

88581_Auto_Edited.docx

Name of Journal: *World Journal of Gastrointestinal Surgery*

Manuscript NO: 88581

Manuscript Type: ORIGINAL ARTICLE

Clinical and Translational Research

Construction and validation of somatic mutation-derived LncRNA signatures of genomic instability to predict prognosis of hepatocellular carcinoma

Construction and validation of genomic instability characteristics of somatic mutation LncRNA in the prognosis of hepatocellular carcinoma

Botao Duan, xuekai zhao, Yangyang Cui, Dezheng Liu, lin wang, Lei Zhou, XingYuan Zhang

Abstract

BACKGROUND

Long non-coding RNAs (LncRNAs) have been found to be a potential prognostic factor for cancers, including hepatocellular carcinoma (HCC). Some LncRNAs have been confirmed as potential indicators to quantify genomic instability (GI). Nevertheless, GI-LncRNAs remains largely unexplored. This study established a genomic instability-derived LncRNA signature (GILncSig) that can predict the prognosis of HCC patients.

AIM

Establish a genomic instability-derived LncRNA signature (GILncSig) that can predict the prognosis of HCC patients.

METHODS

Identifying the GI-LncRNAs by combining LncRNA expression and somatic mutation profiles. Next, GI-LncRNAs were analyzed for functional enrichment. GILncSig was established in the training set by Cox regression analysis, and its predictive ability was verified in testing set and TCGA set. In addition, we explored the effects of GILncSig and TP53 on the prognosis.

RESULTS

A total of 88 GI-LncRNAs were found, and functional enrichment analysis showed that their functions were mainly involved in small molecule metabolism and GI. GILncSig was constructed by 5 LncRNAs (MIR210HG, AC016735.1, AC116351.1, AC010643.1, LUCAT1). In the training set, the prognosis of high-risk patients was significantly worse than that of low-risk patients, and similar results were verified in the testing set and TCGA set. Multivariate Cox regression analysis and stratified analysis confirmed that GILncSig could be used as an independent prognostic factor. The ROC curve analysis of GILncSig showed that its AUC (0.773) was higher than the two LncRNA signatures

published recently. Furthermore, GILncSig may have a better predictive performance than TP53 mutation status alone.

CONCLUSION

We established a GILncSig that can predict the prognosis of HCC patients, which can help to guide the prognostic evaluation and treatment decisions.

Key Words: genomic instability, long noncoding RNA, hepatocellular carcinoma, prognosis, diagnose

Duan B, zhao X, Cui Y, Liu D, wang L, Zhou L, Zhang X. Construction and validation of somatic mutation-derived LncRNA signatures of genomic instability to predict prognosis of hepatocellular carcinoma. *World J Gastrointest Surg* 2023; In press

Core Tip: Identifying the GI-LncRNAs by combining LncRNA expression and somatic mutation profiles. Next, GI-LncRNAs were analyzed for functional enrichment. GILncSig was established in the training set by Cox regression analysis, and its predictive ability was verified in testing set and TCGA set. A total of 88 GI-LncRNAs were found, and functional enrichment analysis showed that their functions were mainly involved in small molecule metabolism and GI. We established a GILncSig that can predict the prognosis of HCC patients, which can help to guide the prognostic evaluation and treatment decisions.

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the cancers with the highest mortality among all malignant tumor, ranking sixth among common cancers [1, 2] In the world, the mortality rate of patients with HCC ranks second among the total mortality of all cancers. The incidence rate and mortality of liver cancer in East Asia, Southeast Asia, Africa and southern Europe are particularly obvious.[3] The incidence rate of HCC is

increasing year by year, which is, of course, related to the improvement of diagnostic mode and the shortening of cancer monitoring interval. In the past, viral hepatitis was the main epidemiological cause of hepatocellular carcinoma, but with the implementation of hepatitis B vaccination and hepatitis C treatment plan worldwide, the annual incidence rate of hepatocellular carcinoma with viral hepatitis as the main cause decreased. And increasing evidence suggests that ²¹ non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH) contribute to the development of HCC and are becoming increasingly common causes of HCC worldwide. With the implementation of viral hepatitis treatment plans, the epidemiological etiology of HCC is likely to shift from viral hepatitis to non-alcoholic steatohepatitis. ^[4-6] As we all know, ¹³ HCC refers to a fairly complex disease. The current prognostic factors for HCC include tumor size, number, vascular invasion, extrahepatic spread, severity of underlying liver disease as defined by bilirubin and portal hypertension, as well as corresponding qualified treatment modalities.^[7] Traditional surgical treatment and locoregional therapies have obvious efficacy for some HCC patients, but some patients still have the possibility of long-term recurrence, with poor prognosis and high mortality.^[8] In systematic treatment regimens, advanced HCC patients can generally be treated with tyrosine kinase inhibitors (TKIs). With the increasing understanding and characterization of the immune characteristics of the tumor microenvironment, immune checkpoint inhibitors (ICI) methods further expand the systemic treatment of HCC. The current emerging systemic comprehensive treatment method combines the above two methods, and there is evidence that the combination therapy of ICI+TKI has achieved certain results. However, existing research evidence suggests that the treatment options currently used in clinical practice are still relatively ineffective. In fact, although the efficacy has significantly improved after the introduction of ICI, the objective response rate to treatment is still largely insufficient. Most patients do not have good returns, and the 5-year OS of metastatic HCC is still unsatisfactory. Currently, efforts should mainly focus on expanding treatment targets and searching for reliable biomarkers as much as possible, which will help adjust treatment choices and avoid the risks and costs

associated with drug ineffectiveness and side effects.^[9] Therefore, new biomarkers are eagerly needed to predict the prognosis of HCC patients.

Genomic instability (GI) has been verified to be one of the characteristics of malignant tumor.^[10] Chromosomal instability (CIN) and microsatellite instability (MIN) are two major types of genomic instability, and more importantly, they are significantly associated with the prognosis of cancer patients.^[11] The underlying mechanism of it may be related to oxidative stress response and the joint defect of DNA damage checkpoint and repair pathway.^[12] It also proves that molecular markers have great potential in quantifying genomic instability. For example, mettu *et al*^[13] analyzed demonstrates that their identified 12-gene genomic instability signature could predict disease outcomes in multiple cancer types with epithelial origins. A mutation-derived Gene signatures of genomic instability that can help in predicting the OS of patients with HCC constructed by Song *et al*^[14] Therefore, these Genomic Instability Signatures potential may be a potential new therapeutic direction for HCC patients.

Long non-coding RNAs (LncRNAs) are a non-protein coding transcript with length greater than or equal to 200 nucleotides.^[15] Now, more and more evidence suggests that LncRNA is becoming a potential regulator for GI and to some extent quantifying the level of genome instability. ^[16, 17] For example, some studies have found that a discovered NORAD or LINC00657 regulates genomic stability by isolating pumilio proteins.^[18] LncRNA dysfunction is closely associated with the occurrence of various tumors, including HCC.^[19] Li *et al* found that LncRNA Ftx overexpression promoted the proliferation, invasion and migration of HCC cells ^[20] Although a considerable number of LncRNAs have been discovered to be related to genomic stability, the clinical application of other GI-related LncRNAs in cancer has largely not been explored. At the same time, it has great potential as a new prognostic biomarker.

Therefore, in our study, we attempted to establish a genomic instability-derived LncRNA signature (GILncSig) that could help predict the prognosis of HCC patients by combining LncRNA expression profile with somatic mutation profile.

MATERIALS AND METHODS

Data sources

The data involved in our research mainly included Clinical characteristics, somatic mutation information, and transcriptome expression data of hepatocellular carcinoma which were extracted from TCGA portal. (<https://portal.gdc.cancer.gov/>). A total of 424 files with mRNA and LncRNA profiles (including 50 normal and 374 tumor tissues), 377 clinical characteristics of HCC and 372 patients with somatic mutation information were obtained. All HCC patients ($n = 343$) were randomly divided into the training set and the testing set (chi-square test showed that there was no statistical difference between the training set and the testing set) for further construction and verification of LncRNA signature.

Identification of genome instability-associated LncRNAs (GI-LncRNAs)

In order to identify GI-LncRNAs, we first calculated the cumulative number of somatic mutations for each patient in HCC samples by combining LncRNA expression profile and somatic mutation profile, and arranged them from large to small. The patients in the top quarter are referred to as genomically unstable (GU) samples, and the patients in the bottom quarter are genomically stable (GS) samples. The differentially expressed LncRNAs (absolute value of fold change was greater than 1, and the adjusted p value of FDR was less than 0.05) between the two groups were defined as genome instability-associated LncRNAs.

Hierarchical cluster analysis was performed on all samples, and differentially expressed LncRNAs were used to identify GU-like group and GS-like group. In order to explore the correlation between GI-LncRNAs and mRNA pairings, the top 10 mRNAs most related to each GI-LncRNAs were screened by Pearson correlation coefficient. On this basis, a co-expression network is established. Subsequently, functional enrichment analysis was performed on the co-expressed LncRNA-associated mRNAs to reveal the potential biological characteristics of GI-LncRNA, including Gene Ontology (GO) terms

and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. ClusterProfiler software in R-version 4.0.2^[21] was used for functional enrichment analysis.

Establishment of the genome instability- derived LncRNA signature (GILncSig)

In the training set, GILncSig formula with risk score was established based on the results of multivariate Cox regression analysis and the expression level of GI-LncRNA. which formula was as follows: $GILncSig(patient) = \sum (expression\ of\ LncRNA_n * coef(LncRNA_n))$, where GILncSig (patient) is a prognostic risk score for the HCC patient, and the LncRNA_n represents the nth of independent prognostic LncRNAs. The coef (LncRNA_n) represents the contribution index of LncRNA_i to prognostic risk score from the Cox regression analyses. ^[21] In the training set, patients' median risk score was used as a dividing line between patients in the high-risk group (high GILncSig) and those in the low-risk group (low GILncSig). The prediction ability of GILncSig was verified by Kaplan-Meier method (P < 0.05 was considered significant), and moreover, the performance was further evaluated by the time-dependent receiver operating characteristic (ROC) curve. All calculations and analyses in this paper are all using R-version 4.0.2.

Validation of the genome instability- derived LncRNA signature (GILncSig)

We first validated the model on a randomly assigned test set and a TCGA set containing all patients. Similar to the training set, we used GILncSig to calculate the risk scores of each patient within the two sets separately and divided them into two groups of high and low risk within the respective sets. The same K-M analysis and ROC curves were used to validate GILncSig between the two groups in each of the two pools. Secondly, we used COX regression analysis to verify whether GILncSig could be distinguished from other clinical features as an independent prognostic factor. We also performed ROC curve analysis of GILncSig with two extant LncRNA signatures predicting HCC prognosis and compared their area under the curve, and then we verified whether GILncSig could be applied to patients with different clinical

characteristics using K-M analysis. In addition, we also analysed the prognostic value of GILncSig in combination with TP53.

31

RESULTS

Identification of genomic instability-related LncRNAs (GI-LncRNAs) in HCC

According to the number of somatic mutations in each patient, we were able to get the GS group ($n = 90$) and Gu group ($n = 93$). After that, differential expression analysis of LncRNA expression profiles of the two groups was conducted, and 88 different LncRNAs with statistical significance were obtained ($|\text{fold change}| > 1$ and false discovery rate (FDR) adjusted $P < 0.05$). Of these, 56 LncRNAs were found to be upregulated and 32 to be downregulated. The heat map (Figure 1A) shows the top 20 LncRNAs with the largest differential expression. Unsupervised hierarchical clustering analysis was performed on all HCC samples based on the expression levels of 88 differentially expressed LncRNAs, and 374 samples were divided into two groups, which were shown in the Figure 1B. The level of somatic mutations in the GU-like group was significantly higher than in the GS-like group. (Figure 1C). In addition, the expression of H2AX was compared between the two groups. In addition, we found that the expression of H2AX in GU-like group was significantly higher than that in GS-like group ($p < 0.01$, Figure 1D). H2AX has been found to promote rapid division of cancer cells and is significantly associated with genomic instability. [22]

Next, we used functional enrichment analysis to predict the potential functions of these GI-LncRNAs. We screened the top 10 protein-coding genes (PCGs) with the strongest correlation with LncRNA. On this basis, an LncRNA-mRNA co-expression network was constructed (Figure 2A). The GO analysis of co-expressed LncRNA-associated mRNAs showed that mRNAs and LncRNA-corrected PCG in the network were significantly enriched in the metabolic process, including small molecule catabolic process and fatty acid metabolic process ($p < 0.05$, Figure 2B). In terms of KEGG pathway analysis, 22 significantly rich pathways were found, including pyrimidine metabolism, purine metabolism, and folate biosynthesis ($p < 0.05$, Figure 2C). The results of functional

enrichment analysis showed that 88 differentially expressed LncRNAs could participate in a variety of cancer-related biological processes by interfering with a variety of metabolic pathways, among them, gene instability could be affected by interfering with gene synthesis.

³ Development of a genomic instability-derived LncRNA signature (GILncSig) outcome prediction in the training set

To further explore the prognostic effects of these GI-LncRNAs, we randomly divided all HCC patients into two groups: training set ($n = 172$) and testing set ($n = 171$). Univariate Cox regression analysis was performed on the samples in the training set to analyse the association between overall survival (OS) and LncRNA expression levels of 88 GI-LncRNAs in the training set. A total of 13 LncRNAs were discovered to be significantly correlated with the prognosis of HCC patients ($P < 0.05$, Figure 3A). Next, multivariate Cox regression analysis was conducted for the 13 LncRNAs. Finally, 5 of the 13 candidates LncRNAs (MIR210HG, AC016735.1, AC116351.1, AC010643.1 and LUCAT1) in multivariate Cox analysis has kept the prognostic significance, is considered to be independent prognostic factors. Finally, on this basis, a genomic instability-derived LncRNA signature (GILncSig) was constructed that could be used to assess the prognostic risk of HCC patients. The formula are as follows: $GILncSig = (0.0867 \times \text{expression level of MIR210HG}) + (0.0454 \times \text{expression level of AC016735.1}) + (0.1316 \times \text{expression level of AC116351.1}) + (0.3036 \times \text{expression level of AC010643.1}) + (0.2557 \times \text{expression level of LUCAT1})$. In GILncSig, the coefficients of 5 LncRNAs were all positive, and their high expression was associated with poor prognosis. It means that these LncRNA tended to be risk factors.

Risk scores were calculated for all patients in the training set by using GILncSig. The risk scores equal to or higher than the median value of the risk score was the high-risk group, and the rest were the low-risk group. Log-rank tests and Kaplan-Meier analysis showed that patients in the low-risk group had significantly better survival outcomes than those in the high-risk group ($P < 0.001$ Figure 3B). The 5-year survival

rates of the two groups were 9.3% (high-risk group) and 19.8% (low-risk group). The ROC curve analysis of GILncSig over time is shown in Figure 3C, and the area under the curve (AUC) is 0.773. At the same time, GILncSig expression level, somatic mutation count and expression level of H2AX, UBQLN4 genes (a newly discovered driver of genomic instability^[23]) were also observed to change with the increase of the risk score (Figure 3D). For patients with high scores, the expression of miR210HG AC016735.1, AC116351.1, AC010643.1 and LUCAT1 were up-regulated. Compared with the low-risk group, somatic mutations were more frequent in the high-risk group ($P = 0.0011$ Figure 3E). In addition, the expression of UBQLN4 and H2AX were higher in high-risk patients than in low-risk patients ($P < 0.01$ Figure 3F).

Independent validation of GILncSig on the RNA-seq platform of HCC data.

Subsequently, in order to examine the credibility of the prognostic performance of GILncSig, we used the independent testing set of 171 patients to test it. Similarly, using GILncSig to calculate the risk score of patients in the testing set, and patients were also divided into high-risk group ($n = 76$) and low-risk group ($n = 95$) according to the same method as in the training set, and the KM analysis also showed significant differences between the two groups. The OS rate in low-risk group was significantly better than that in high-risk group ($P = 0.013$, Figure 4A). The 5-year survival rate in the high-risk group was 3.95%, which was lower than that in the low-risk group (12.63%). In the testing set, the ROC curves analysis of GILncSig over time showed that the AUC was 0.679 (Figure 4B). Similar to the training set, the expression of GILncSig as well as somatic mutation count and the expression of H2AX, UBQLN4 in the testing set were mostly positively correlated with the risk value ($p < 0.01$ Figure 4C). The somatic cell mutation rate of the high-risk group in the testing set was slightly higher than that of the low-risk group ($P = 0.18$ Figure 4D). The expression level of UBQLN4 and H2AX in low-risk group was significantly lower than that in high-risk group (Figure 4E, $P < 0.01$).

Similarly, we divided all patients in the TCGA set into the high-risk group ($n = 162$) and low-risk group ($n = 181$) and used the same method to verify the performance of

GILncSig. As expected, we got similar but more meaningful results. The overall survival rate and 5-year survival rate (6.79% to 16.02%) of patients in high-risk group were lower than those in low-risk group ($P < 0.01$ Figure 4F). The ROC curves analysis of GILncSig in the TCGA set over time showed that the AUC value was 0.730 (Figure 4G). Figure 4H shows the expression of GILncSig, somatic mutation count and the expression of UBQLN4, H2AX in the TCGA set. As expected, the somatic cell mutation rate and the expression levels of UBQLN4 and H2AX in the high-risk group were significantly higher than those in the low-risk group (Figure 4I $P = 0.0011$; $P < 0.01$, respectively).

Comparison of prediction ability of GILncSig with existing LncRNA Signatures

Next, the predictive performance of GILncSig in our study was compared with two published LncRNA signatures for predicting HCC prognosis: 6-LncRNA signature derived from Gu's study (hereinafter referred to as GuLncSig)^[24] and 4-LncRNA signature derived from Wu's study (hereinafter referred to as WuLncSig)^[25] using the same TCGA patient cohort. On this basis, using ROC curve analysis to evaluate the prognostic performance of these signatures. As shown in the Figure 5, the AUC of the GILncSig is 0.736, which is higher than that of GuLncSig (AUC = 0.664) and WuLncSig (AUC=0.725). These results may indicate that GILncSig has better prognostic performance than the two recently published LncRNA signatures.

Independence of the GILncSig from other clinical factors

To verify whether GILncSig can be used as an independent clinical variable to evaluate the prognosis of HCC patients. Multivariate Cox regression analyses were performed for age, sex, grade, stage, and prognostic risk score based on GILncSig. Finally, GILncSig was observed to be significant as an independent prognostic factor ($P < 0.05$ Table 1). To determine whether GILncSig can be applied to different clinical traits, we first divided the TCGA group into groups older than 65 years old ($n = 141$) and younger than or equal to 65 years old ($n = 235$) according to age, and then the risk scores of

patients in each age group were calculated by GILncSig. Patients in each group were divided into high-risk and low-risk groups according to the median risk score. The results showed significant differences in survival between the two groups ($P < 0.01$ Figure 6A). Next, TCGA patients were also divided into male group ($n = 255$) and female group ($n = 122$) according to gender, and then patients in each group were divided into high-risk group and low-risk group by GILncSig. In the male group, the difference in OS between the high and low risk groups was considered significant and meaningful, whereas in the female group, the result was not satisfactory (Figure 6B, $P < 0.001$, $P = 0.952$). We next used the same method to divide patients into two groups according to other different clinical conditions, such as grade, stage, T stage, M stage, N stage, and then divided them into the high and low risk group using GILncSig according to these groups. As expected, Figure 6C-G shows that in most clinical subgroups, the OS of low-risk patients was significantly better than that of high-risk patients, including Grade1-2 ($p < 0.001$), M0 ($p < 0.001$), N0 ($p < 0.001$), T1-2 ($P = 0.002$), and Stage I-II ($P = 0.006$). Nevertheless, the results in the M1, N1-3 and stage III-IV were seemingly meaningless ($p > 0.1$), the p value was only slightly significant in Grade3-4 (0.089), T3-4 ($P = 0.085$).

These findings may mean that GILncSig can be used as a reliable independent prognostic factor to predict the prognosis of HCC patients. And appears to be a better predictor of prognosis for HCC patients in the early stages of the disease

Further exploration of the predictive power of GILncSig

TP53 mutation is the most common mutation in HCC, and it affects the progression and prognosis of it^[26]. Mutations in TP53 are closely related to poor survival in HCC patients, and can be used as an independent prognostic biomarker in HCC^[27]. As shown in the Figure 7A-C, the percentage of patients with TP53 mutations was 51%, 43% and 47% in the high-risk groups of the training set, testing set and TCGA set, respectively, which were significantly higher than the proportion of 21%, 12% and 16% in the low-risk group in each set. This suggests that GILncSig is also associated with

TP53 mutation status. In addition, K-M survival analysis of TCGA patients was further performed in combination with TP53 mutation status and GILncSig. As expected, patients in the TP53 wild-type combined with genomic stable-like group had the best prognosis and those in the TP53 mutant combined with genomic unstable-like group had the worst prognosis. Patients in the same TP53 mutation status had a better prognosis than those in the genomic instability-like group. ($P = 0.009$, Figure 7D). These results suggest that GILncSig may have a more reliable predictive power for HCC patients than TP53 mutation status alone.

DISCUSSION

¹⁹ In the past few years, a large number of studies have been conducted on the initiation, diagnosis and treatment of HCC.^[28, 29] At present, traditional clinicopathological features are still used as a tool to predict the prognosis of HCC.^[30] Imaging examination is indispensable for the diagnosis of liver cancer, but the sensitivity of imaging examination will be greatly reduced due to the small lesions and insignificant symptoms of early liver cancer.^[31] In recent decades, among all biomarkers for the diagnosis of HCC, AFP is the most widely used and relatively reliable. Abnormal plasma AFP level is closely related to the HCC malignancy.^[32] But due to its lack of sensitivity and specificity, its results are not satisfactory in the diagnosis of early liver cancer.^[33] Therefore, it is urgent to find new reliable prognostic indicators to evaluate the prognosis of HCC patients.

¹⁰ In recent years, with the rapid development of high-throughput sequencing technology, GI-LncRNAs have been gradually identified as potential prognostic indicators.^[16, 17] ² It is reported that genomic instability is one of the ubiquitous characteristics of cancer^[10, 34, 35]. It also has great potential as one of the prognostic factors of HCC patients.^[12] In addition, aberrant expression of LncRNAs may affect cell proliferation, tumor progression or metastasis, suggesting that LncRNA may also become a new prognostic factor for HCC by affecting GI^[36]. A considerable number of research have found that some LncRNAs are associated with gene instability, thus

affecting the prognosis of cancer, such as MANCR^[37], CCAT2^[38] and NORAD^[17]. Nevertheless, it is still difficult to identify GI-LncRNAs, its significance in predicting the clinical outcome of HCC is unclear, and their potential as a new prognostic marker remains to be explored. Thus, we constructed a computational framework for identifying genome instability-related LncRNAs by combining LncRNA expression with tumor mutant phenotype.

In this study, we first obtained 88 GI-LncRNAs by comprehensive analysis of the LncRNA profile and somatic mutation downloaded from TCGA database. PCGs closely associated with LncRNAs were identified and analyzed for functional enrichment. Through KEGG and GO pathway analysis, we found that its biological process and biological pathway mainly involved small molecule catabolic process and fatty acid metabolic process, pyrimidine metabolism, purine metabolism, and folate biosynthesis. Pyrimidine metabolism, purine metabolism and folate biosynthesis are involved in DNA synthesis. Dysfunction related to DNA damage will lead to cell cycle imbalance and genomic instability.^[39] In addition, Fanconi anemia pathway is composed of a complex DNA damage signal and repair network, which is very important to prevent genomic instability.^[40]

In addition, we obtained five GI-related LncRNAs (MIR210HG, AC016735.1, AC116351.1, AC010643.1 and LUCAT1), and further explored the GI-related LncRNAs plays the role in predicting the clinical outcome of HCC patients. Based on this, GILncSig was established. Subsequently, GILncSig was used to divide the patients into two groups with high and low risk. In the training set, patients in the low-risk group survived longer than those in the high-risk group, and the independent TCGA set, testing set further validated this result. The ROC curves of GILncSig in the three groups mentioned above were respectively 0.773, 0.679 and 0.736, which means that GILncSig has excellent prognostic ability. In all HCC cohorts, we found that the number of somatic mutations was higher in the high-risk group than in the low-risk group. In addition, the expression of UBQLN4 and H2AX was significantly higher in high-risk patients than that in the low-risk patients either. UBQLN4 is an identified driver of gene

instability in a variety of cancers, and its overexpression in HCC tissues leads to poor prognosis.^[23, 41] A recent study³⁰ indicated that HCC patients with high expression of MIR210HG had a worse prognosis than those with low expression.^[42] LUCAT1 has also been found to be directly associated with the development and progression of cancers, including HCC, and its inhibition of ANXA2 phosphorylation in hepatocellular carcinoma promotes tumorigenesis.^[43,44] AC010643.1 and AC116351.1 have been used as key components of the recently published LncRNA signatures for predicting HCC prognosis, suggesting that they have great potential as new prognostic markers.^[25, 45, 46] But little is known about AC016735.1. In general, these 5 LncRNAs play a crucial role in the pathogenesis of cancer and show their prognostic value potential. TP53 is a common mutation site of cancer, and its mutation type is significantly associated with lower survival rate of HCC patients.^[47, 48] According to⁴ GILncSig, the mutation rate of TP53 in high-risk patients was significantly higher than that in low-risk patients. In addition,² there was a significant difference in survival between high-risk and low-risk patients with TP53 mutations. Therefore, it is of great significance for personalized prognostic evaluation of HCC patients.

Many previous studies have used similar methods to find prognosis related LncRNA and establish LncRNA signatures to predict the prognosis of HCC, such as the study of Huang *et al* and Wu *et al*^[25, 49] Moreover, since all data used in this study were collected from TCGA database, similar results could be obtained when searching for GI-LncRNAs and exploring their functional pathways. The difference is that all HCC patients were divided into the training set and the testing set according to the principle of random grouping. As a result, the calculated prognosis related LncRNAs were different, and the established formula of GILncSig was also different. In addition, the AUC of the GILncSig in this study was relatively high. Subsequently, GILncSig showed good performance in both the independent testing set and TCGA set either. Although this study quantified the GI index of HCC patients and established GILncSig to assess patient outcomes, there are still some limitations that need to be further investigated. Firstly, GILncSig is based on a single TCGA database, which requires an independent,

large and comprehensive comprehensive database for further verification. Due to the limited availability of LncRNAs of HCC samples in GEO database, we did not use GEO database for further study. In addition, GILncSig is determined using the computational framework based on mutation hypothesis. In the future, *in vivo* or *in vitro* experiments are needed to verify its mechanism in the development of liver cancer.

CONCLUSION

In general, we established a computational framework for identifying genome instability-related LncRNAs by combining LncRNA expression with tumor mutant phenotype. It can be used as an independent biomarker to predict the clinical outcome of HCC patients. This is helpful for prognosis assessment and further clinical decision making in HCC patients.

ARTICLE HIGHLIGHTS

Research background

Long stranded non coding RNA (LncRNA) has been found to be a potential prognostic factor for cancer, including hepatocellular carcinoma (HCC). Some LncRNAs have been confirmed as potential indicators for quantifying genomic instability (GI). However, GI LncRNA has yet to be largely explored. This study established the LncRNA features derived from genomic instability (GILncSig), which can predict the prognosis of HCC patients.

Research motivation

We established a GILncSig that can predict the prognosis of HCC patients, which can help to guide the prognostic evaluation and treatment decisions.

Research objectives

Establishing genomic instability derived LncRNA signature (GILncSig) for predicting the prognosis of hepatocellular carcinoma patients. At present, the treatment of liver

cancer has achieved certain results. However, existing research evidence suggests that the treatment options currently used in clinical practice are still relatively ineffective. The objective effective rate of treatment is still largely insufficient, and most patients have not received good returns. The 5-year OS of metastatic hepatocellular carcinoma is still not ideal. At present, the work should mainly focus on expanding treatment targets and searching for reliable biomarkers as much as possible, which will help adjust treatment choices and avoid the risks and costs associated with drug ineffectiveness and side effects. Therefore, there is an urgent need for new biomarkers to predict the prognosis of HCC patients.

Research methods

Identifying the GI-LncRNAs by combining LncRNA expression and somatic mutation profiles. Next, GI-LncRNAs were analyzed for functional enrichment. GILncSig was established in the training set by Cox regression analysis, and its predictive ability was verified in testing set and TCGA set. In addition, we explored the effects of GILncSig and TP53 on the prognosis.

Research results

A total of 88 GI-LncRNAs were found, and functional enrichment analysis showed that their functions were mainly involved in small molecule metabolism and GI. GILncSig was constructed by 5 LncRNAs (MIR210HG, AC016735.1, AC116351.1, AC010643.1, LUCAT1). In the training set, the prognosis of high-risk patients was significantly worse than that of low-risk patients, and similar results were verified in the testing set and TCGA set. Multivariate Cox regression analysis and stratified analysis confirmed that GILncSig could be used as an independent prognostic factor. The ROC curve analysis of GILncSig showed that its AUC (0.773) was higher than the two LncRNA signatures published recently. Furthermore, GILncSig may have a better predictive performance than TP53 mutation status alone.

Research conclusions

We established a GILncSig that can predict the prognosis of HCC patients, which can help to guide the prognostic evaluation and treatment decisions.

Research perspectives

To find new reliable biomarkers to predict the prognosis of HCC patients, adjust the treatment plan, and avoid the risks and costs associated with drug ineffectiveness and side effects.

17%

SIMILARITY INDEX

PRIMARY SOURCES

- | | | |
|---|---|----------------|
| 1 | www.ncbi.nlm.nih.gov
Internet | 159 words — 3% |
| 2 | nursing.usp-pl.com
Internet | 105 words — 2% |
| 3 | academic.oup.com
Internet | 64 words — 1% |
| 4 | www.frontiersin.org
Internet | 56 words — 1% |
| 5 | Yijie Ma, Tong Luo, Dong Dong, Xubo Wu, Yilin Wang.
"Characterization of long non-coding RNAs to reveal
potential prognostic biomarkers in hepatocellular carcinoma",
Gene, 2018
Crossref | 34 words — 1% |
| 6 | Liang Huang, Yu Xie, Shusuan Jiang, Weiqing Han,
Fanchang Zeng, Daoyuan Li. "The lncRNA Signatures
of Genome Instability to Predict Survival in Patients with Renal
Cancer", Journal of Healthcare Engineering, 2021
Crossref | 32 words — 1% |
| 7 | www.aging-us.com
Internet | 31 words — 1% |

8 Yingchun Liang, Fangdie Ye, Zhang Cheng, Yuxi Ou, Lujia Zou, Yun Hu, Jimeng Hu, Haowen Jiang. 30 words — 1%

"Calculated identification of mutator-derived lncRNA signatures of genomic instability to predict the clinical outcome of muscle-invasive bladder cancer", Cancer Cell International, 2021

Crossref

9 Wang Jianfeng, Wang Yutao, Bi Jianbin. "Long non-coding RNAs correlate with genomic stability in prostate cancer: A clinical outcome and survival analysis", 28 words — 1%

Genomics, 2021

Crossref

10 f6publishing.blob.core.windows.net 27 words — < 1%

Internet

11 Xianyu Dai, Hongliang Cao, Hongjie Wang, Rong Zhong et al. "A Robust Ferroptosis-Related 24 words — < 1%

Prognostic Model Associated With Immune Infiltration, Tumor Mutation Burden in Bladder Cancer", Research Square Platform LLC, 2022

Crossref Posted Content

12 aging-us.com 23 words — < 1%

Internet

13 coek.info 20 words — < 1%

Internet

14 www.scilit.net 18 words — < 1%

Internet

15 Junyu Long, Anqiang Wang, Yi Bai, Jianzhen Lin, Xu Yang, Dongxu Wang, Xiaobo Yang, Yan Jiang, Haitao Zhao. "Development and validation of a TP53-associated 17 words — < 1%

immune prognostic model for hepatocellular carcinoma",
EBioMedicine, 2019

Crossref

16 Mingjun Xu, Ting Ma, Shanping Shi, Jingjun Xing, Yang Xi. "Development and Validation of a Mutational Burden-Associated LncRNA Signature for Improving the Clinical Outcome of Hepatocellular Carcinoma", Life, 2021

17 words — < 1%

Crossref

17 Wenbin Shen, Wei Jiang, Shuang Ye, Min Sun, Huijuan Yang, Boer Shan. "Identification of epigenetic genes for predicting prognosis and immunotherapy response of ovarian cancer", Japanese Journal of Clinical Oncology, 2022

17 words — < 1%

Crossref

18 pdfs.semanticscholar.org

Internet

17 words — < 1%

19 Chao Chen, Zhiyu Jie, Kainan Wang. "Fatigue life evaluation of high-strength steel wires with multiple corrosion pits based on the TCD", Journal of Constructional Steel Research, 2021

15 words — < 1%

Crossref

20 bmccancer.biomedcentral.com

Internet

15 words — < 1%

21 cris.biu.ac.il

Internet

15 words — < 1%

22 Lan Lei, Yu-yin Lan, Li-Ting Chen, Zhi-Bing Wu. "6-phosphofructo-2-kinase is a potential prognostic gene in gastric cancer and associated with immune infiltration", Research Square Platform LLC, 2022

14 words — < 1%

Crossref Posted Content

23 Aierpati Maimaiti, Xixian Wang, Yinan Pei, Nuerbiye Nuermaimaiti et al. "Identification and validation of a novel eight mutant-derived long non-coding RNAs signature as a prognostic biomarker for genome instability in low-grade glioma", Aging, 2021

13 words — < 1%

Crossref

24 Dan-Ping Huang, Mian-Mian Liao, Jing-Jing Tong, Wei-Qu Yuan et al. "Construction of a genome instability-derived lncRNA based risk scoring system for the prognosis of hepatocellular carcinoma", Aging, 2021

13 words — < 1%

Crossref

25 Rama K.R. Mettu, Ying-Wooi Wan, Jens K. Habermann, Thomas Ried, Nancy Lan Guo. "A 12-Gene Genomic Instability Signature Predicts Clinical Outcomes in Multiple Cancer Types", The International Journal of Biological Markers, 2018

13 words — < 1%

Crossref

26 Tingting Cui, Jiantao Guo, Zhixia Sun. "A computational prognostic model of lncRNA signature for clear cell renal cell carcinoma with genome instability", Expert Review of Molecular Diagnostics, 2021

13 words — < 1%

Crossref

27 e-century.us

Internet

13 words — < 1%

28 rcastoragev2.blob.core.windows.net

Internet

13 words — < 1%

29 www.karger.com

Internet

13 words — < 1%

30 www.science.gov

Internet

13 words — < 1%

31 Aman Xu, Baichuan Zhou, Xiaodong Zhu, Lixiang Zhang, Panquan Luo, Zhijian Wei, Wenxiu Han. "Identification and validation of mutator-derived lncRNA signatures of genome instability for predicting prognosis of Pancreatic Cancer patients", Research Square Platform LLC, 2022 12 words — < 1%
Crossref Posted Content

32 Chen-Rui Guo, Yan Mao, Feng Jiang, Chen-Xia Juan, Guo-Ping Zhou, Ning Li. "Computational detection of a genome instability-derived lncRNA signature for predicting the clinical outcome of lung adenocarcinoma", Cancer Medicine, 2021 12 words — < 1%
Crossref

33 Ziyu Xun, Yanyu Wang, Junyu Long, Yiran Li, Xu Yang, Huishan Sun, Haitao Zhao. "Development and validation of a genomic instability-related lncRNA prognostic model for hepatocellular carcinoma", Frontiers in Genetics, 2023 12 words — < 1%
Crossref

34 www.researchsquare.com 12 words — < 1%
Internet

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 12 WORDS
EXCLUDE MATCHES < 12 WORDS