Vol.63: e20190511, 2020
http://dx.doi.org/10.1590/1678-4324-2020190511
ISSN 1678-4324 Online Edition

Article - Human and Animal Health

Identification of Ten Long Noncoding RNAs as Biomarkers for Hepatocellular Carcinoma

Chao Lin[1,2]
https://orcid.org/0000-0002-7658-9136

Shui Liu[2]
https://orcid.org/0000-0002-6379-9230

Dan Zhang[2]
https://orcid.org/0000-0002-7363-4157

Yien Xiang[2]
https://orcid.org/0000000179941302

Xiaoxiao Yao[2]
https://orcid.org/0000-0002-1045-8627

Xuewen Zhang [2*]
https://orcid.org/0000-0002-3619-2222

[1]China-Japan Union Hospital of Jilin University, Department of Hepatobiliary-Pancreatic Surgery, Changchun City, Jilin Province, China; [2]Jilin University Second Hospital, Department of Hepatobiliary and Pancreatic Surgery, Changchun City, Jilin Province, China.

Received: 2019.11.13; Accepted: 2020.04.01.

*Correspondence: zhangxw@jlu.edu.cn; Tel.: +86-0431-84995114

HIGHLIGHTS

- The gene expressions from 374 tumor patients and 50 normal patients in TCGA were analyzed.
- The abnormal expressions of 387 differentially expressed lncRNAs (DElncRNAs) were identified.
- 10 lncRNAs, whose main functions are cellular metabolic capacity and proliferation, were found.
- The key lncRNAs may be used as new biomarkers for the prognosis of HCC.

Abstract: Long-chain non-encoded RNAs (lncRNAs) are important in many life activities and can participate in the occurrence of hepatocellular carcinoma (HCC). Moreover, lncRNAs can be used as basis for developing new strategies to hinder liver cancer. To investigate the utility of lncRNAs in HCC as potential biomarkers for early detection and diagnosis, we mined genomic data from the Cancer Genome Atlas (TCGA), and analyzed the gene expressions from 374 tumor patients and 50 normal patients. The abnormal expressions of 387 differentially expressed lncRNAs (DElncRNAs) were identified from a total of 3099 lncRNAs. Moreover, 18 modules were divided based on WGCNA, and 2 of the 18 modules were positively correlated with stage and grade, and negatively correlated with survival time. Finally, 10 lncRNAs were found and their main functions are the enhancement of cellular metabolic capacity and cell proliferation. These 10 lncRNAs may serve as novel prognostic markers and therapeutic targets, and may help guide subsequent studies on HCC.

Keywords: long noncoding RNAs; hepatocellular carcinoma; biomarkers.

INTRODUCTION

Liver cancer is the third leading cause of cancer mortality worldwide and has an extremely poor prognosis [1-4]. The incidence of liver cancer in men is higher than in women [5]. Moreover, hepatocellular carcinoma (HCC) is the most common form of liver cancer. The major pathogenic promoters of HCC include viral/alcohol-related liver disease, obesity, type 2 diabetes, and nonalcoholic fatty liver disease[6]. The

underlying molecular mechanisms remain unclear, however. Therefore, exploring convenient and accurate methods for preventing and even healing HCC is important.

Non-encoded RNA refers to functional RNA molecules that cannot be translated into proteins, and they include small interferometric RNA, microRNA, piRNA, and long-chain non-encoded RNA (lncRNA). Long-chain non-encoded RNA is a non-encoded RNA with a length greater than 200 nucleotides[7,8]. Studies have shown that lncRNA is important in many life activities, such as dose-compensation effect, epigenetic regulation, cell cycle regulation, and cell differentiation regulation; thus, it has become a hotspot of genetics research. Studies on the expressions of lncRNAs, such as UFC1[9], lnc00210[10], AB209630[11], and TP73-AS1[12], have been associated with cell proliferation and metastasis in HCC. Moreover, lncRNA have been reported to participate in the occurrence of HCC. For example, lincRNA SNHG20[13], HOTTIP/HOXA13[14] and HOTAIR[15] were explored and could be predictors for the recurrence of liver cancer. These findings provided evidence that the innovative utilization of lncRNAs as biomarkers for HCC disease progression and tumor suppression therapy will be a promising approach for therapeutic options.

Numerous shared databases have been developed for cancer research as networks have become established. Tumor Genome Mapping (TCGA) uses genomic analysis techniques, which are dominated by large-scale sequencing, to understand the molecular mechanisms of cancer through extensive collaboration. Compared with single biomarker prediction of HCC, multi-biomarker prediction based on lncRNA can improve prediction accuracy. In our study, the analysis of differential lncRNA between HCC and adjacent normal tissue based on TCGA database provided credible biomarkers for HCC.

MATERIAL AND METHODS

Acquisition of clinical data and RNA-seq data

Patient data from 374 patients with HCC were provided by the TCGA database (https://cancergenome.nih.gov/). In addition to patient samples with incomplete clinical information, complete data from 358 patients were provided. Moreover, lncRNA and mRNA expression data from 374 tumor and 50 paracancerous samples were downloaded from TCGA.

Different Expression Genes Analysis

The edgeR packages of bioconductor analysis tool for R was applied to detect the differentially expressed genes (DEGs), which were based on a range of statistical methodologies based on the negative binomial distributions, including empirical Bayes estimation, exact tests, generalized linear models, and quasi-likelihood tests. For mRNA, the differential expression genes (DEGs) with $|logFC|>1$ and FDR <0.05 were considered significant; and for lncRNA, DEGs with $|logFC|>3$ and FDR <0.05 were considered significant.

Weighted Correlation Network Analysis

Before using the R package “WGCNA” for analysis, we integrated and normalized the differentially expressed mRNA and all lncRNA in one matrix. Only 358 tumor samples with comprehensive clinical data were included in the study. Patient clinical information included gender, race, clinical stage, survival time, clinical grade, survival status, and age. Moreover, WGCNA was used for scale-free network topology analysis of the RNA-seq expression data of HCC samples. The analytical approach aimed to find co-expressed gene modules and explore the association between gene networks and phenotypes of interest, as well as core genes in the network. The correlation coefficient between genes and genes was based on the Pearson method. Based on the mRNA co-expressed with lncRNA, we predicted the biological function of lncRNA. In this study, we were more interested in genes whose expression levels were up-regulated in tumors.

Functional Gene Enrichment Analysis

Metacape (http://metascape.org/) was used for gene enrichment analysis, including gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). FDR<0.05 was considered statistically significant.

Cox Regression

The relationship between lncRNA expression level imbalance and patient survival was evaluated by univariate Cox regression analysis. Considering the large number of genes, to make the results highly credible, we believed that P value <0.01 was significant. After Wayne's analysis, we found 10 lncRNAs as

biomarkers for liver cancer. Multivariate Cox regression analysis was performed to evaluate these 10 lncRNAs. According to the significant survival lncRNA Cox regression coefficient, each patient was scored for risk and divided into low- and high-risk groups according to median risk. Furthermore, Cox regression analysis relied on the “survival” package in R to complete.

RESULTS

A total of 3099 lncRNAs from 50 normal samples and 374 HCC samples were obtained from TCGA. The expression levels of all lncRNAs standardized via TMM were displayed in the heatmap (Figure 1A). EdgeR, a R package to process the differential expression analysis of RNA-seq expression profiles with biological replication, was performed to identify 387 DElncRNAs, including 382 upregulated and 5 downregulated genes. As shown in the volcano plot, red dots indicated upregulated genes and green dots indicated downregulated genes, which were significantly regulated > 8-fold ($|\log_2 FC|>3$) in HCC samples and FDR<0.05. Black dots show genes with inconspicuous regulation (Figure 1B). Apparently, lncRNA is highly expressed in tumors. Thus, our follow-up study only focused on the high expression of lncRNA.

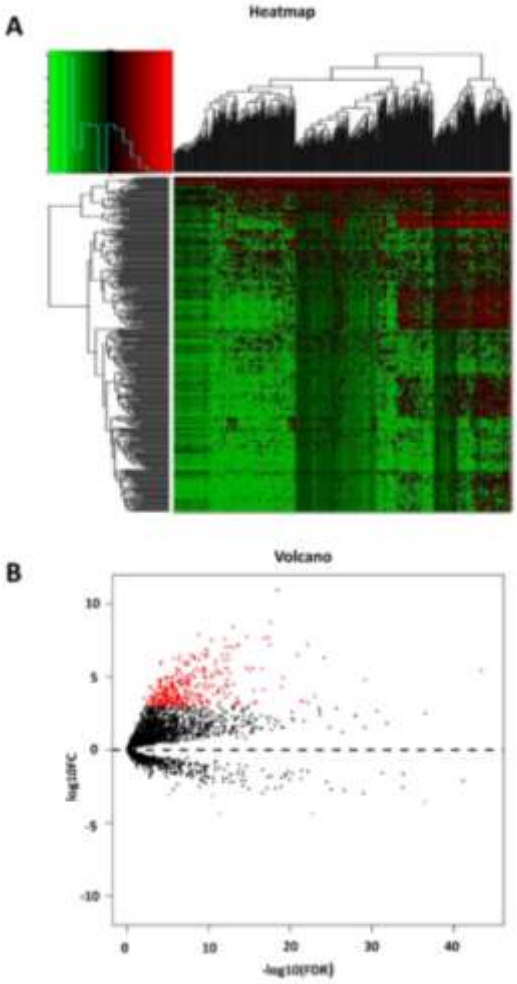

Figure 1. Heatmap and volcano plot showing differentially expressed lncRNAs of HCC compared with paracancerous samples. (A) Heatmap exhibits the expression of lncRNAs in the matrix. Those samples and genes with similar expression states were clustered together. (B) Volcano plots, the red dots indicate upregulated lncRNAs and green dots indicate downregulated lncRNAs, whose $|\log_2 FC|<3$ and FDR<0.05. The black dot marks the no-significant-changed lncRNAs. The X axis represents $\log_2 FC$ and the Y axis is the value of $\log_{10} FDR$.

To investigate which lncRNAs were associated with important clinical indicators, we performed weight co-expression analysis. First, we determined the DEmRNA by using EdgeR (Figure S1A and Figure S1B). Afterward, 8851 DEmRNAs and all 3099 lnRNAs were integrated into a matrix, which was used to perform WGCNA. Based on the characteristics of the data, β=4 was selected as the appropriate soft ‐ thresholding value to construct a scale-free network (Figure 2A). As a result, these genes were divided into 18 modules (Figures 2B and 2C). Combining the clinical data of patients, we calculated the correlation between each module and clinical information. What we were most concerned about in this clinical information were the

clinical stage, survival time, and clinical grade. Interestingly, the blue and turquoise modules were positively correlated with stage and grade, and negatively correlated with survival time (Figure 2D). In other words, as the gene expression levels of these 2 modules increased, the patient's clinical stage and clinical grade increased and survival time decreased. Therefore, the lncRNAs in these two modules are potential biomarkers for HCC.

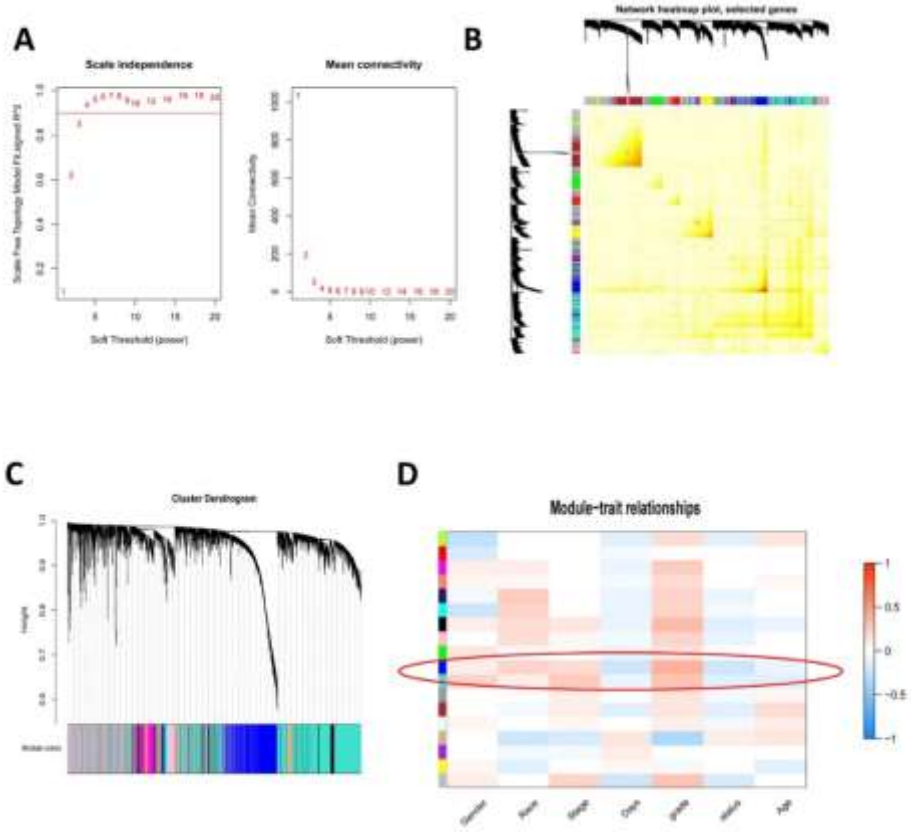

Figure 2. (A) Selection of soft thresholds in WGCNA. 4 was selected as a soft threshold. (B) Network heatmap plot to visualize the genetic correlation within the modules. The results show that the genes of the same modules tend to cluster together, which justifies the rationality of the module division. (C) Tree diagram represents the distant relationship between genes and genes. The similar genes are clustered into the consent module. (D) Module-clinical features correlation analysis. Red means positive correlation, blue means negative correlation and The shades of color represent the degree of correlation. For a cancer research, the most important clinical features are clinical stage, survival time, and clinical grade.

To further study the relationship between lncRNA expression and patient survival, we performed cox regression analysis (proportional hazards analysis) of all lncRNAs. Those lncRNAs with P values <0.01 and HR values >1 were considered significant. So far, we had three lncRNA sets, including DElncRNAs, the lncRNAs in blue and turquoise modules, and the lncRNAs significantly associated with survival time. We analyzed these 3 sets using Wayne analysis. The results showed the presence of 10 lncRNAs in these three sets simultaneously (Figure S2). These 10 lnRNAs were MIR137HG, BX322234.2, C10orf91, LINC02154, LINC01096, PICSAR, AC090921.1, AP003469.2, AP003469.2 and LINC01559.1 Their information in univariate cox regression are shown in Table 1. To further demonstrate the importance of these 10 genes, we performed a multivariate cox regression model with the expression levels of these 10 lncRNAs as variables. According to the regression results, we divided patients into a high-risk group (n=179) and a low-risk group (n=179). The results showed that the survival time of high-risk patients are significantly lower than those of low-risk patients (P < 0.001, Figure 3A). Moreover, ROC analysis showed that AUC (Area Under Curve) was 0.787. These indicated that these lncRNAs are potential HCC biomarkers, and their high expression levels have strong biological significance in the development of HCC.

Table 1. Univariate cox regression of 10 lncRNAs shows significant correlation with survival. Hazard ratio (HR), HR = risk function $h_1(t)$ of the exposed group / risk function $h_2(t)$ of the non-exposed group, t refers to the same time point. HR > 1 means the mortality risk of exposed group is higher than non-exposed group.

Gene	HR	P-value
MIR137HG	1.290015508	0.000179584
BX322234.2	1.206519509	0.002257706
C10orf91	1.188059694	4.79E-06
LINC02154	1.156755704	0.002126805
LINC01096	1.177976544	0.008431249
PICSAR	1.270247912	5.44E-06
AC090921.1	1.176592219	0.007129329
AP003469.2	1.180233484	0.00034846
LINC01224	1.13121042	0.000504966
LINC01559	1.105267388	0.002786875

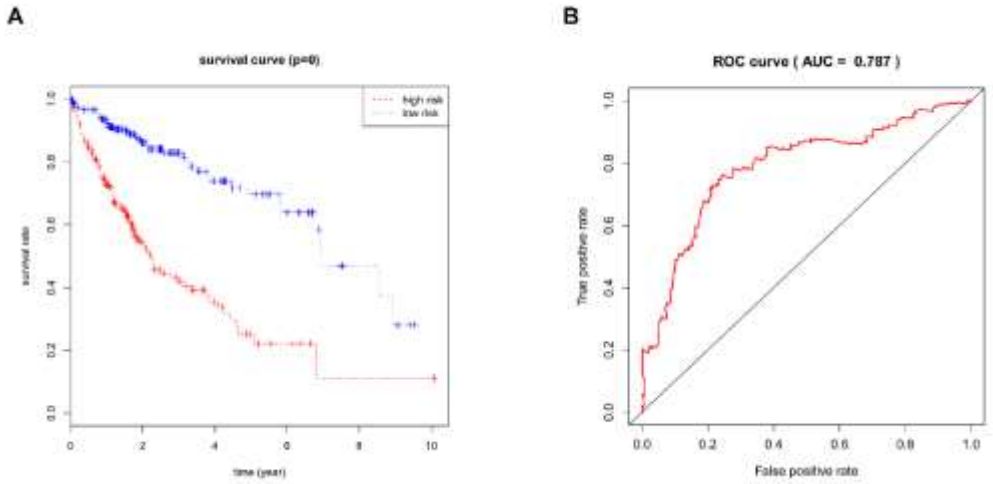

Figure 3. Kaplan-Meier (KM) method and receiver operating characteristic (ROC) determine the sensitivity and specificity of 10 lncRNA multivariate cox models. (A) KM plot shows that the survival time of the high-risk group is significantly lower than thatc of the low-risk group (P < 0.001). (B) ROC curve shows that the overall prediction accuracy of the model has reached 0.787, which proved that the results of this analysis are reliable.

To explore the function of these 10 lncRNAs, we needed to construct their co-expression networks with mRNA. In the previous WGCNA, we obtained the co-expression relationship of genes in the blue and turquoise modules. We selected the top 30 genes with the largest correlation coefficients of lncRNA for co-expression network construction. Of the 10 lncRNAs, 3 of them (LINC01559, LINC01224, MIR137HG) were in the blue module and 7 (BX322234.2, C10orf91, LINC02154, LINC01096, PICSAR, AC090921.1, AP003469.2) were in the turquoise module. Afterward, two co-expression networks were constructed (Figures 4A and 4C). The mRNAs in the co-expression network were likely to be positively regulated by lncRNAs in various ways. Thus, the functions of these mRNAs were the underlying mechanism of lncRNAs. We separately performed enrichment analysis of the mRNAs in the two networks. The most enriched functions and pathways were visualized based on the FDR value. The results seemed to indicate that the main function of the turquoise module was to enhance cellular metabolic capacity (Figure 4B), whereas the blue module primarily enhanced cell proliferation (Figure 4D).

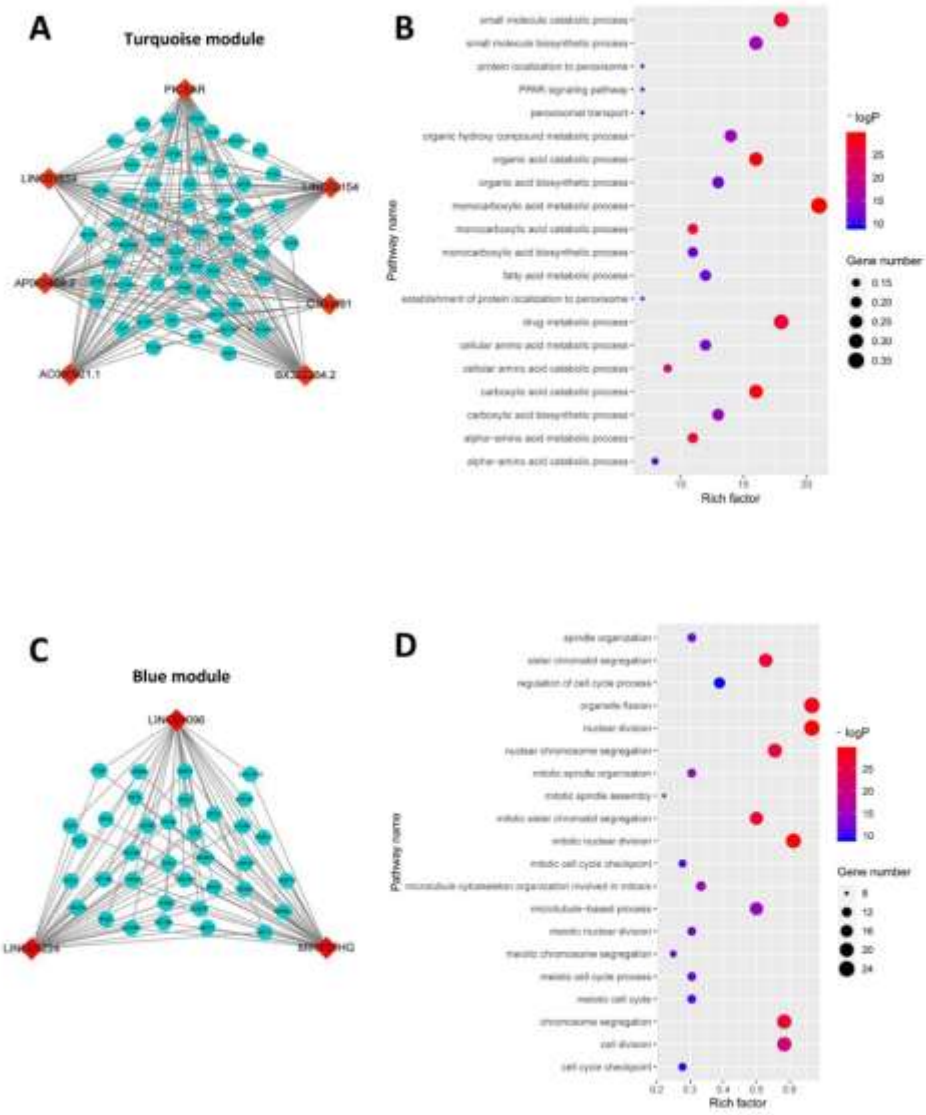

Figure 4. (A) Co-expression network diagram of 7 lncRNAs in the turquoise module. The red nodes are lncRNAs, and the green ones are mRNAs. (B) Gene enrichment analysis (GO and KEGG) of the mRNA in the turquoise module co-expression network. We showed all results with P less than 0.05. Rich factor = the number of mRNAs enriched / the total number of mRNA in network.(C) Co-expression network diagram of 3 lincRNAs in the blue module. The red nodes are lncRNAs, and the green ones are mRNAs. (D) Gene enrichment analysis (GO and KEGG) of the mRNAs in the blue module co-expression network. We also showed all results with P less than 0.05.

DISCUSSION

HCC is a worldwide disease with a survival rate of less than 5 years due to clinical conditions, such as easy transfer and complex cirrhosis of the liver [16]. Exploring the effective detection of biological markers of HCC is conducive to improving the understanding of its pathogenesis. Recently, research has proved the essential roles of lncRNAs in the pathogenesis and progression of HCC [17,18]. In our study, we screened key lncRNAs related to HCC through the TCGA database, and helped us elucidate the molecular mechanisms of HCC at the genomic level.

Recent studies have shown that lncRNAs can interact with other cellular macromolecules, including DNA, proteins, and RNA, and induce many important phenotypes in diverse cancers[19]. Our research is based on TCGA data and bioinformatics and sought to analyze the differential expression of lncRNA (DElncRNA). The RNA expression spectrum data from 374 tumor patients and 50 normal patients were downloaded from the TCGA project. The abnormal expression levels of 387 DElncRNAs were identified from a total of 3099 lncRNAs. Finally, our study focused on the high expression of lncRNA.

In recent years, more lncRNAs are being found and are closely related to the development and prognosis of HCC[16, 20]. lncRNA HULC was the first lncRNA-regulating gene expression that was reported in patients

with HCC [21]. lncRNA ANRIL has been found to be significantly up-regulated when regulating the proliferation ability in HCC tissues[22]. Upregulation of long noncoding RNA ZEB1-AS1 promotes tumor metastasis and predicts poor prognosis in HCC [23]. Long noncoding RNA CCAT1 promotes HCC progression by functioning as a sponge [24]. A novel lncRNA, TCONS_00006195, represses HCC progression by inhibiting enzymatic activity of ENO1[25]. Many lncRNAs have been proven to have key roles in the occurrence and development of HCC, and these will help us further understand the pathogenesis of liver cancer. In our study, 18 modules were divided based on WGCNA, and 2 of the 18 modules were positively correlated with stage and grade, and negatively correlated with survival time. Finally, 10 lncRNAs were found based on patient survival, namely, MIR137 host gene (MIR137HG), BX322234.2, long intergenic non-protein coding RNA 2870 (C10orf91), long intergenic non-protein coding RNA 2154 (LINC02154), long intergenic non-protein coding RNA 1096 (LINC01096), P38 inhibited cutaneous squamous cell carcinoma associated lincRNA (PICSAR), AC090921.1, AP003469.2, AP003469.2, and long intergenic non-protein coding RNA 1559 (LINC01559). Most of these lncRNAs relate closely with cancer. MIR137HG is correlated with the overall survival of patients with muscle-invasive bladder cancer[26]. C10orf91 is closely related with overall survival non-small cell lung cancer[27]. LINC02154 is risk factor for laryngeal cancer. LINC01559 accelerates pancreatic cancer cell proliferation and migration[28]. The co-expression network and function of these 10 lncRNAs were analyzed. Their main functions are the enhancement of cellular metabolic capacity and cell proliferation. These lncRNAs may act as competing endogenous RNA to crosstalk with mRNAs and miRNAs. The lncRNA-miRNA-mRNA networks play an important role in cellular metabolic capacity and cell proliferation.

However, the study was limited to bioinformatics results, and the results lack “wet-lab” verification. The further research will involve testing in cancer cell lines and animal models.

CONCLUSION

In conclusion, combining multiple reliable data platforms with existing knowledge is effective in explaining clinical problems. The key lncRNA confirmed by experiments can be used as new biomarkers for the prognosis of HCC. Moreover, targeted drugs can be designed based on the key lncRNAs and key signaling pathways to obtain the best cancer treatment. Therefore, the studying lncRNA will effectively guide us in comprehending the HCC mechanism.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Eggert, Wolter, Ji, Ma, Yevsa, Klotz, et al. Distinct Functions of Senescence-Associated Immune Responses in Liver Tumor Surveillance and Tumor Progression. Cancer Cell. 2016; 30(4):533-47.
2. Koduru, Leberfinger, Kawasawa, Mahajan, Gusani, Sanyal, et al. Non-coding RNAs in Various Stages of Liver Disease Leading to Hepatocellular Carcinoma: Differential Expression of miRNAs, piRNAs, lncRNAs, circRNAs, and sno/mt-RNAs. Sci Rep. 2018; 8(1):7967.
3. Okrah, Tarighat, Liu, Koeppen, Wagle, Cheng, et al. Transcriptomic analysis of hepatocellular carcinoma reveals molecular features of disease progression and tumor immune biology. NPJ Precis Oncol. 2018;2:25.
4. Bray, Ferlay, Soerjomataram, Siegel, Torre and Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68(6):394-424.
5. Bosch, Ribes, Dã-Az and Clã©Ries. Primary liver cancer: worldwide incidence and trends. Gastroenterology. 2004; 127(5):S5-S16.
6. Marengo, Rosso and Bugianesi. Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. Annu Rev Med. 2016; 67:103-17.
7. Yao, Ma, Xue, Wang, Li, Liu, et al. Knockdown of long non-coding RNA XIST exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR-152. Cancer Lett. 2015; 359(1):75-86.
8. Blanco, Jimbo, Wulfkuhle, Gallagher, Deng, Enyenihi, et al. The mRNA-binding protein HuR promotes hypoxia-induced chemoresistance through posttranscriptional regulation of the proto-oncogene PIM1 in pancreatic cancer cells. Oncogene. 2016; 35(19):2529-41.
9. Liu, Wu and Cao. Abstract 3954: lincRNA-UFC1 facilitates cell proliferation and tumor growth in hepatocellular carcinoma by upregulating HuR/β-catenin expression. Cancer Research. 2015; 75(15 Supplement):3954.
10. Fu, Zhu, Qin, Zhang, Lin, Ding, et al. Linc00210 drives Wnt/β-catenin signaling activation and liver tumor progression through CTNNBIP1-dependent manner. Molecular Cancer. 2018;17(1):73.
11. Li and Liu. Long non-coding RNA AB209630 suppresses cell proliferation and metastasis in human hepatocellular carcinoma. Experimental & Therapeutic Medicine. 2017; 14(4):3419-24.

12. Li, Yan, Yun, Fu, Tang, Rui, et al. The long non-coding RNA TP73-AS1 modulates HCC cell proliferation through miR-200a-dependent HMGB1/RAGE regulation. Journal of Experimental & Clinical Cancer Research Cr. 2017; 36(1):51.
13. Liu, Lu, Xiao, Jiang, Qu and Ni. Long non-coding RNA SNHG20 predicts a poor prognosis for HCC and promotes cell invasion by regulating the epithelial-to-mesenchymal transition. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2017; 89:857.
14. Luca, Matter, Salvatore, Leila, Christian, Alfredo, et al. Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. Hepatology. 2014; 59(3):911-23.
15. Yang, Zhou, Wu, Lai, Xie, Zhang, et al. Overexpression of Long Non-coding RNA HOTAIR Predicts Tumor Recurrence in Hepatocellular Carcinoma Patients Following Liver Transplantation. Annals of Surgical Oncology. 2011; 18(5):1243-50.
16. Wang, Pu, Yao, Lu and Deng. Four long noncoding RNAs as potential prognostic biomarkers for hepatocellular carcinoma. J Cell Physiol. 2018;
17. Guttman and Rinn. Modular regulatory principles of large non-coding RNAs. Nature. 2012; 482(7385):339-46.
18. Peng, Yuan, Zhang, Peng, Zhang, Pan, et al. The emergence of long non-coding RNAs in hepatocellular carcinoma: an update. J Cancer. 2018; 9(14):2549-58.
19. Hanahan and Weinberg. Hallmarks of cancer: the next generation. Cell. 2011; 144(5):646-74.
20. Ponting, Oliver and Reik. Evolution and functions of long noncoding RNAs. Cell. 2009; 136(4):629-41.
21. Tschernatsch, Guelly and Moustafa. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology. 2007; 132(1):330-42.
22. Huang, Chen, Qi, Xia, Sun, Xu, et al. Long non-coding RNA ANRIL is upregulated in hepatocellular carcinoma and regulates cell apoptosis by epigenetic silencing of KLF2. J Hematol Oncol. 2015; 8:50.
23. Li, Xie, Shen, Cheng, Shi, Wu, et al. Upregulation of long noncoding RNA ZEB1-AS1 promotes tumor metastasis and predicts poor prognosis in hepatocellular carcinoma. Oncogene. 2016; 35(12):1575-84.
24. Deng, Yang, Xu and Zhang. Long noncoding RNA CCAT1 promotes hepatocellular carcinoma progression by functioning as let-7 sponge. Journal of Experimental & Clinical Cancer Research. 2015; 34(1):1-10.
25. Galluzzi, Vitale, Aaronson, Abrams, Adam, Agostinis, et al. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. Cell Death Differ. 2018; 25(3):486-541.
26. Lyu, Xiang, Zhu, Huang, Yuan and Zhang. Integrative analysis of the lncRNA-associated ceRNA network reveals lncRNAs as potential prognostic biomarkers in human muscle-invasive bladder cancer. Cancer Manag Res. 2019; 11:6061-77.
27. Wang, Yin, Zhang, Zheng, Yang, Zhang, et al. The construction and analysis of the aberrant lncRNA-miRNA-mRNA network in non-small cell lung cancer. J Thorac Dis. 2019; 11(5):1772-8.
28. Lou, Zhao, Gu, Li, Tang, Wu, et al. LINC01559 accelerates pancreatic cancer cell proliferation and migration through YAP-mediated pathway. J Cell Physiol. 2020; 235(4):3928-38.