Identification of Potential Diagnostic and Prognostic Pseudogenes in Hepatocellular Carcinoma Based on Pseudogene-miRNA-mRNA Competitive Network

Lijun Yan* Chaosen Yue* Yingchen Xu Xincen Jiang Lijun Zhang Jixiang Wu

* Lijun Yan and Chaosen Yue contributed equally to this work

Corresponding Authors: Lijun Zhang, e-mail: tzhanglijun2019@126.com, Jixiang Wu, e-mail: trwjixiang2019@126.com

Source of support: This research was funded by the Beijing Hospitals Authority Youth Programme (code: QML20180203) and the Beijing Natural Science Foundation (code: 7194248)

Background: It is widely known that hepatocellular carcinoma (HCC) has high rates of morbidity and mortality. A large number of studies have indicated that pseudogenes have an important effect on the carcinogenesis of HCC. Pseudogenes can play a role through the ceRNA network. There have been numerous studies on lncRNA-miRNA-mRNA and circRNA-miRNA-mRNA networks. However, the pseudogene-miRNA-mRNA network in HCC has rarely been researched or reported on.

Material/Methods: The Cancer Genome Atlas (TCGA) database was researched and differences between selected genes were studied. A pseudogene-miRNA-mRNA network was then constructed and clustering of pseudogenes was studied. The diagnostic value of the selected pseudogenes, their functions, and pathways were investigated using available databases to understand their possible pathogenic mechanism in HCC. The protein-protein interaction network of target genes was found and the top 10 hub genes were identified. Expression of hub genes in HCC tissues was then detected by RT-qPCR.

Results: By analyzing the gene difference and clinical data of HCC, we constructed a ceRNA network composed of 4 pseudogenes, 8 miRNAs, and 30 mRNAs. The pseudogenes AP000769.1, KRT16P1, KRT16P3, and RPLP0P2 were all correlated with the diagnosis and prognosis of HCC. Functional analyses through the Kyoto Encyclopedia of Genes and Genomes and the Gene Ontology databases indicated that pseudogenes can affect the physiological process of HCC through the p53 pathway. The top 10 hub genes identified were all highly expressed in HCC tissues and affected the patient survival rate.

Conclusions: In this study, 4 pseudogenes related to the diagnosis and prognosis of liver cancer were found through the construction of a ceRNA network. These 4 pseudogenes might constitute new therapeutic targets for liver cancer patients.

MeSH Keywords: Carcinoma, Hepatocellular • Molecular Targeted Therapy • Pseudogenes

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/921895
Background

An ever-growing research community focuses on hepatocellular carcinoma (HCC). However, the pathogenesis of HCC is still not very clear and there is still no effective treatment for the disease. Although there are several treatment options, such as surgical resection, radiofrequency ablation, arterial chemobolization, and liver transplantation, survival time of HCC patients is still rather short and the recurrence rate is still high [1]. Finding potential diagnostic biomarkers or effective therapeutic targets is a top priority for research.

In the mammalian genome, only 1–2% of the genes perform protein-coding functions [2,3]. The remaining transcribed non-coding RNAs (ncRNAs) can be divided into short ncRNAs (miRNAs, siRNAs, snoRNAs, rRNAs, tRNAs, and piRNAs) and long ncRNAs (lincRNAs, antisense RNAs, pseudogenes, and circRNAs) [4]. Identification of long ncRNA (IncRNA) [5], circRNA [6,7], and miRNA [8-12] has offered new opportunities in HCC treatments. Initially, pseudogenes were not considered to have transcriptional activity or protein-coding capacity, based on gene duplication, mutation, insertion of retrotransposon, and so on, but an increasing number of studies have shown that pseudogenes are involved in the development of tumors. For instance, research has confirmed that phosphatase and tensin homolog pseudogene 1 (PTENP1) is associated with esophageal squamous cell carcinoma [13], oral squamous cell carcinoma [14], head and neck squamous cell carcinoma [15], and melanoma [16]. Double homeobox A pseudogene 10 (DUXAP10) is also reported to be related to various cancers, including colorectal cancer [17], pancreatic cancer [18], and gastric cancer [19], whereas ferritin heavy chain (FTH1)-pseudogene-miRNA network is oncogenic in prostate cancer [20] and uveal melanoma cells [21]. The function and underlying mechanisms of pseudogenes largely remains to be determined. Pseudogenes can affect gene expression through the RNAi pathway [22–25]. Pseudogenes can also interact with RNA-binding protein (RBP) to reduce mRNA stability of the parent gene [26,27]. There is also a hypothesis that pseudogenes competitively bind to miRNA through ceRNA networks, causing abnormal expression of target genes, and leading to tumorigenesis. For instance, PTENP1 affects the expression of its parent gene PTEN by binding to miR-106b and miR-93 in gastric cancer [28]. Similarly, the tumor suppressor candidate-2 pseudogene (TUSC2P) promotes TUSC2 function by binding to multiple miRNAs [29].

Our study is the first to search for pseudogenes that might be significant for diagnosis and prognosis of hepatocellular carcinoma, based on the pseudogene-miRNA-mRNA network.

Material and Methods

Analysis of gene expression based on The Cancer Genome Atlas database

The Cancer Genome Atlas (TCGA) database contains abundant tumor types and plentiful gene data. It is possible to find different oncogenes and tumor suppressor genes via this database, assisting in understanding the mechanism of occurrence and development of cancers. We downloaded the entire gene data from TCGA database, selected mRNAs and pseudogenes, and then used R software to analyze their differences. |log2FC| > 2.0 was set as the cutoff criterion, and P<0.01 was regarded as significant. In the same way, using the TCGA database, we performed miRNA expression analysis, with the same cutoff and significance criteria.

Construction of ceRNA network

We used the miRcode database to predict the relationship between miRNAs and pseudogenes. The target prediction databases miRDB, miRTarBase, and TargetScan were then consulted to predict the target genes of miRNA. Finally, we used Cytoscape software to construct the pseudogene-miRNA-mRNA network.

Potential prognostic value of selected pseudogenes

Clinical data of HCC patients, including age, sex, survival time, and survival status, were downloaded from TCGA. The pseudogenes obtained from the ceRNA network were combined with the clinical data, and then clustered and analyzed by weighted correlation network analysis (WGCNA). Survival time and survival status of HCC patients were used to analyze the prognostic value of pseudogenes in the ceRNA network using R software. P<0.01 was considered to be significant, and pseudogenes with significant prognostic value were selected for further analyses.

Diagnostic analysis of potential pseudogenes

Receiver operating characteristic (ROC) analysis was performed using SPSS Statistics version 22 (IBM Corporation, Armonk, NY, USA) to sort out genes with diagnostic value for HCC patients. Area under the curve (AUC) greater than 0.5 was considered to have diagnostic significance. AUCs closer to 1 indicate the pseudogene has greater diagnostic significance. In our study, to find pseudogenes with higher diagnostic significance, pseudogenes with AUC greater than 0.75 were considered to have an obvious diagnostic significance.

Construction of a Sankey diagram

Origin 2019 software was used to construct a Sankey diagram of pseudogenes with diagnostic and prognostic value for HCC, along with their matched miRNAs and mRNAs.
Kyoto Encyclopedia of Genes and Genomes pathway and
gene ontology function analyses of target genes

The WEB-based Gene Set Analysis Toolkit (WebGestalt) was
used to study the Kyoto Encyclopedia of Genes and Genomes
(KEGG) pathway and gene ontology (GO) function of target
genes. The GO functions included biological process (BP), cel-
lular component (CC), and molecular function (MF).

Protein-protein interaction analysis and hub gene
screening

Protein-protein interaction (PPI) network prediction for target
genes was carried out using the String database. First, to con-
struct the PPI network picture, target genes were inserted into
the String database, and the interaction score was set to 0.150.
The resultant file was exported to Cytoscape software and the
top 10 hub genes were selected using the Cytohubba plugin.
Figure 1. Dysregulation of pseudogenes, miRNA, and mRNA in HCC patients, based on TCGA database. (A) 374 pseudogenes are upregulated and 34 are downregulated in HCC. (B) 123 miRNAs are upregulated and 3 are downregulated in HCC. (C) 1937 mRNAs are upregulated and 211 are downregulated in HCC. TCGA – The Cancer Genome Atlas; HCC – hepatocellular carcinoma.

Figure 2. Construction of ceRNA network. Pseudogenes-miRNA-mRNA network was constructed using Cytoscape software. In the presented network figure, diamonds represent pseudogenes, round rectangles represent miRNAs, and ellipses represent mRNAs. The interrelationships between genes are represented by connecting lines. Blue stands for downregulated genes and red for upregulated genes.
Survival analysis in relation to the hub genes

Survival time and status data in relation to HCC were downloaded from TCGA database and used to analyze the association between them and the selected hub genes. The significance level was set at $P<0.01$.

Quantitative real-time polymerase chain reaction (RT-qPCR)

All HCC and adjacent normal liver tissues used in the study were obtained from Beijing Tongren Hospital, which is affiliated with Capital Medical University. This study was approved by the hospital’s Ethics Committee. Total RNA was extracted from the tissues using TRIzol reagent (Vazyme, Nanjing, China). The GoScript™ reverse transcription kit (Promega, WI, USA) was used to synthesize the cDNA according to the manufacturer’s protocol. Results were illustrated using GraphPad Prism 8.

Statistical analysis

Statistical analysis of pseudogene expression and survival was conducted using SPSS 22.0 software (IBM Corporation, Armonk, NY, USA) and GraphPad Prism8 (GraphPad Software Inc., USA), respectively. Statistical differences between 2 groups were examined using the t test. The prognosis of HCC patients was analyzed using the Kaplan-Meier method. Differences were considered statistically significant at $P<0.05$.

Results

Dysregulation of pseudogenes, miRNA, and mRNA in HCC

We located 50 liver normal samples and 374 HCC samples with pseudogene and mRNA expression information in TCGA database. We also found 50 normal samples and 375 HCC samples with miRNA expression information. Through representation difference analysis, we found 374 pseudogenes, 123 miRNAs, and 1937 mRNAs that were upregulated, and 34 pseudogenes, 3 miRNAs, and 211 mRNAs that were downregulated. (Figure 1).

Construction of ceRNA network

The ceRNA network was constructed in Cytoscape using 118 pseudogenes, 16 miRNAs, and 35 mRNAs (Figure 2). Among these genes, 100 pseudogenes, 15 miRNAs, and 30 mRNAs were upregulated, and 18 pseudogenes, 1 miRNA, and 5 mRNAs were downregulated.

Prognostic and diagnostic value of pseudogenes

First, the pseudogenes in the ceRNA network were clustered using WGCNA. We found that some pseudogenes were related to age and survival of patients (Figure 3). Using the RStudio software package to analyze clinical data of HCC patients, 6 pseudogenes associated with survival were identified: AP000769.1,
Figure 4. Survival analysis with relation to the selected pseudogenes. Six pseudogenes showed great significance in relation to prognosis of HCC. (A) AP000769.1, (B) KRT16P1, (C) KRT16P3, (D) PSMC1P10, (E) RPLP0P2, (F) TPRXL. HCC – hepatocellular carcinoma.
Figure 5. Diagnostic potential analysis of selected pseudogenes. ROC analysis was used to estimate diagnostic value of the 6 potential pseudogenes. Four pseudogenes with AUC greater than 0.75 had distinct diagnostic significance. (A) AP000769.1, (B) KRT16P1, (C) KRT16P3, (D) RPLP0P2. The other 2 pseudogenes had AUC smaller than 0.75 and were considered to have no diagnostic significance for HCC. (E) PSMC1P10, (F) TPRXL. ROC – receiver operating characteristic; AUC – area under the ROC curve; HCC – hepatocellular carcinoma.
Table 1. Position and classification of the 4 identified pseudogenes.

<table>
<thead>
<tr>
<th>Pseudogene</th>
<th>Dysregulation</th>
<th>Genome location</th>
<th>Ensemble ID</th>
<th>Gene type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP000769.1</td>
<td>Up</td>
<td>chr11: 65,455,258-65,466,720</td>
<td>ENSG00000173727</td>
<td>Transcribed unprocessed pseudogene</td>
</tr>
<tr>
<td>RPLP0P2</td>
<td>Up</td>
<td>chr11: 61,615,036-61,639,449</td>
<td>ENSG00000243742</td>
<td>Transcribed processed pseudogene</td>
</tr>
<tr>
<td>KRT16P1</td>
<td>Down</td>
<td>chr17: 18,432,051-18,442,895</td>
<td>ENSG00000214856</td>
<td>Transcribed unprocessed pseudogene</td>
</tr>
<tr>
<td>KRT16P3</td>
<td>Down</td>
<td>chr17: 20,501,513-20,512,357</td>
<td>ENSG00000214822</td>
<td>Transcribed unprocessed pseudogene</td>
</tr>
</tbody>
</table>

Figure 6. Construction of Sankey diagram. Origin 2019 software was used to generate a Sankey diagram of the diagnostic and prognostic potential of the selected pseudogenes and their matched miRNAs and mRNAs in HCC. HCC – hepatocellular carcinoma.

KEGG pathway enrichment and GO functional annotation for target genes

To better understand how pseudogenes in the ceRNA network affect tumorigenesis and development of HCC, we conducted KEGG pathway enrichment and GO functional annotation of the target genes. Using KEGG pathway enrichment (Figure 7) we found that the pseudogenes might affect the cell cycle (Figure 8A) and p53 pathway (Figure 8B), which might lead to HCC. On the other hand, a large number of GO terms were identified (Figure 7). GO functional annotation showed that the 4 pseudogenes might affect biological regulation and cell proliferation in HCC.

PPI network analysis and hub genes

PPI network was constructed to explore the interactions between the target genes (Figure 9A). We then inserted the target genes into Cytoscape and used the Cytohubba plugin to find the top 10 hub genes (CCNB1, CCNE1, CDC25A, CEP55, CLSPN, E2F7, EZH2, KIF23, PBK, RRM2) (Figure 9B).

Survival analysis and expression of hub genes

Survival analysis of the 10 hub genes showed that all had prognostic significance for HCC patients (Figure 10; P<0.01). Using RT-qPCR, we analyzed the expression of the hub genes in HCC and normal adjacent liver tissues, and found that the hub genes were all highly expressed in HCC tissues (Figure 11; P<0.01).

Discussion

A growing number of studies have focused on HCC because of its high incidence and mortality rates. Published reports indicate that pseudogenes might play a role in the tumorigenesis of HCC [30–33].

The ceRNA regulatory mechanism was first put forward in 2011 [34]. Subsequent studies suggested that the ceRNA mechanism plays an important role in the progression and development of different cancers. Studies on lncRNA-miRNA-mRNA [35–38] and circRNA-miRNA-mRNA networks have been conducted. To the best of our knowledge, we are the first to focus on the pseudogene-miRNA-mRNA network to find pseudogenes.
Figure 7. KEGG pathway and GO function analyses of target genes. KEGG pathway and GO function were analyzed by the WebGestalt. (A) GO terms in BP, (B) GO terms in CC, (C) GO terms in MF, (D) volcano plot of KEGG pathway. KEGG – Kyoto Encyclopedia of Genes and Genomes; GO – gene ontology; WebGestalt – WEB-based Gene Set Analysis Toolkit; BP – biological process; CC – cellular component; MF – molecular function.
Figure 8. Potentially HCC-related KEGG pathways. (A) Cell cycle, (B) p53 signaling pathway HCC – hepatocellular carcinoma; KEGG – Kyoto Encyclopedia of Genes and Genomes.
related to the diagnosis and prognosis of HCC. We also assessed their potential pathogenic mechanism in the disease.

We first used the HCC-related dysregulated pseudogenes, miRNAs, and mRNAs obtained from the TCGA database to construct a ceRNA network. We then combined the obtained pseudogenes with clinical data of HCC patients and conducted pseudogenes cluster analysis. Survival and ROC analyses on the pseudogenes led us to identify 4 pseudogenes that had significant value for the diagnosis and prognosis of HCC patients. These included the upregulated AP000769.1 and RPLP0P2, and the downregulated KRT16P1 and KRT16P3. It is well known that each pseudogene can bind multiple miRNAs. Similarly, a single miRNA can bind multiple pseudogenes. In our study, we found that the pseudogene RPLP0P2 could bind to hsa-miR-519d, hsa-miR-424, hsa-miR-373, hsa-miR-372, hsa-miR-205, and hsa-miR-183. Both KRT16P1 and KRT16P3 could bind to hsa-miR-205. Furthermore, hsa-miR-519d [43], hsa-miR-424 [44], hsa-miR-373 [45], hsa-miR-372 [46], hsa-miR-217 [47], hsa-miR-205 [48], hsa-miR-183 [49], and hsa-miR-182 [50], which were included in our ceRNA network, have all been confirmed to play a role in the development of HCC.

The p53 pathway affects both the occurrence and development of HCC [51]. In our study, 30 mRNAs related to the pseudogenes were used to analyze the KEGG pathway and GO function to clarify the pathogenic mechanism associated with the 4 identified pseudogenes in HCC. The GO analysis results indicate that the 4 pseudogenes are related to the physiological process of HCC, including biological regulation, cell proliferation, and growth. KEGG pathway analysis showed that the p53 signaling pathway plays a role in HCC and affects the cell cycle. We consider that the pseudogenes might affect the physiological process of HCC through the p53 signaling pathway.

The top 10 hub genes were selected through Cytoscape. We analyzed patient survival with relation to the expression of each hub gene. The results indicate that all 10 hub genes (CCNB1, CCNE1, CDC25A, CEP55, CLSPN, E2F7, EZH2, KIF23, PBK, RRM2) are highly expressed in HCC patients, and all were found to be associated with HCC patient survival. The results further confirm that pseudogenes can reduce the degradation of target genes by competitive binding to miRNA.

**Conclusions**

In our study, based on construction of a ceRNA network, we found 4 pseudogenes (AP000769.1, RPLP0P2, KRT16P1, KRT16P3) with diagnostic, prognostic, and possibly therapeutic significance for HCC patients. Furthermore, we investigated the potential molecular mechanism of the 4 pseudogenes in HCC, some of which might become promising targets for liver cancer treatment. The 4 pseudogenes might prove to be novel biomarkers to guide tailored therapy for patients with HCC. However, more experiments are needed to verify our conclusions.

**Conflicts of interest**

None.
Figure 10. Survival analysis with relation to the selected target genes. All target genes showed great significance as prognostic markers for HCC (P<0.01). (A) CCNB1, (B) CCNE1, (C) CDC25A, (D) CEP55, (E) CLSPN, (F) E2F7, (G) EZH2, (H) KIF23, (I) PBK, and (J) RRM2. HCC – hepatocellular carcinoma.
Figure 11. Expression of target genes. All target genes were upregulated in HCC tissues. (A) CCNB1, (B) CCNE1, (C) CDC25A, (D) CEP55, (E) CLSPN, (F) E2F7, (G) EZH2, (H) KIF23, (I) PBK, (J) RRM2. (** P<0.01; *** P<0.001). HCC – hepatocellular carcinoma.

