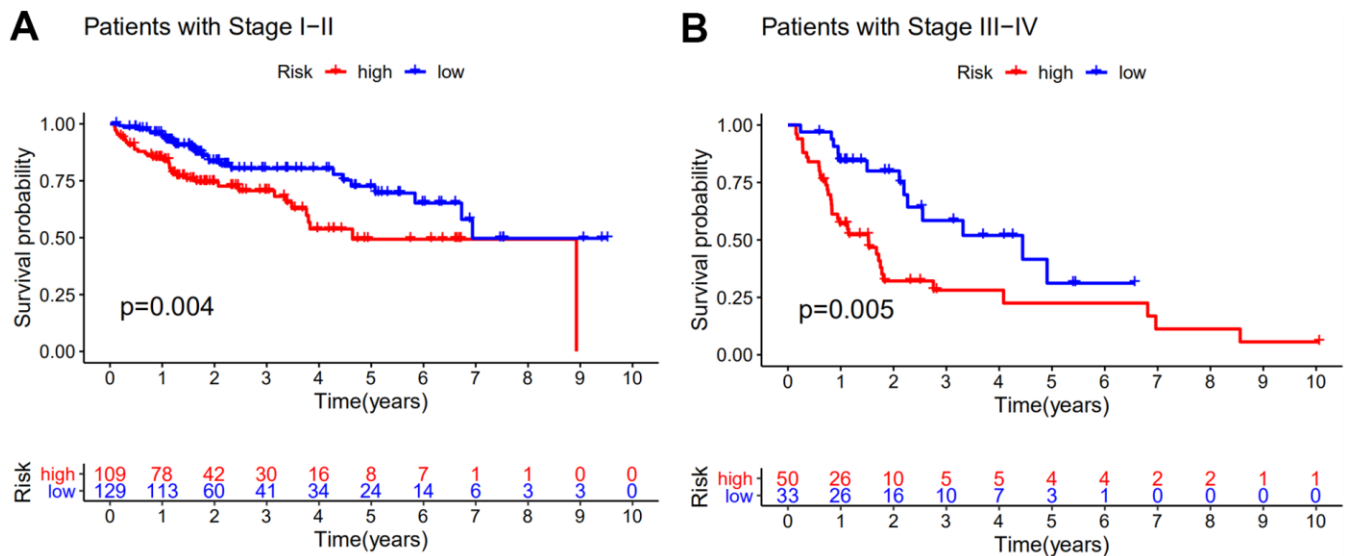

Figure 4. Stratification analyses by stage. Kaplan–Meier curve analysis of overall survival of patients in high- and low-risk groups for early-stage patients (**A**) and late-stage patients (**B**). Statistical analysis was performed using the log-rank test and univariate Cox analysis.

Comparison of the GILncSig with existing lncRNA-related signatures in prognosis value

To further illustrate the performance of GILncSig in our study, we recruited two recently published lncRNA signatures for survival prediction of HCC patient. Li's study included 12-lncRNA signature (LilncSig) and Ma's study included 4-lncRNA signature (MalncSig) [28, 29]. For LilncSig, Li et al. analyzed 12 pairs of HCC and adjacent normal mucosal tissues and identified 3900 differentially expressed lncRNAs as candidate biomarkers for the prognosis of HCC [28]. The 12-lncRNA signature was constructed using the least absolute shrinkage and selection operator (LASSO) cox regression method [8]. For MalncSig, the probe expression profiles of 225 HCC samples and 220 paired non-tumor tissue samples were derived from Gene Expression Omnibus (GEO)-GSE14520 [29]. Univariate cox regression and LASSO model were applied to screen lncRNAs linking to the overall survival. Then the multivariate Cox regression model was implemented to construct the prognostic score model [20]. Comparison analysis was performed between GILncSig and two recruited lncRNA signatures. As shown in Figure 5, the AUC of overall survival (OS) for the GILncSig is 0.728, which is significantly higher than that of LilncSig (AUC = 0.619) and MalncSig (AUC = 0.575). In addition, GILncSig consisted of 3 lncRNAs while MalncSig/

LilncSig included 4/12 lncRNA for outcome prediction. Based on the AUC and lncRNA number, our optimized GILncSig significantly outperformed the two recently published lncRNA signatures in the sense of prognostic performance.

Comparison of the GILncSig with TP53 mutation status in prognosis value.

It is reported that mutation of TP53 gene is associated with worse survival in cancer patients [30]. Statistic results suggested that patients in the high-risk group have a significantly higher percentage of TP53 mutations than patients in the low-risk group among the training set, testing set and all TCGA samples (Figure 6A) ($P<0.001$). The proportions of TP53 mutation in the high-risk group of the training set, testing set and TCGA set were 49%, 42% and 45% respectively. In the low-risk group of the training set, testing set and TCGA set, TP53 mutation proportions were 20%, 12% and 16% respectively. TP53 transcriptionally targets hundreds of genes and regulates the expression of gene contributed to cell cycle, apoptosis, DNA repair proteins and metabolic [31, 32]. Recent studies had demonstrated that mutation of TP53 increased genomic instability and served as an independent prognostic marker [33–37]. Therefore, we continued to compare the performance of the GILncSig and TP53 mutation status in prognosis value. According to the
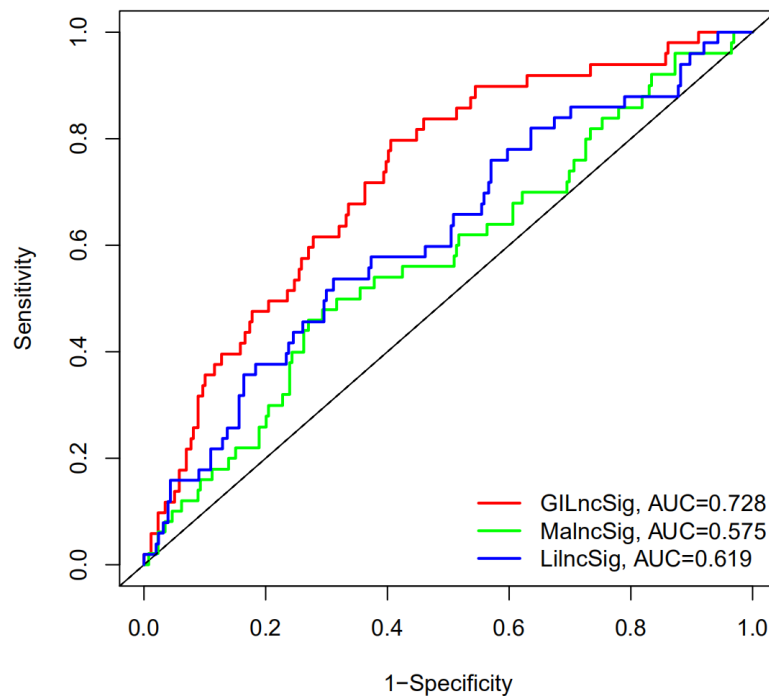

Figure 5. The ROC analysis of overall survival (OS) for the LilncSig and MalncSig. The AUC of OS for the GILncSig, LilncSig and MalncSig is 0.728. 0.619 and 0.575, respectively.

GILncSig and TP53 mutation status, we classified all samples into TP53 Mutation/GS−like group, TP53 Mutation/GU−like group, TP53 Wild/GS−like group and TP53 Wild/GU−like group. Figure 6B revealed the survival curve of four risk groups. With regard to GU−like patients, the survival of TP53 Mutation group was more closely resembles that of TP53 Wild group. However, for patients with TP53 mutation, the survival of GU−like group was not similar to the GS−like group. Moreover, the survival outcome of patients in both GU−like group and TP53 mutation group were significantly worse than that in both GS−like group and TP53 wild group, indicating that the GILncSig and TP53 mutation status exhibited better prognostic performance than TP53 mutation status alone.

Unfavorable impact of AC145343.1 on HCC

Of the GILncSig, AC145343.1 served as a risky factor and the most important lncRNA for prognostic prediction according to the regression coefficient. Hence, we further assessed the function of AC145343.1 with regards to HCC. Firstly, we sought to characterize the potential prognostic ability of AC145343.1 indicated by survival analysis using The Encyclopedia of RNA Interactomes (ENCORI) as described in our previous study [38]. A significant decline of survival time in high AC145343.1 set were observed (Figure 7A). To evaluate the phenotype effect of AC145343.1 *in vitro*, we transfected siRNA into HepG2 cell to establish AC145343.1-downregulated cell line. To assess the proliferation inhibitory potential of AC145343.1 in HCC, we employed CCK-8 and colony

formation assay in HepG2 with/without AC145343.1 downregulation. After AC145343.1 silencing, HepG2 exhibited obviously lower cell viability and significantly decreased colony area compared to the control group (Figure 7B, 7C). Invasion and migration roles of AC145343.1 were further observed, manifested by transwell assays and wound healing array. Transwell assay indicated silencing AC145343.1 remarkably decreased the number of HepG2 that migrated across the transwell chamber (Figure 7D). Wound healing array also revealed that AC145343.1-downregulated HepG2 exerted a significant delay in wound healing when compared with the control group (Figure 7E). Taken together, these results supported AC145343.1 knockdown repressed the proliferative, migratory and invasive abilities of HepG2 cells.

DISCUSSION

The onset is hidden, the early symptoms are not obvious and the clinical manifestations of HCC are quite different making early diagnosis and prognosis difficult. Imaging examination plays an important role in the diagnosis of HCC, but the diagnostic sensitivity is greatly reduced and restricted when the lesion is small [39]. AFP is currently the most widely used biomarker for the diagnosis of HCC with low sensitivity and specificity [13]. Therefore, it is necessary and urgent to look for the new prognostic evaluation indicators in order to improve the prognosis of HCC. With the rapid development of high-throughput sequencing technology, genomic instability-related lncRNA is gradually being discovered to be a potential biomarker of prognostic evaluation indicator [18, 19].
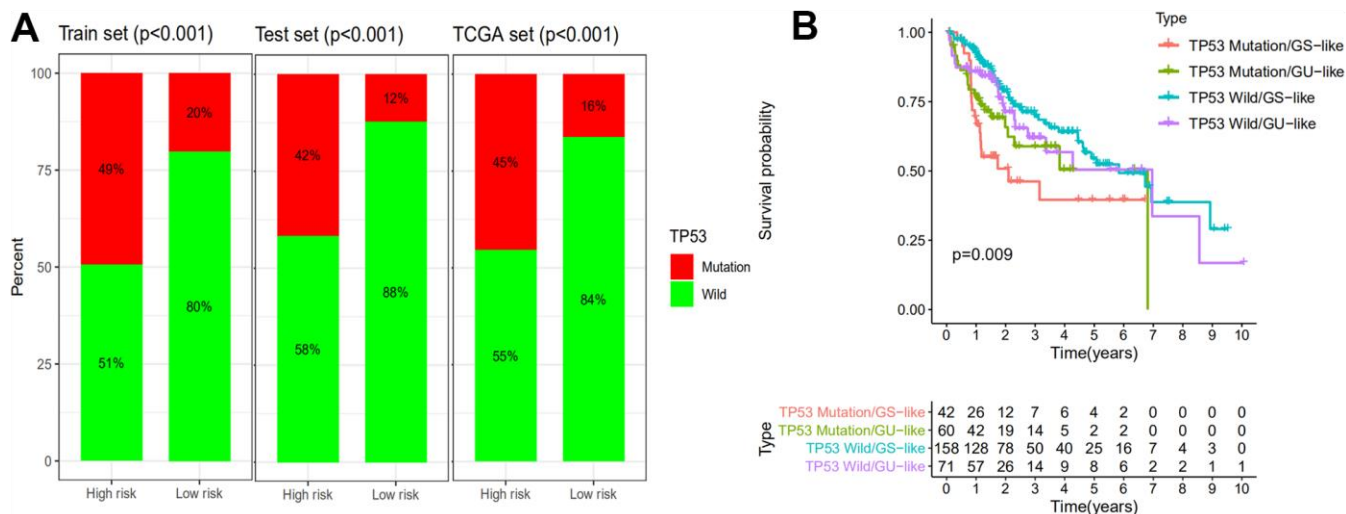

Figure 6. Comparison of the GILncSig with TP53 mutation status in prognosis value. (**A**) The proportion of TP53 mutation in high- and low-risk groups in the training set, testing set and the TCGA set. (**B**) Kaplan–Meier curve analysis of overall survival is shown for patients classified according to TP53 mutation status and the GILncSig. Statistical analysis was performed using the log-rank test.

LncRNA plays an important regulatory role in chromosome modification, nuclear transcription and cytoplasmic post-transcriptional processing, and can be used as a tissue factor of subcellular structure to regulate the location or activity of the protein [40, 41]. Abnormal expression of lncRNA is closely related to human diseases, especially in tumors. The abnormal expression of lncRNA has been found in HCC, and it is involved in tumor growth, infiltration, metastasis and recurrence suggesting that lncRNA may become a new prognostic marker in the occurrence and development of HCC [42, 43]. Studies have shown that lncRNA is an emerging regulator of genomic instability, such as BGL3 and NORAD [18, 44]. However, the identification of lncRNAs related to genomic instability and their prognosis and clinical significance for HCC are still unclear. Thus, we constructed the GILncSig with a computational frame-work integrating somatic mutation information and lncRNA expression profiles to confirm the roles of lncRNAs related to genomic instability in prognosis for HCC.

First, we integrated the lncRNA profile with the somatic mutation profile of HCC for a comprehensive analysis, and obtained 88 lncRNAs with |fold change| > 1.5 and FDR adjusted $P<0.05$ related to genome instability. Then, PCGs closely related to lncRNAs were screened out to perform Go and KEGG pathway analysis. We found that biological processes and biological pathways were mainly involved in various small molecule metabolic/catabolic process, a variety of metabolism pathway, HIF-1 signal pathway, biosynthesis of amino acids and so on. The normal cell cycle is the basic condition to ensure the correct sequence, integrity and fidelity of life activities and study suggests metabolic
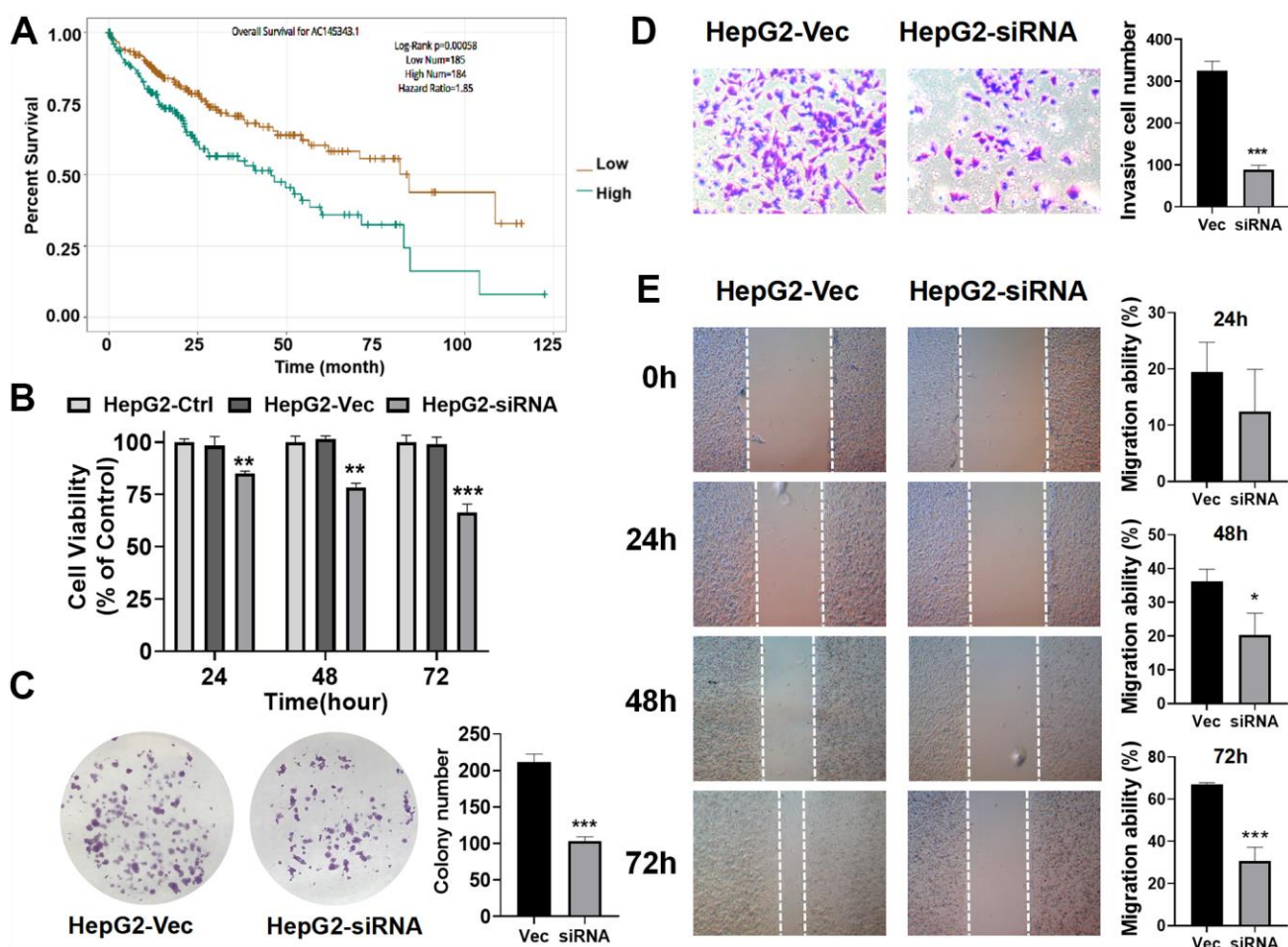

Figure 7. Unfavorable impact of AC145343.1 on HCC *in vitro*. (**A**) Kaplan–Meier curve of expression level of AC145343.1 on HCC patients using ENCORI. (**B**) Cell viability of HepG2 was significantly inhibited after AC145343.1 silencing. (**C**) Colony formation number was significantly decreased after AC145343.1 silencing compared to the control group. (**D**) Invasive ability of HepG2 was suppressed after AC145343.1 silencing manifested by transwell experiment. (**E**) Wound healing array demonstrated that AC145343.1-downregulated HepG2 cell exhibited obviously delay in wound healing when compared with the control group. $^*P<0.05$, $^{**}P<0.01$, $^{***}P<0.01$.

dysfunction linking to DNA damage causes dysregulated cell cycle, which will lead to genomic instability [45, 46]. In addition, research shows that metalloproteinase SPRTN involved in biosynthesis of amino acids and metabolism regulates covalent DNA-protein crosslinks to prevent genome instability and carcinogenesis [47]. HIF, a major participant in sensing and adapting to hypoxia, is closely related to genome instability and cancer progression [48]. The results of Go and KEGG pathway analysis further proved PCGs closely related to lncRNAs were mainly involved in genome instability, which is an important feature of cancer pathogenesis [46]. We further explored the roles of GILncRNAs in the prognosis prediction of HCC patients and received 3 candidate GILncRNAs including AC145343.1, AC004862.1 and ZFPM2-AS1 by using multivariate cox regression analysis. Patients was divided into two groups according to GILncSig score. The low risk group showed a longer survival time compared with high risk group with significance in the training set. The same result was validated in the testing set and TCGA set by using Kaplan–Meier and time-dependent ROC curve analysis. It is worth noting that the AUC of ROC curve for GILncSig in the training set, testing set and TCGA set were respectively achieved 0.781, 0.665 and 0.728 revealing outstanding performance of GILncSig for prognosis prediction. Meanwhile, multivariate cox regression analysis showed the predictive significance of GILncSig was independent of other clinical factors, further elaborating the reliability of GILncSig for HCC prognosis prediction. Moreover, we found the number of somatic mutation in high risk group was higher than low risk group with significant in testing set and TCGA set, demonstrating GILncSig was significantly associated with HCC mutator phenotype, which is important for assessing genome instability and prognosis. After systematically reviewing the literature, we found that ZFPM2-AS1 upregulated in HCC tissues and involved in cell cycle progression with miR-653 binding sites can reverse the inhibitory effect of miR-653 on the proliferation and metastasis of HCC cells by regulating the target gene GOLM1 of miR-653, and regulate the process of HCC by binding to miR-139 to regulate the expression of GDF10 [49, 50]. However, no previous reports describing the function of AC145343.1 and AC004862.1 until now. We found that the lncRNA AC145343.1 is located in chromosome 17q24 [51]. Genetic variations in the region of chromosome 17q24 are predictors of prostate cancer and lung malignancy risk [52–54]. The AC004862.1 gene is localized in chromosome 7q21.11. Previous studies showed frequent amplification of 7q21 was found in Barrett's and gastric cardia cancers and associated with early neoplastic lesions [55, 56]. The amplicon in the 7q21

area is known for breast cancer metastasis-related susceptibility loci in previous genome wide analysis [57]. In this study, we first propose that lncRNA AC145343.1 has great potential to serve as a risk factor and lncRNA AC004862.1 is recognized as a protective factor for HCC prognosis performing crucial role in the development of HCC. However, further research is needed to understand the deeper mechanisms. The present results in our study and available published literatures reveal the GILncSig has great potential to perform prognosis prediction and are very likely to become indicators of genome instability for HCC patients at the same time.

To further confirm the performance in prognosis prediction of GILncSig, we recruited two recently published lncRNA signatures of survival prediction for HCC patient [28, 29]. We found that the AUC for the GILncSig with lower number of lncRNA was higher than that of LilncSig and MalncSig, indicating our GILncSig possessed more optimized prognostic effect. Studies show that TP53 mutation increases genomic instability and serves as an independent prognostic marker [33–37]. Cells with DNA damage can avoid apoptosis then transform into cancer cells in the event of TP53 mutation. In HCC, TP53 alterations are found to be associated with serum AFP level, tumor stage, vascular invasion, tumor differentiation and Child-Pugh class [58–61]. Meanwhile, HCC patients with TP53 mutations have shorter OS and relapse-free survival times [62]. Consistent with previous research, patients in high risk group showed a higher TP53 mutation rate than those in low risk group in training set, testing set and TCGA set, strongly showing GILncSig can reflect the TP53 mutation status. Then, we further compared the prognosis value of GILncSig and TP53 mutation. The survival curve of TP53 Mutation/GU−like group was more closely resembles that of TP53 Wild/GU−like group but not that similar to the TP53 Mutation/GS−like group, indicating that GILncSig rather TP53 mutation was more closely associated with the overall survival of HCC patients, suggesting TP53 mutation status alone does not perform well for predicting outcome of GU−like patients. Remarkably, patients in both GU−like group and TP53 mutation group had a shorter survival time compared to that in both GS−like group and TP53 wild group, indicating that the GILncSig and TP53 mutation status exhibited better prognostic performance than TP53 mutation status alone. Finally, AC145343.1 was regarded as the most relevant one for outcome prediction. A significant decline of survival time in patients with high AC145343.1 was observed, indicating AC145343.1 exerts pro-cancer effect in human HCC among the lncRNAs in GILncSig. Therefore, we conducted the

molecular biology experiments *in vitro* to validate the effect of AC145343.1 on HCC. CCK-8 and colony formation assay showed that cell viability and colony area of HepG2 cells with AC145343.1 silencing were significantly reduced, which revealed that AC145343.1 silencing contributed to inhibit the proliferative of HepG2 cells. Further, transwell assays and wound healing array showed that the migration ability of HepG2 cells with AC145343.1 silencing were significantly suppressed, which confirmed that AC145343.1 silencing repressed the migration and invasion of HepG2 cells. All in vitro results confirmed that AC145343.1 tended to be a critical risky factor for the survival of HCC patients.

We have provided preliminary evidence for evaluating the relationship between GILncSig and the prognosis of HCC. GILncSig may be of great significance in predicting the degree of genome instability and prognosis of HCC patients. But it still has certain limitations for clinical purposes. Although the prognostic value and independence of lncRNA AC145343.1, ZFPM2-AS1 and lncRNA AC004862.1 on HCC have been verify in the training set, testing set and TCGA set and we proved that AC145343.1 is a high risk factor for HCC *in vitro,* more data sets, *in vivo* experiments, *in vitro* experiments and clinical experiments are still necessary to verify the accuracy, repeatability and the mechanism in regulating genome instability of GILncSig in the future.

MATERIALS AND METHODS

Data collection

Clinical characteristics, RNA sequencing (RNA-Seq) data and somatic mutation variation information of HCC patients were obtained from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/). 377 samples with RNA expression profiles, survival information and common clinical characteristics were extracted for next analysis. Clinical and pathological characteristics of all samples were shown in Supplementary Data 1. Somatic mutation data of 364 HCC patients were also downloaded from TCGA. After integration of 377 samples with their somatic mutation data, 343 patients remained. These patients were randomly separated into two groups according to previous study, named training set and testing set respectively [63]. The training set with 172 HCC patients was used to recognize clinical outcome-related lncRNA signature and establish prognostic risk model. The testing set with 171 HCC patients was used to evaluate the performance of prognostic risk model of training group.

Genome instability-associated lncRNAs analysis

Somatic acquired genomic instability is one of the hallmarks of malignancy cancer [64]. Aberrant lncRNA levels are contributed to abnormal mutation and expression of genes involved in both tumor initiation and progression [65]. To recognize genome instability-associated lncRNAs (GILncRNAs), we constructed a computational frameworks integrating somatic mutator information and lncRNA expression profiles of tumor genome (Figure 8): (i) total cumulative somatic mutations of each sample was calculated; (ii) samples were ranked based on cumulative number of somatic mutations from high to low; (iii) The top 25% samples in the ranking list were set as genomic unstable (GU)-like group while the last 25% were set as genomically stable (GS)-like group; (iv) lncRNA expression profiles between the GU group and GS group were filtered; (v)
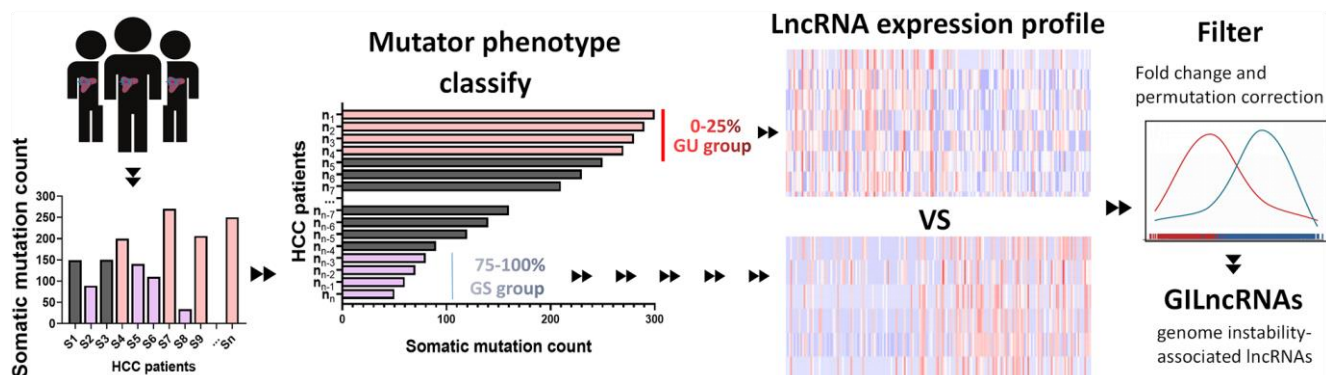

Figure 8. Computational overview of genomic instability-related lncRNAs. Somatic mutations of each HCC sample were counted. Samples were divided into two groups, GU group (patients' mutator phenotype ranked in the top 25%) and GS group (patients' mutator phenotype ranked in the last 25%). Genomic instability-related lncRNAs were examined according to the difference of lncRNA expression profile between GU group and GS group.

GILncRNAs were identified according to the criteria: |fold change| > 1.5 and false discovery rate (FDR) adjusted $P<0.05$.

Statistical analysis

To identify GU group and GS group using profiles of differentially expressed lncRNAs, hierarchical cluster analyses were carried out based on Euclidean distances and Ward's linkage method. Univariate and multivariate Cox proportional hazard regression analysis was performed to evaluate the prognostic value of the expression level of GILncRNAs. To obtain outcome prediction model using GILncRNAs, we combined the regression coefficients from the multivariate regression analysis with the expression data of identified prognostic lncRNAs and constructed the following equation of genome instability-derived lncRNA signature (GILncSig) according to previous study [63]:

$$GILncSig\,(sample) = \sum_{i=1}^{n} coef\,(lncRNA_i)*expr\,(lncRNA_i)$$

GILncSig (sample) in the formula is a risk score of the prognosis of HCC patient. lncRNAi indicates the ith prognostic lncRNA and expr (lncRNAi) represents the expression level of lncRNAi of HCC patient. coef (lncRNAi) is the contribution index of lncRNAi to prognostic risk score which was acquired from the regression coefficient of multivariate Cox analysis. Patients were further equally divided into high-risk group and low-risk group according the risk score. High-risk group of unfavorable prognosis with high GILncSig and low-risk group with low GILncSig were obtained to verified the performance of outcome prediction model. To demonstrate the survival rate and median survival of different group, we performed the Kaplan–Meier analysis and $P<0.05$ was considered to be meaningful. Independence validation among GILncSig and other important clinical factors were evaluated using multivariate Cox regression and stratified analysis. Hazard ratio (HR) and 95% confidence interval (CI) were obtained through Cox analysis. R-version 3.5.2 was used to compute ROC curves and ROC AUC to validate the performance of the constructed GILncSig.

Functional enrichment analysis

To unreal the co-expressed lncRNA-mRNA pairs, Pearson Correlation Coefficients were calculated based on the expression profile between every differentially expressed lncRNA and mRNA. The top 10 items were defined as significant co-expressed mRNAs of lncRNAs. To disclose the biological functional properties associated with the proposed GILncRNAs, functional

enrichment analysis of significant co-expressed mRNAs including Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was applied. The enrichment analysis was performed using clusterProfiler software in R-version 3.5.2 [63].

Cell proliferation and colony formation assays

Human liver cancer cell lines HepG2 were purchased from the Cell Bank of the Type Culture Collection of the Chinese Academy of Sciences, Shanghai Institute of Biochemistry and Cell Biology. Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin and 1% streptomycin (Gibco Life Technologies, Lofer, Austria), then incubated in a humidified incubator with settled parameters (37° C, 5% CO_2). AC145343.1 specific siRNA and negative control siRNA were obtained from Vigene Biosciences and transfected into HepG2 using X-tremegene siRNA transfection reagent (Roche Diagnostics, Shanghai, China) according to the standard guidelines.

Cell proliferation ability of HepG2 with/without AC145343.1 downregulation was evaluated by the Cell Counting KIT-8 (CCK-8, KeyGEN BioTECH, Nanjing, China) according to manufacturer's guidelines. Briefly, 3×10^3 cells suspended were seeded in each well of 96-well plates and incubated overnight for cell attachment. The CCK-8 cell proliferation reagent (10 μl) was added to each well at 24 h, 48 h and 72 h. 4 h after CCK-8 administration, cell proliferation ability is detected. For colony formation assay, 1×10^3 cells suspended were seeded in each well of 6-well plate and incubated for 2 weeks. Cell colonies were fixed with 4% formaldehyde solution and stained with crystal violet for image visualization.

Wound healing assay and transwell assay

Migration ability of HepG2 with/without AC145343.1 downregulation was observed by wound healing assay. 5×10^5 cells suspended were seeded into a 6-well plate. When the cells permeated 90% of the plate, a 'wound' in cell monolayers was scratched using a 1 ml pipette tip. Then images of wounds were captured by an inverted microscope at 0 h, 24 h, 48 h and 72 h. Migration ability was calculated by calculating the wound confluence parameter. Invasive ability of cells was further accessed by transwell assay. The transwell chambers were placed in a 24-well plate which contained complete cultured medium (10% FBS), then added a layer of Matrigel before cell seeding. 8×10^4 cells were seeded into upper transwell chambers and inserted into 300 μl serum-free medium. The cells of the upper surface were removed using a cotton swab 24 h

later. Then the cells on the bottom surface were fixed with 4% formaldehyde solution and stained with 0.5% hematoxylin solution for 20 min. Images of invaded cells were collected by an inverted microscope.

Availability of data and materials

The datasets used during the current study are available from TCGA database (https://cancergenome.nih.gov/), and Supplementary Materials.

AUTHOR CONTRIBUTIONS

GDT, YJQ conceived and designed the study. DPH, JPL and JJT collected and processed data. DPH, JJT and MML analyzed data and prepared tables and figures. YHZ, DTP and WQY conducted the experiments and drafted the manuscript. YJQ and GDT revised the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

Editorial note

&This corresponding author has a verified history of publications using a personal email address for correspondence.

REFERENCES

1. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, Bonaventure A, Valkov M, Johnson CJ, Estève J, Ogunbiyi OJ, Azevedo E Silva G, Chen WQ, et al, and CONCORD Working Group. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet. 2018; 391:1023–75.
   https://doi.org/10.1016/S0140-6736(17)33326-3
   PMID:29395269

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68:394–424.
   https://doi.org/10.3322/caac.21492 PMID:30207593

3. Qian YY, Liu ZS, Zhang Z, Levenson AS, Li K. Pterostilbene increases PTEN expression through the targeted downregulation of microRNA-19a in hepatocellular carcinoma. Mol Med Rep. 2018; 17:5193–201.
   https://doi.org/10.3892/mmr.2018.8515
   PMID:29393488

4. Wei L, Lee D, Law CT, Zhang MS, Shen J, Chin DW, Zhang A, Tsang FH, Wong CL, Ng IO, Wong CC, Wong CM. Genome-wide CRISPR/Cas9 library screening identified PHGDH as a critical driver for Sorafenib resistance in HCC. Nat Commun. 2019; 10:4681.
   https://doi.org/10.1038/s41467-019-12606-7
   PMID:31615983

5. Tzartzeva K, Obi J, Rich NE, Parikh ND, Marrero JA, Yopp A, Waljee AK, Singal AG. Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta-analysis. Gastroenterology. 2018; 154:1706–18.e1.
   https://doi.org/10.1053/j.gastro.2018.01.064
   PMID:29425931

6. Gillespie JR, Uversky VN. Structure and function of alpha-fetoprotein: a biophysical overview. Biochim Biophys Acta. 2000; 1480:41–56.
   https://doi.org/10.1016/S0167-4838(00)00104-7
   PMID:11004554

7. Mizejewski GJ. Biological role of alpha-fetoprotein in cancer: prospects for anticancer therapy. Expert Rev Anticancer Ther. 2002; 2:709–35.
   https://doi.org/10.1586/14737140.2.6.709
   PMID:12503217

8. Park SJ, Jang JY, Jeong SW, Cho YK, Lee SH, Kim SG, Cha SW, Kim YS, Cho YD, Kim HS, Kim BS, Park S, Bang HI. Usefulness of AFP, AFP-L3, and PIVKA-II, and their combinations in diagnosing hepatocellular carcinoma. Medicine (Baltimore). 2017; 96:e5811.
   https://doi.org/10.1097/MD.0000000000005811
   PMID:28296720

9. Gao J, Xu AF, Zheng HY, Chen G, Wang MC, Lou GQ. [Clinical application studies on AFP-L3 detected by micro-spin column method]. Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi. 2010; 24:461–63.
   PMID:21604577

10. Taketa K, Endo Y, Sekiya C, Tanikawa K, Koji T, Taga H, Satomura S, Matsuura S, Kawai T, Hirai H. A collaborative study for the evaluation of lectin-reactive alpha-fetoproteins in early detection of hepatocellular carcinoma. Cancer Res. 1993; 53:5419–23.
   PMID:7693340

11. Aoyagi Y, Isemura M, Yosizawa Z, Suzuki Y, Sekine C, Ono T, Ichida F. Fucosylation of serum alpha-fetoprotein in patients with primary hepatocellular carcinoma. Biochim Biophys Acta. 1985; 830:217–23.
https://doi.org/10.1016/0167-4838(85)90277-8
PMID:2411292

12. Sato Y, Nakata K, Kato Y, Shima M, Ishii N, Koji T, Taketa K, Endo Y, Nagataki S. Early recognition of hepatocellular carcinoma based on altered profiles of alpha-fetoprotein. N Engl J Med. 1993; 328:1802–806.
https://doi.org/10.1056/NEJM199306243282502
PMID:7684823

13. Luo P, Wu S, Yu Y, Ming X, Li S, Zuo X, Tu J. Current Status and Perspective Biomarkers in AFP Negative HCC: Towards Screening for and Diagnosing Hepatocellular Carcinoma at an Earlier Stage. Pathol Oncol Res. 2020; 26:599–603.
https://doi.org/10.1007/s12253-019-00585-5
PMID:30661224

14. Andor N, Maley CC, Ji HP. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. Cancer Res. 2017; 77:2179–85.
https://doi.org/10.1158/0008-5472.CAN-16-1553
PMID:28432052

15. Tubbs A, Nussenzweig A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell. 2017; 168:644–56.
https://doi.org/10.1016/j.cell.2017.01.002
PMID:28187286

16. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013; 153:17–37.
https://doi.org/10.1016/j.cell.2013.03.002
PMID:23540688

17. Ottini L, Falchetti M, Lupi R, Rizzolo P, Agnese V, Colucci G, Bazan V, Russo A. Patterns of genomic instability in gastric cancer: clinical implications and perspectives. Ann Oncol. 2006 (Suppl 7); 17:vii97–102.
https://doi.org/10.1093/annonc/mdl960
PMID:16760303

18. Munschauer M, Nguyen CT, Sirokman K, Hartigan CR, Hogstrom L, Engreitz JM, Ulirsch JC, Fulco CP, Subramanian V, Chen J, Schenone M, Guttman M, Carr SA, Lander ES. Publisher Correction: the NORAD lncRNA assembles a topoisomerase complex critical for genome stability. Nature. 2018; 563:E32.
https://doi.org/10.1038/s41586-018-0584-2
PMID:30279576

19. Aguilera A, García-Muse T. Causes of genome instability. Annu Rev Genet. 2013; 47:1–32.
https://doi.org/10.1146/annurev-genet-111212-133232 PMID:23909437

20. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature. 2011; 470:284–88.
https://doi.org/10.1038/nature09701
PMID:21307942

21. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2016; 17:47–62.
https://doi.org/10.1038/nrg.2015.10 PMID:26666209

22. Huarte M, Rinn JL. Large non-coding RNAs: missing links in cancer? Hum Mol Genet. 2010; 19:R152–61.
https://doi.org/10.1093/hmg/ddq353 PMID:20729297

23. Zhang F, Ding L, Cui L, Barber R, Deng B. Identification of long non-coding RNA-related and -coexpressed mRNA biomarkers for hepatocellular carcinoma. BMC Med Genomics. 2019 (Suppl 1); 12:25.
https://doi.org/10.1186/s12920-019-0472-0
PMID:30704465

24. Zeng YL, Guo ZY, Su HZ, Zhong FD, Jiang KQ, Yuan GD. Diagnostic and prognostic value of lncRNA cancer susceptibility candidate 9 in hepatocellular carcinoma. World J Gastroenterol. 2019; 25:6902–15.
https://doi.org/10.3748/wjg.v25.i48.6902
PMID:31908394

25. Zhang L, Yang F, Yuan JH, Yuan SX, Zhou WP, Huo XS, Xu D, Bi HS, Wang F, Sun SH. Epigenetic activation of the MiR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma. Carcinogenesis. 2013; 34:577–86.
https://doi.org/10.1093/carcin/bgs381
PMID:23222811

26. Yang Z, Zhou L, Wu LM, Lai MC, Xie HY, Zhang F, Zheng SS. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol. 2011; 18:1243–50.
https://doi.org/10.1245/s10434-011-1581-y
PMID:21327457

27. Tam AS, Sihota TS, Milbury KL, Zhang A, Mathew V, Stirling PC. Selective defects in gene expression control genome instability in yeast splicing mutants. Mol Biol Cell. 2019; 30:191–200.
https://doi.org/10.1091/mbc.E18-07-0439
PMID:30462576

28. Li F, Bai L, Li S, Chen Y, Xue X, Yu Z. Construction and evaluation of a prognosis lncRNA model for hepatocellular carcinoma. J Cell Biochem. 2021; 122:983–92.
https://doi.org/10.1002/jcb.29608 PMID:32348599

29. Ma L, Deng C. Identification of a novel four-lncRNA signature as a prognostic indicator in cirrhotic hepatocellular carcinoma. PeerJ. 2019; 7:e7413.
https://doi.org/10.7717/peerj.7413

PMID:31396449

30. Watanabe T, Nanamiya H, Kojima M, Nomura S, Furukawa S, Soeda S, Tanaka D, Isogai T, Imai JI, Watanabe S, Fujimori K. Clinical implication of oncogenic somatic mutations in early-stage cervical cancer with radical hysterectomy. Sci Rep. 2020; 10:18734.
https://doi.org/10.1038/s41598-020-72518-1
PMID:33127935

31. Engeland K. Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. Cell Death Differ. 2018; 25:114–32.
https://doi.org/10.1038/cdd.2017.172 PMID:29125603

32. Vousden KH, Prives C. Blinded by the Light: The Growing Complexity of p53. Cell. 2009; 137:413–31.
https://doi.org/10.1016/j.cell.2009.04.037
PMID:19410540

33. Gasco M, Shami S, Crook T. The p53 pathway in breast cancer. Breast Cancer Res. 2002; 4:70–76.
https://doi.org/10.1186/bcr426 PMID:11879567

34. Pharoah PD, Day NE, Caldas C. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. Br J Cancer. 1999; 80:1968–73.
https://doi.org/10.1038/sj.bjc.6690628
PMID:10471047

35. Huszno J, Grzybowska E. *TP53* mutations and SNPs as prognostic and predictive factors in patients with breast cancer. Oncol Lett. 2018; 16:34–40.
https://doi.org/10.3892/ol.2018.8627 PMID:29928384

36. Børresen-Dale AL. TP53 and breast cancer. Hum Mutat. 2003; 21:292–300.
https://doi.org/10.1002/humu.10174 PMID:12619115

37. Walerych D, Napoli M, Collavin L, Del Sal G. The rebel angel: mutant p53 as the driving oncogene in breast cancer. Carcinogenesis. 2012; 33:2007–17.
https://doi.org/10.1093/carcin/bgs232
PMID:22822097

38. Huang DP, Zeng YH, Yuan WQ, Huang XF, Chen SQ, Wang MY, Qiu YJ, Tong GD. Bioinformatics Analyses of Potential miRNA-mRNA Regulatory Axis in HBV-related Hepatocellular Carcinoma. Int J Med Sci. 2021; 18:335–46.
https://doi.org/10.7150/ijms.50126 PMID:33390802

39. Lin D, Yang HI, Nguyen N, Hoang J, Kim Y, Vu V, Le A, Chaung K, Nguyen V, Trinh H, Li J, Zhang J, Hsing A, et al. Reduction of chronic hepatitis B-related hepatocellular carcinoma with anti-viral therapy, including low risk patients. Aliment Pharmacol Ther. 2016; 44:846–55.
https://doi.org/10.1111/apt.13774 PMID:27549411

40. Chen DL, Chen LZ, Lu YX, Zhang DS, Zeng ZL, Pan ZZ,

Huang P, Wang FH, Li YH, Ju HQ, Xu RH. Long noncoding RNA XIST expedites metastasis and modulates epithelial-mesenchymal transition in colorectal cancer. Cell Death Dis. 2017; 8:e3011.
https://doi.org/10.1038/cddis.2017.421
PMID:28837144

41. Chu H, Xia L, Qiu X, Gu D, Zhu L, Jin J, Hui G, Hua Q, Du M, Tong N, Chen J, Zhang Z, Wang M. Genetic variants in noncoding PIWI-interacting RNA and colorectal cancer risk. Cancer. 2015; 121:2044–52.
https://doi.org/10.1002/cncr.29314 PMID:25740697

42. Arun G, Spector DL. MALAT1 long non-coding RNA and breast cancer. RNA Biol. 2019; 16:860–63.
https://doi.org/10.1080/15476286.2019.1592072
PMID:30874469

43. Jin Y, Feng SJ, Qiu S, Shao N, Zheng JH. LncRNA MALAT1 promotes proliferation and metastasis in epithelial ovarian cancer via the PI3K-AKT pathway. Eur Rev Med Pharmacol Sci. 2017; 21:3176–84.
PMID:28770968

44. Hu Z, Mi S, Zhao T, Peng C, Peng Y, Chen L, Zhu W, Yao Y, Song Q, Li X, Li X, Jia C, Pei H. BGL3 lncRNA mediates retention of the BRCA1/BARD1 complex at DNA damage sites. EMBO J. 2020; 39:e104133.
https://doi.org/10.15252/embj.2019104133
PMID:32347575

45. Yousefzadeh M, Henpita C, Vyas R, Soto-Palma C, Robbins P, Niedernhofer L. DNA damage-how and why we age? eLife. 2021; 10:10.
https://doi.org/10.7554/eLife.62852
PMID:33512317

46. Wenzel ES, Singh AT. Cell-cycle Checkpoints and Aneuploidy on the Path to Cancer. In Vivo. 2018; 32:1–5.
https://doi.org/10.21873/invivo.11197
PMID:29275292

47. Zhao S, Kieser A, Li HY, Reinking HK, Weickert P, Euteneuer S, Yaneva D, Acampora AC, Götz MJ, Feederle R, Stingele J. A ubiquitin switch controls autocatalytic inactivation of the DNA-protein crosslink repair protease SPRTN. Nucleic Acids Res. 2021; 49:902–15.
https://doi.org/10.1093/nar/gkaa1224
PMID:33348378

48. Wang Y, Li G, Deng M, Liu X, Huang W, Zhang Y, Liu M, Chen Y. The multifaceted functions of RNA helicases in the adaptive cellular response to hypoxia: from mechanisms to therapeutics. Pharmacol Ther. 2021; 221:107783.
https://doi.org/10.1016/j.pharmthera.2020.107783
PMID:33307143

49. Zhang XW, Li QH, Xu ZD, Dou JJ. STAT1-induced

regulation of lncRNA ZFPM2-AS1 predicts poor prognosis and contributes to hepatocellular carcinoma progression via the miR-653/GOLM1 axis. Cell Death Dis. 2021; 12:31.
https://doi.org/10.1038/s41419-020-03300-4
PMID:33414427

50. He H, Wang Y, Ye P, Yi D, Cheng Y, Tang H, Zhu Z, Wang X, Jin S. Long noncoding RNA ZFPM2-AS1 acts as a miRNA sponge and promotes cell invasion through regulation of miR-139/GDF10 in hepatocellular carcinoma. J Exp Clin Cancer Res. 2020; 39:159.
https://doi.org/10.1186/s13046-020-01664-1
PMID:32795316

51. Kanagalingam T, Solomon L, Vijeyakumaran M, Palikhe NS, Vliagoftis H, Cameron L. IL-2 modulates Th2 cell responses to glucocorticosteroid: A cause of persistent type 2 inflammation? Immun Inflamm Dis. 2019; 7:112–24.
https://doi.org/10.1002/iid3.249 PMID:30994266

52. Muller S, Mazen A, Martinage A, Van Regenmortel MH. Use of histone antibodies for studying chromatin topography and the phosphorylation of chromatin subunits. EMBO J. 1984; 3:2431–36.
https://doi.org/10.1002/j.1460-2075.1984.tb02150.x
PMID:6499836

53. Shen H, Zhu M, Wang C. Precision oncology of lung cancer: genetic and genomic differences in Chinese population. NPJ Precis Oncol. 2019; 3:14.
https://doi.org/10.1038/s41698-019-0086-1
PMID:31069257

54. Fradet Y. Biomarkers in prostate cancer diagnosis and prognosis: beyond prostate-specific antigen. Curr Opin Urol. 2009; 19:243–46.
https://doi.org/10.1097/MOU.0b013e32832a08b5
PMID:19325493

55. Riegman PH, Vissers KJ, Alers JC, Geelen E, Hop WC, Tilanus HW, van Dekken H. Genomic alterations in malignant transformation of Barrett's esophagus. Cancer Res. 2001; 61:3164–70.
PMID:11306503

56. van Dekken H, Alers JC, Riegman PH, Rosenberg C, Tilanus HW, Vissers K. Molecular cytogenetic evaluation of gastric cardia adenocarcinoma and precursor lesions. Am J Pathol. 2001; 158:1961–67.
https://doi.org/10.1016/S0002-9440(10)64666-4
PMID:11395372

57. Konetzke GW. [Criteria and references for the recognition of malignant tumors as occupational disease]. Z Arztl Fortbild. 1985; 79:89–90.
PMID:3993091

58. Park NH, Chung YH, Youn KH, Song BC, Yang SH, Kim JA, Lee HC, Yu E, Lee YS, Lee SG, Kim KW, Suh DJ. Close correlation of p53 mutation to microvascular invasion in hepatocellular carcinoma. J Clin Gastroenterol. 2001; 33:397–401.
https://doi.org/10.1097/00004836-200111000-00011
PMID:11606857

59. Terris B, Laurent-Puig P, Belghitti J, Degott C, Hénin D, Fléjou JF. Prognostic influence of clinicopathologic features, DNA-ploidy, CD44H and p53 expression in a large series of resected hepatocellular carcinoma in France. Int J Cancer. 1997; 74:614–19.
https://doi.org/10.1002/(SICI)1097-0215(19971219)74:6<614::AID-IJC10>3.0.CO;2-5
PMID:9421358

60. Yuan RH, Jeng YM, Chen HL, Lai PL, Pan HW, Hsieh FJ, Lin CY, Lee PH, Hsu HC. Stathmin overexpression cooperates with p53 mutation and osteopontin overexpression, and is associated with tumour progression, early recurrence, and poor prognosis in hepatocellular carcinoma. J Pathol. 2006; 209:549–58.
https://doi.org/10.1002/path.2011 PMID:16739096

61. Atta MM, el-Masry SA, Abdel-Hameed M, Baiomy HA, Ramadan NE. Value of serum anti-p53 antibodies as a prognostic factor in Egyptian patients with hepatocellular carcinoma. Clin Biochem. 2008; 41:1131–39.
https://doi.org/10.1016/j.clinbiochem.2008.06.006
PMID:18616935

62. Liu J, Ma Q, Zhang M, Wang X, Zhang D, Li W, Wang F, Wu E. Alterations of TP53 are associated with a poor outcome for patients with hepatocellular carcinoma: evidence from a systematic review and meta-analysis. Eur J Cancer. 2012; 48:2328–38.
https://doi.org/10.1016/j.ejca.2012.03.001
PMID:22459764

63. Bao S, Zhao H, Yuan J, Fan D, Zhang Z, Su J, Zhou M. Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. Brief Bioinform. 2020; 21:1742–55.
https://doi.org/10.1093/bib/bbz118
PMID:31665214

64. Ganmore I, Smooha G, Izraeli S. Constitutional aneuploidy and cancer predisposition. Hum Mol Genet. 2009; 18:R84–93.
https://doi.org/10.1093/hmg/ddp084 PMID:19297405

65. Shi X, Sun M, Liu H, Yao Y, Song Y. Long non-coding RNAs: a new frontier in the study of human diseases. Cancer Lett. 2013; 339:159–66.
https://doi.org/10.1016/j.canlet.2013.06.013
PMID:23791884

SUPPLEMENTARY MATERIALS

Supplementary Tables

Supplementary Table 1. The common clinical features characteristic between the training set and testing set.

Variables	Type	Total	Test	Train	Pvalue
Age	<=65	216(62.97%)	107(62.57%)	109(63.37%)	0.967
	>65	127(37.03%)	64(37.43%)	63(36.63%)	
Gender	FEMALE	110(32.07%)	63(36.84%)	47(27.33%)	0.0763
	MALE	233(67.93%)	108(63.16%)	125(72.67%)	
Grade	G1-2	214(62.39%)	106(61.99%)	108(62.79%)	1
	G3-4	124(36.15%)	61(35.67%)	63(36.63%)	
Stage	Stage I-II	238(69.39%)	115(67.25%)	123(71.51%)	0.9667
	Stage III-IV	83(24.2%)	41(23.98%)	42(24.42%)	
T	T1-2	252(73.47%)	126(73.68%)	126(73.26%)	0.9524
	T3-4	88(25.66%)	43(25.15%)	45(26.16%)	
M	M0	245(71.43%)	113(66.08%)	132(76.74%)	0.8991
	M1	3(0.87%)	2(1.17%)	1(0.58%)	
N	N0	239(69.68%)	111(64.91%)	128(74.42%)	0.9081
	N1-3	3(0.87%)	2(1.17%)	1(0.58%)	

Supplementary Table 2. Univariable Cox regression analysis.

Gene symbol	HR	95% CI	P-value
CASC9	1.068	1.012-1.126	0.017
MIR210HG	1.157	1.062-1.261	9.90E-04
LUCAT1	1.174	1.078-1.278	2.33E-04
PRRT3-AS1	1.091	1.028-1.156	0.037
KCNMB2-AS1	1.235	1.082-1.409	1.72E-03
AC145343.1	1.378	1.072-1.772	0.012
AL731684.1	1.141	1.037-1.255	0.007
AC004862.1	0.811	0.683-0.964	0.017
ZFPM2-AS1	1.148	1.095-1.204	1.14E-08

Supplementary Data

Please browse Full Text version to see the data of Supplementary Data 1, 2.

Supplementary Data 1. Clinical and pathological characteristics of all samples.

Supplementary Data 2. 88 GILncRNAs.