Development of genomic instability-associated long non-coding RNA signature: a prognostic risk model of clear cell renal cell carcinoma

Hua Jiang
Zhongda Hospital Affiliated to Southeast University

Ming Chen (✉ 3075818401@qq.com)
Zhongda Hospital Affiliated to Southeast University

Research Article

Keywords: clear cell renal cell carcinoma, TCGA, long non-coding RNA, prognosis, bioinformatics

Posted Date: January 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1291990/v1

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

Background: Renal cell carcinoma is one of the most common cancers in the urinary system. Renal clear cell carcinoma is the most lethal of all pathological subtypes. Genomic instability was recently reported to be relative with occurrence and development of renal cells. The biological roles of long non-coding RNAs (lncRNAs) in tumorigenesis have been increasingly valued, and various lncRNAs were found to be oncogenes or cancer suppressors. Here, we identified a novel genomic instability-associated lncRNAs (GILncs) model for ccRCC patients to predict overall survival (OS).

Methods: The Cancer Genome Atlas database was applied to obtain full transcriptome data, somatic mutation profiles and clinical characteristics. The differential expressed lncRNAs between genome-unstable (GU) like group and genome-stable (GS) like group were defined as GILncs, with |logFC| >1 and adjusted $p$ value< 0.05 of false discovery rate. All samples were allocated into GU-like type and GS-like type based on expression of GILncs using hierarchical cluster analyses. A genomic instability-associated lncRNA signature (GILncSig) was constructed using parameters of included lncRNAs. Quantitative real-time PCR analysis was used to detect the expression of included lncRNAs in vitro. Validation of the risk model was performed by Kaplan–Meier (K-M) log-rank test, Time-dependent receiver operating characteristic curve analysis and multivariate Cox regression analysis.

Results: 46 lncRNAs were identified as GILncs. LINC00460, AL139351.1 and AC156455.1 were employed for GILncSig calculation based on the results of multivariate Cox analysis. GILncSig was confirmed as an independent index for predicting OS and concerned with malignant progression of ccRCC. Additionally, the GILncSig Score presented better efficiency and accuracy than two lncRNA-related models published before.

Conclusion: GILncSig Score was qualified as a critical indicator independent from other clinical factors for predicting prognostic risk of ccRCC patients.

Background

Renal cell carcinoma (RCC) accounts for approximately 85% of urinary cancers generating from kidney, and the according morbidity is growing continuously over the recent years[1]. Nearly 100,000 patients die from RCC annually all over the world, and more than 170,000 deaths were observed globally in 2018 according to recent statistics[2, 3]. Histologically, RCC can be classified into five subtypes with unique characteristics, including clear cell, papillary, chromophobe, collecting duct, and unclassified renal cell carcinoma. About 70~80% of patients were diagnosed with ccRCC after tumor biopsy or nephrectomy[4]. Generally, ccRCC with metastases was associated with high mortality, and over 25% of patients first diagnosed with ccRCC were reported to be distant metastatic with a 5-year survival rate of 0~10%[5]. Besides, the number of deaths from clear cell RCC (ccRCC) accounts for the most among all subtypes of kidney cancers.
Genomic instability has been widely acknowledged as a trigger to carcinogenesis which requests therapeutic intervention. Usually, cancer genomic instability functions as a promotor which improves the development of carcinomatous characteristics\cite{6, 7}. The more frequently the genetic alterations arise, the more likely the genomic instability occur during cell cycle\cite{8}. High level of genomic instability with numerous somatic mutations could lead to malignant progression, distant metastasis and poor prognosis in multiple cancers\cite{9-11}. It was currently reported that genomic instability of various critical genes in RCC cells can affect metabolic features of tumor and disturb the process of cell division, thus resulting in the malignancy progression\cite{8, 12}.

IncRNAs are defined as transcripts comprising over 200 nucleotides which are unable to encode proteins. They function as regulators equipped with diverse biological functions in tumor-associated signaling pathways correlated genes, including epigenetic regulation, transcriptional regulation, and post-transcriptional regulation\cite{13, 14}. In multiple cancers, such as breast cancer, prostate cancer, colorectal cancer as well as RCC, aberrant IncRNAs were constantly detected, which indicated different processes of tumorigenesis. For example, dysregulated Lipid-associated IncRNAs could be regarded as a predictor for the poor prognosis of cancer\cite{15}. Moreover, IncRNA-URRCC, which overexpresses in RCC samples, was found to be related to poor prognosis and the acceleration of cell proliferation and invasion in ccRCC\cite{16}.

Nowadays, although aberrantly expressed IncRNAs and genomic instability are both considered to play core roles in the cancerization of renal cells, it is still unclear whether there is a clinical association between them\cite{9-14}. Notably, evidence unraveling that IncRNAs’ critical roles in genomic instability maintenance, but the genomic instability-associated IncRNAs and their prognostic significance in cancer patients remained to be identified. Therefore, we attempted to establish a risk-score model aimed at predicting long-term outcomes of ccRCC patients in this research, based on statistics from The Cancer Genome Atlas (TCGA) database.

Results

Sample cluster dependent on Genomic instability-associated IncRNAs

The research process was shown in Figure 1. Detailed clinical characteristics of ccRCC patients were described in Table 1, including age, gender, grade, clinical stage and TNM, and no difference was detected between subgroups. Firstly, 170 samples were divided into GU group and GS group, on basis of cumulative number of somatic mutations, and the expressions of 46 genes were significantly different between the two groups (Figure 2A). With the hierarchical cluster analyses, all samples were clustered based on the expression levels of qualified IncRNAs. 53 and 486 samples were classified into GS-like type and GU-like type respectively (Figure 2B). Consistently, the frequency of somatic mutation was significantly higher in GU-like type than in GS-like type (Figure 2C). For reverse verification, we compared the expression of the novel GILnc AC021744.1 in two groups, which was shown to highly express in GU
subgroup. As was expected, AC021744.1 expressed significantly higher in GU-like type compared with GS-like type (Figure 2D).

**LncRNA-mRNA co-expression net and biological function prediction**

Interactions between lncRNAs and mRNAs were visualized in a network consisting of nodes and lines (Figure 3A). Green nodes represented critical GILncs, and red nodes represented the 10 most related mRNAs regulated by lncRNAs. The lines connecting them showed the degree of relevancy. To further predict the potential biological roles of mRNAs which they play in ccRCC, the functional enrichment analysis was employed for Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways research. As is displayed in Figure 3B, the ‘monovalent inorganic cation homeostasis’ (GO:0055067, \( p \text{ value}=1.58 \times 10^{-6} \)) and ‘ear development’ (GO:0043583, \( p \text{ value}=6.98 \times 10^{-5} \)) were GO terms of biological process (BP) that most genes are involved in. In terms of cellular component (CC), the number of genes that enriched in ‘apical part of cell’ (GO:0045177, \( p \text{ value}=5.70 \times 10^{-6} \)) was the greatest, reaching 20. As far as molecular function (MF) is concerned, most of genes are likely to participate in ‘monovalent inorganic cation transmembrane transporter activity’ (GO:0016324, \( p \text{ value}=0.0001 \)). What is more, KEGG enrichment analysis was also performed to find biological signal pathways in which GILncs might be involved. In accordance with the bubble diagram (Figure 3C), included genes were mostly inclined to concentrate on KEGG pathways of ‘Human papillomavirus infection’ (n=11, \( p \text{ value}=0.026 \)) and ‘PI3K-Akt signaling pathway’ (n=11, \( p \text{ value}=0.039 \)).

**Establishment of GILncSig and prognostic risk model**

Using multivariate Cox regression analysis, 3 lncRNAs closely related to OS were selected for model establishment, including LINC00460, AL139351.1 and AC156455.1. To validate the result, human renal cells were used to confirm the high expression of included lncRNAs.

Next, all ccRCC samples were randomly allocated to training set (n=257) and testing set (n=256) for further validation of the prognostic risk model. The GILncSig was utilized as the index for risk group classification. In order to figure out the association between risk score and prognosis, we performed survival analyses of both training set and testing set. K-M curves of OS were drafted to compare survival outcomes (Figure 4A, 4F) in two sets, and both of them showed significantly better OS in low-risk group than high-risk group (log-rank \( p < 0.001 \)). Areas under the curve (AUC) values of receiver operating characteristic (ROC) curves were applied to assess the reliability of our model (Figure 4B, 4G). AUC values of training set (AUC=0.691) and testing set (AUC=0.689) were simultaneously close to 0.7, indicating relatively high effectiveness for prognosis prediction. The correlation between prognostic risk score and AC021744.1 expression was illustrated in expplots (Figure 4C, 4H) and heatmaps (Figure 4D, 4I). The results suggested that as the risk score increases, the expression level of lncRNA AC021744.1 would upgrade consistently in both training set and testing set. Additionally, lncRNA LINC00460, AL139351.1 and AC156455.1 were found to be upregulated when risk score was higher, which further clarified these
three lncRNAs as the high-risk genes. The mutplots (Figure 4E, 4J) showed that for both training set and testing set, higher somatic mutation counts came with higher risk scores.

Relevance between risk score and somatic mutation count/AC021744.1 expression

As demonstrated in two sets of box plots in Figure 5, the associations between prognostic risk score and mutation counts or AC021744.1 expression level were further assessed in pairs. A significantly positive trend could be illustrated between mutation frequency and risk of prognosis in training set ($p$ value=0.0067) (Figure 5A), while none significant association was detected in testing set (Figure 5C). No association between risk score and AC021744.1 expression level was found in training set, neither in testing set (Figure 5B, 5D).

Validation of GILncSig as independent prognostic factor from clinical characteristics

To further investigate whether GILncSig could be identified as an independent predictor of OS, we comprehensively researched on several potential prognostic indicators with integration of univariate and multivariate Cox regression analysis. Factors of interest included age, gender, pathological grade, clinical stage and GILncSig (Table 2). The consequence of Cox regression analysis revealed that the GILncSig Score could further be identified as an independent prognostic factor affecting OS from other potential factors in testing set (HR (hazard ratio) 1.110, 95% CI (confidence interval) 0.987-1.248, $p$ value=0.019) and TCGA set (HR 1.015, 95% CI 1.007-1.023, $p$ value<0.000). Similarly, pathological grade showed significant association with survival outcomes in testing (HR 1.633, 95% CI 1.146-2.327, $p$ value=0.007) and TCGA set (HR 1.428, 95% CI 1.131-1.802, $p$ value=0.003), as well. Meanwhile, age and clinical stage could significantly influence survival of ccRCC patients in all sets.

In addition, we performed a series of K-M log-rank analyses to further validate its consistency as an independent prognostic factor in different subgroup samples. In each pair of subgroups, survival time was positively correlated with risk score, under the effects of age (Figure 6A, 6B), gender (Figure 6C, 6D), pathological grade (Figure 6E, 6F) and clinical stage (Figure 6G, 6H).

To conclude the results above, the GILncSig Score could be deemed as a prognostic predictor with consistent independence, which was negatively correlated with OS of ccRCC patients.

Relationship between risk and oncogene mutation

PBRM1 has been recognized as a classic oncogene taking part in genomic instability, of which the high mutation frequency was related to occurrence of ccRCC[19]. In order to figure out whether the new signature of risk estimation in ccRCC is capable of predicting prognosis, the single genetic mutation count was calculated in training set and testing set. The proportion of mutation appeared to be larger in high-risk group in both of the two sets (Figure 7A, 7B), whereas no statistical significance was probed (training set $p$=0.235, testing set $p$=0.276).

Expression Level of GILnc in different cell lines
As the figure showed, compared with normal renal cell, the expressions of LINC00460, AL139351.1 and AC156455.1 were all higher in RCC cells 786-O and Caki-2 tested by qPCR (Figure 7C).

Comparison of IncRNA-related prognostic prediction models

In order to evaluate the efficiency of our prognostic model of GILncSig, we compared it with other two published IncRNA-related prognostic prediction models for ccRCC. AUCs of the three ROC curves corresponding to the three models represented relative predictive accuracy. As was illustrated in Figure 7C, the AUC of ZengLncSig[20], a six-IncRNA-based risk model, was 0.500. For another five immune-related IncRNA signature of Sun[21], the AUC was 0.679. Accordingly, our GILncSig of 3 IncRNAs with the AUC of 0.688 exhibited the most effective prediction of prognosis for patients with ccRCC.

Discussion

CcRCC is the most prevalent histological subtype of kidney cancer, taking a major part in yearly mortality related to cancers[4, 5]. While IncRNAs have become increasingly notable for its multiple biological roles in tumorigenesis process in ccRCC, researches on their functions of prognostic risk prediction were still insufficient[22-24]. The occurrence of somatic mutation in cancer-related genes has been considered as a critical inducement of RCC, and there is an increasing number of etiological studies on genomic instability. For instance, Wang W, et al put forward that genomic instability existing in the DNA repair gene Ku70 was a considerable cause of RCC in 2012[25]. Moreover, genomic DNA hypomethylation was proved as a hallmark of RCC risk by Mendoza-Pérez J, et al, which promoted the genomic instability of the global genome, offering further hypothesis on the etiology of RCC tumorigenesis[26]. Till 2019, Renzo G DiNatale, et al demonstrated that the mutation on TCEB1 could diminish Von Hippel-Lindau (VHL) gene's suppressive effects on ccRCC. Therefore, molecular events contributing to high genomic instability were proved to enforce the aggressiveness of ccRCC and associated with more clinical adverse outcomes[27-29]. In recent years, several studies investigated into the association between survival outcomes and different clusters of IncRNAs in ccRCC samples, and established risk models to validate the predictive ability of IncRNAs. According to former researches, an immune-related signature successfully predicted clinical outcomes through synthesizing 5 IncRNAs (AC008105.3, LINC02084, AC243960.1, AC093278.2, and AC108449.2) extracted data from TCGA database by Sun Z, et al [21]. Zeng JH, et al proposed a practical six-IncRNA prognostic risk model (CTA-384D8.35, CTD-2263F21.1, LINC01510, RP11-352G9.1, RP11-395B7.2, RP11-426C22.4) based on expression levels of involved non-coding genes in ccRCC samples [20]. However, prognostic models with the theme of GILncs were rarely reported. Therefore, when constructing predictive models for prognostic risk of ccRCC patients, we were encouraged to bring out an original index, GILncSig Score, and to carry out a comprehensive analysis to validate relations among GILnc and clinical outcomes[30-32].

We eventually screened out 46 GILncs equipped with differential mutation frequency in 539 samples from TCGA database with outright open access. After seriatim referring to correlative literatures, we found that very few studies discussed on non-coding genes AC016405.3, AC114803.1, AC156455.1,
AL139351.1, OSTM1-AS1 and AC015977.2 in GIlnSig computation. Nevertheless, highly-mutant lncRNA AC021744.1 was reported to be an indicator of poor prognosis among patients with hepatocellular carcinoma (HCC). Overexpression of AC021744.1 could directly lead to shorter recurrence-free survival (RFS) time because of more severe fibrosis of livers[33]. As the only known oncogene relevant to genomic instability in our model, we took AC021744.1 into the next research as a representative. In accordance with individual distribution of 46 genome-unstable lncRNAs, all 539 samples were clustered as GS-like type (n=53) and GU-like type (n=486). In addition, in Bao S’s study in 2019[34], AC021744.1 was also found to be a gene instability-related lncRNA, which was significantly correlated with the gene instability-driving gene UBQLN4 and played an important role in the occurrence and development of breast cancer. Hence, we took AC021744.1 as a representative genomic instability associated lncRNA for further validation. We found that with the presence or absence of genomic instability, levels of somatic mutation count respectively appear low and high, verifying the positive association among included parameters.

In the further detailed construction of co-expression network among genomic instability-related lncRNAs and regulated mRNAs, the relevancy between lncRNAs and mRNAs were quantitative synthesized and analyzed. We unfortunately did not obtain further detailed co-regulation prediction from any other databases when attempted to look up for more prediction of co-expression among genes and the lncRNAs of interest. We assumed it was because the functions and signaling pathways of genomic instability-associated lncRNAs were rarely studied or reported, based on literature retrieval. Due to the limitations, more original sequencing statistics are anticipated for further co-expression or ceRNA analyses.

In the meantime, GO and KEGG enrichment analyses of susceptible mRNAs were conducted for biological function forecast. Noticeably, mRNAs targeted downstream were highly enriched at the sites of the ‘PI3K-Akt signaling pathway’, in line with the result of KEGG analysis. PI3K/Akt signaling pathway has been mentioned in numerous studies on carcinomas, which was reported to participate in the malignant progression and lead to poor prognosis of various cancers, including RCC[35]. Hence, the impacts of genomic instability-related lncRNAs on prognosis of ccRCC patients could be viewed from another aspect. When the upregulated and downregulated gene was analyzed separately in GO analyses, we found that among the 46 genomic instability-associated lncRNAs, most upregulated lncRNAs significantly enriched on GO terms while the enrichments of few candidate downregulated lncRNAs were insignificant. Then we synthesized the enrichment generally in Figure 3B to summarize the enrichment of GIlns on GO terms, no matter it was significant or not.

After the establishment of GIlnSig model, samples could be accordingly divided into high-risk group and low-risk group. Evidently, the somatic mutation frequency and expression of high-risk genes, LINC00460, AL139351.1 and AC156455.1, grew when the prognostic risk was higher. Meanwhile, ccRCC manifested more malignant attributes leading to poorer OS concluded by the incremental expression level of cancer-associated gene, AC021744.1. As no biological function of AL139351.1 and AC156455.1 have been found, attention should be paid to LINC00460, a dysregulated lncRNA reported in RCC in 2018[36]. It has been validated that LINC00460 functions as a competing endogenous RNAs in co-expression and
promotes the malignant development of multiple cancers, including prostate cancer[37], skin cancer[38], hepatocellular cancer[39], colorectal cancer[40] and so on, except for RCC. Coincidentally, Zhang D, et al researched on LINC00460 as well. When identifying a three-lncRNA prognostic model[41], they speculated LINC00460’s competing endogenous feature for its overexpression in ccRCC but did not mention its potential genomic instability. Controversially, the mutation frequency and expression did not exhibit to be consistently differential when quantitative analyses were performed (Figure 5), probably due to the uncertain carcinogenic mechanism in ccRCC. Additionally, our GILncSig Score model possessed great independence as a prognostic predictor from other significant clinical factors, such as age, gender, pathological grade and clinical stage, showing its compatibility for all kinds of ccRCC patients with different clinical characteristics. Besides, when comparing our model with former lncRNA-related signatures, it was worth noting that our GILncSig model showed better prognosis efficiency with higher AUC of 0.688 (Figure 7C).

It was widely acknowledged that RCC was insensitive to radiotherapy and chemotherapy, remedies targeted to specific genes or immune checkpoint inhibitors (ICIs) are now being explored [42, 43]. Since Braun DA, et al. [43] clinically validated the alternations of PBRM1 as a biomarker of ICI response in RCC in 2019, and Carril-Ajuria L, et al. [19] demonstrated the prognostic and predictive value of PBRM1 in ccRCC in the same year, more and more researchers have payed attention to exploring relevant genes which could be hallmarks of immune targeted therapy of PBRM1-mutant ccRCC. For example, in the recent 2021 conference of European Urology, Hagiwara M, et al. [44] brought out that poly ADP-ribose polymerase-1 could be a marker of the efficacy of immunotherapy for patients with PBRM1-mutant ccRCC, of which the higher expression suggested poorer prognosis and higher drug resistance. In our study, we found 46 differentially expressed GILncs between genome-unstable and genome-stable like groups, among which LINC00460, AL139351.1 and AC156455.1 were validated as significant independent prognostic factors and taken into the construction of the risk model. Furthermore, a partially positive correlation between the risk score and the mutation of PBRM1 in either training set or testing set. The mutation of PBRM1 were more likely to appear in high-risk group for both training and testing set, while no statistical significance was discovered. Though the difference was insignificant, this still leaded us to the potential of the risk model to predict response to ICI treatment for patients with PBRM1-mutated ccRCC. We assumed that it owed to the limits of sample size and it remained to be confirmed in the future under the condition of larger sample size or experiments in vivo/vitro.

However, some limitations could not be neglected. In the study, we identified 46 genomic instability-associated lncRNAs based on the somatic mutation counts and expression of lncRNAs, and three prognostic genomic instability-associated lncRNAs were taken into the construction of the risk model with further regression analyses. As the lack of other independent cohorts to perform validation, we applied the same datasets to set training group and test group for internal validation. Admittedly, by changing the distribution of samples, the distribution density of repeated samples would increase, thus we anticipate more data from independent cohorts for further validation.
In addition, due to the lack of extra independent dataset for external validation, the qPCR analysis comparing the mRNA expression in ccRCC cell lines and normal renal tubular epithelial cell was applied. As a result, GILncs of the risk model were expressed significantly higher in the tumor cell lines. Through this result, we externally validated the overexpression of GILncs play potential roles in tumor development and predicting high risk, which was consist with the conclusion described in Figure 4D and I that the expressions of LINC00460, AL139351.1 and AC156455.1 were upregulated in groups of high risk.

Last but not least, despite the fact that the GILncSig model was able to predict clinical outcomes more efficiently, objective shortages inevitably existed in our study. Since the data were extracted retrospectively from databases, the outcomes may be not accurate enough, and prospective validations based on experimental results would make the results more persuasive. For example, lncRNA AC021744.1 was identified as an oncogene related to poor prognosis of HCC patients[33], but its carcinogenesis in RCC still lacks further proof from laboratory. Consequently, the arguments using AC021744.1 could not completely support the correlation of the risk score and OS. Also, functions of AL139351.1 and AC156455.1 included in GILncSig remain to be discovered.

To sum up, GILncSig Score, an original index calculated with the coefficients and expression levels of GILncs, is qualified to be a critical indicator independent from other clinical factors for predicting prognostic risk of ccRCC patients.

**Conclusion**

We established and validated a prognostic risk model called GILncSig, based on expressions of three genomic instability-associated lncRNAs LINC00460, AL139351.1 and AC156455.1. Besides, the higher the GILncSig scores were, the poorer prognosis the ccRCC patients would have.

**Methods**

**Data Retrieval and Sample Classification**

We obtained full transcriptome data, somatic mutation profiles and clinical characteristics of 539 patients from TCGA database (https://cancergenome.nih.gov/). To identify the possible relations between lncRNAs and genomic instability, lncRNA expression profiles and somatic mutation profiles were analyzed. After somatic mutation frequency calculation of every sample, all the samples were ranked in a descending order. The top 25% and the bottom 25% of them were allocated to genome-unstable (GU) like group and genome-stable (GS) like group, respectively. The differences of lncRNA expression levels between the two groups were evaluated with significance analysis of microarrays. Differential expressed lncRNAs with |logFC| >1 and adjusted p < 0.05 in false discovery rate were defined as GILncs. According to expression levels of specific lncRNAs, all samples were separately assigned to GU-like type and GS-like type using hierarchical cluster analyses.

**Co-expression net and Gene functional exploration**
Co-expression regulatory net model was constructed in order to analyze the correlation among 46 lncRNAs and corresponding susceptible mRNAs. The relevancy degrees were measured by Pearson correlation coefficients per cluster. Furthermore, enrichment analyses of most related mRNAs were conducted for function prediction, in terms of GO terms and KEGG pathway[17]. Statistics were processed and visualized with ‘clusterProfiler’ R package[18].

**Quantitative real-time PCR analysis**

RNA extraction kits (OMEGA, China) was used to extracted RNA of kidney tissue. The specific primers used were as follows: 5’ ACGCAGTGGATGAGAACGAA (LINC00460 forward) and 5’ GGGGTGACTTCAGAATGCGT (LINC00460 reverse); 5’CTTCACATTCTACACGCTCTCCT (AL139351.1 forward) and 5’GGTGTGGGTGAAGTAAAGAAAGC (AL139351.1 reverse); 5’CTCAGCTTCTACAGCCTCTCTCCT (AC156455.1 forward) and 5’CGTGTTGAGGACTACAGAG (AC156455.1 reverse). mRNA expressions were normalized to GAPDH expression. Every experiment was repeated at least three times.

**Prognosis-related statistics analysis**

To identify prognosis-related GILncs independently from other features of ccRCC patients, univariate and multivariate Cox proportional hazard regression analyses were performed. The qualified prognosis-related lncRNAs were taken into prognostic model construction, of which \( p \) values<0.05 was considered statistically significant.

Combining the expression of prognosis-related GILncs and coefficients from multivariate Cox regression, a GILncSig for predicting prognosis was founded, and the computation formula is as follows:

\[
\text{GILncSig (patient)} = \sum_{i=1}^{n} \text{coef (lncRNA}_i \text{)} \ast \text{expr (lncRNA}_i \text{)}
\]

GILncSig (patient) represents a prognostic risk score for ccRCC patients. LncRNA\(_i\) represents the \(i\)th prognosis-related lncRNA. Expr (lncRNA\(_i\)) is the lncRNA\(_i\) expression level, and coef (lncRNA\(_i\)) is the contribution of lncRNA\(_i\) to the risk score derived from multivariate Cox regression coefficients. Classification between low-risk group with low GILncSig and high-risk group with GILncSig relying on the risk cutoff was computed by the median score of the patients in the training set.

Accuracy of the predictive model for each group was evaluated by the K-M log-rank test, Time-dependent ROC curve analysis and multivariate Cox regression analysis. KM survival curves were analyzed to determine correlations among all parameters, including clinical characteristics and GILncSig. HR, 95% CI and \( p \)-value were standards for identifying independent prognostic indicators. ROC curves were utilized to evaluate the predictive effectiveness of the genome unstable lncRNA-based risk scores for prognosis of ccRCC patients. A two-sided \( P \)-value < 0.05 threshold was considered statistically significant. All statistical analyses were conducted with R version 4.0.3.
Abbreviations

IncRNAs: long non-coding RNAs

GILncs: genomic instability-associated IncRNAs

OS: overall survival

GU: genome-unstable

GS: genome-stable

GILncSig: genomic instability-associated IncRNA signature

K-M: Kaplan–Meier

RCC: renal cell carcinoma

ccRCC: clear cell RCC

TCGA: The Cancer Genome Atlas

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

BP: biological process

CC: cellular component

MF: molecular function

AUC: areas under the curve

ICIs: immune checkpoint inhibitors

HCC: hepatocellular carcinoma

HR: hazard ratio

CI: confidence interval

ROC: receiver operating characteristic

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that there is no competing of interest.

Acknowledgements
None declared.

Author's Contribution
Conceptualization, Ming Chen; Data curation, Hua Jiang; Funding acquisition, Ming Chen; Writing – review & editing, Hua Jiang. both authors approved the submission of the manuscript.

Data Availability Statement
Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

Funding Support
This study was funded by Jiangsu Provincial Key Research and Development Program (BE2019751).

References


44. Hagiwara M, Fushimi A, Matsumoto K, Oya M. **The Significance of PARP1 as a biomarker for Predicting the Response to PD-L1 Blockade in Patients with PBRM1-mutated Clear Cell Renal Cell Carcinoma.** EUR UROL 2021, S0302-2838(21)02054-6.

**Tables**

Table 1: Clinical information for three sets in ccRCC patients
<table>
<thead>
<tr>
<th>Covariates</th>
<th>Training set (n = 140)</th>
<th>Testing set (n=136)</th>
<th>TCGA set (n = 276)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, n (%)</td>
<td>&lt;65 164 (64.3)</td>
<td>176 (68.75)</td>
<td>169 (66.28)</td>
<td>0.276</td>
</tr>
<tr>
<td></td>
<td>≥65 93 (36.19)</td>
<td>80 (31.25)</td>
<td>173 (33.72)</td>
<td></td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td>Female 90 (35.02)</td>
<td>86 (33.59)</td>
<td>176 (34.31)</td>
<td>0.805</td>
</tr>
<tr>
<td></td>
<td>Male 167 (64.98)</td>
<td>170 (66.41)</td>
<td>337 (65.69)</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>G1-2 111 (43.19)</td>
<td>120 (46.88)</td>
<td>231 (45.03)</td>
<td>0.5005</td>
</tr>
<tr>
<td></td>
<td>G3-4 141 (54.86)</td>
<td>133 (51.95)</td>
<td>274 (53.41)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknow 5 (1.95)</td>
<td>3 (1.17)</td>
<td>8 (1.56)</td>
<td></td>
</tr>
<tr>
<td>Stage, n (%)</td>
<td>Stage I/II 95 (67.9)</td>
<td>91 (66.9)</td>
<td>311 (60.62)</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>Stage III/IV 103 (40.08)</td>
<td>96 (37.50)</td>
<td>199 (38.79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknow 2 (0.78)</td>
<td>1 (0.39)</td>
<td>3 (0.58)</td>
<td></td>
</tr>
<tr>
<td>T stage, n (%)</td>
<td>T1/T2 160 (62.26)</td>
<td>169 (66.02)</td>
<td>329 (64.13)</td>
<td>0.4263</td>
</tr>
<tr>
<td></td>
<td>T3/T4 97 (37.74)</td>
<td>87 (33.98)</td>
<td>184 (35.87)</td>
<td></td>
</tr>
<tr>
<td>M stage, n (%)</td>
<td>M0 194 (75.49)</td>
<td>213 (83.20)</td>
<td>407 (79.34)</td>
<td>0.0339</td>
</tr>
<tr>
<td></td>
<td>M1 48 (18.68)</td>
<td>30 (11.72)</td>
<td>78 (15.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknow 15 (5.84)</td>
<td>13 (5.08)</td>
<td>28 (5.46)</td>
<td></td>
</tr>
<tr>
<td>N stage, n (%)</td>
<td>N0 112 (43.58)</td>
<td>117 (45.7)</td>
<td>229 (44.64)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>N1/N2 8 (3.11)</td>
<td>8 (3.12)</td>
<td>16 (3.12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknow 137 (53.31)</td>
<td>131 (51.17)</td>
<td>268 (52.24)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Univariate and Multivariate Cox regression analysis of the GILncSig and OS in different sets
<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariable model</th>
<th>Multivariable model</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HR</td>
<td>95% CI</td>
<td>P-value</td>
<td>HR</td>
<td>95% CI</td>
<td>P-value</td>
</tr>
<tr>
<td>Training set (n = 257)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GILncSig High/Low</td>
<td>1.271</td>
<td>1.148-1.407</td>
<td>0.000</td>
<td>1.110</td>
<td>0.987-1.248</td>
<td>0.081</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.027</td>
<td>1.008-1.046</td>
<td>0.595</td>
<td>1.025</td>
<td>1.004-1.046</td>
<td>0.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender Female/Male</td>
<td>0.889</td>
<td>0.577-1.371</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade (3-4)/(1-2)</td>
<td>2.024</td>
<td>1.527-2.683</td>
<td>0.000</td>
<td>1.268</td>
<td>0.929-1.731</td>
<td>0.135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (III/IV)/(I/II)</td>
<td>1.845</td>
<td>1.546-2.203</td>
<td>0.000</td>
<td>1.656</td>
<td>1.353-2.026</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing set (n = 256)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GILncSig High/Low</td>
<td>1.098</td>
<td>1.060-1.136</td>
<td>0.000</td>
<td>1.044</td>
<td>1.004-1.086</td>
<td>0.030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.030</td>
<td>1.010-1.049</td>
<td>0.003</td>
<td>1.035</td>
<td>1.014-1.057</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender Female/Male</td>
<td>1.039</td>
<td>0.652-1.656</td>
<td>0.872</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade (3-4)/(1-2)</td>
<td>2.538</td>
<td>1.867-3.449</td>
<td>0.000</td>
<td>1.633</td>
<td>1.146-2.327</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (III/IV)/(I/II)</td>
<td>1.948</td>
<td>1.581-2.401</td>
<td>0.000</td>
<td>1.692</td>
<td>1.329-2.154</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCGA set (n = 513)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GILncSig High/Low</td>
<td>1.112</td>
<td>1.079-1.146</td>
<td>0.000</td>
<td>1.015</td>
<td>1.007-1.023</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.029</td>
<td>1.015-1.042</td>
<td>0.000</td>
<td>1.031</td>
<td>1.016-1.046</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender Female/Male</td>
<td>0.964</td>
<td>0.703-1.323</td>
<td>0.821</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade (3-4)/(1-2)</td>
<td>2.268</td>
<td>1.845-2.787</td>
<td>0.000</td>
<td>1.428</td>
<td>1.131-1.802</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (III/IV)/(I/II)</td>
<td>1.896</td>
<td>1.658-2.168</td>
<td>0.000</td>
<td>1.664</td>
<td>1.428-1.938</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figures

Figure 1
The flow chart of research processes

Figure 2

Figure 3
Regulatory network and functional analyses. A) Co-expression network of GI LNc and targeted mRNAs. B) Bar plots of GO terms enrichment analysis. C) The bubble plot of KEGG pathway enrichment analysis. D) mRNA expression of LINC00460, AL139351.1 and AC156455.1 in HK2, 786-O and Caki-2 cells according to qRT-PCR analysis.

Figure 4
Validation of the GI LNc Sig model. A) The survival analysis in training set; B) The ROC curve in training set; C) The expplot analysis in training set; D) The heatmap in training set; E) The mutplot analysis in training set.
set; F) The survival analysis in testing set; G) The ROC curve in testing set; H) The expplot analysis in testing set; I) The heatmap in testing set; J) The mutplot analysis in testing set.

Figure 5

Risk correlation analysis. A, C) Boxplots of correlation between risk levels and somatic mutation count in training group and testing group. B, D) Boxplots of correlation between risk levels and AC021744.1 expression in training group and testing group.

Figure 6

Predictive consistency analyses of GILncSig in populations with different clinical characteristics. A, B) KM curves: Comparisons of OS between ccRCC patients with high and low risk, in old and young groups. C, D) KM curves: Comparisons of OS between ccRCC patients with high and low risk, in male and female groups. E, F) KM curves: Comparisons of OS between ccRCC patients with high and low risk, in pathologically early and advanced groups. G, H) KM curves: Comparisons of OS between ccRCC patients with high and low risk, in clinically early and advanced groups.
Figure 7

A, B) Carcinogenic function analysis of GILncSig based on single geng mutation of PBRM1: Proportional bar plots of mutation type and wild type of PBRM1 in samples of training set and testing set. C) Comparison among three lncRNA-based prognostic models based on AUCs: ROCs of GILncSig, SunLncSig and ZengLncSig.