Original Article

Construction of ceRNA network and identification of two differentially expressed circRNAs in hepatocellular carcinoma by bioinformatic analysis

Hongwei He1, Zhong Shen2, Qiyun Gu1, Rong Xie2

1Department of Hepatobiliary Surgery, Jinshan Branch of The Sixth People’s Hospital of Shanghai, China; 2Department of General Surgery, Jinshan Branch of The Sixth People’s Hospital of Shanghai, China

Received August 11, 2020; Accepted October 23, 2020; Epub November 1, 2020; Published November 15, 2020

Abstract: Covalently closed circular RNAs (circRNAs) display dysregulated expression in several types of cancer. However, their functions remain largely unclear. In this work, datasets GSE125469 and GSE128274 of hepatocellular carcinoma (HCC) were selected from Gene Expression Omnibus (GEO) database. To identify differentially expressed genes (DEGs) in HCC and adjacent tissues, we used R package DESeq for analysis. Then, 15 DEcircRNAs, 65 DEmiRNAs, and 2084 DEmRNAs were identified comparing HCC and normal tissues. Next, to predict the target relationship of circRNA-miRNA and miRNA-mRNA in DEGs, we use the databases CircInteractome and starBase v2.0 for analysis. Finally, the ceRNA network of circRNA-miRNA-mRNA was established by Cytoscape software based on 2 DEcircRNAs (hsa_circ_0007813 and hsa_circ_0089372), 2 DEmiRNAs, and 98 DEmRNAs. In addition, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of DEGs to explore the function of DEGs in HCC. Functional enrichment analyses indicated DEmRNAs might be associated with HCC occurrence and progression. In general, our research reveals an important role of ceRNA’s molecular mechanism in HCC.

Keywords: HCC, circRNA, DEGs, ceRNA

Introduction

Hepatocellular carcinoma (HCC) is the most common primary liver cancer, with high morbidity and high mortality worldwide. There were more than 850,000 incident cases and 800,000 deaths due to liver cancer globally in 2015, according to the Global Burden of Disease Study 2015 [1]. Increasing evidence from observational studies suggest that chronic alcohol consumption [2], diabetes mellitus [3], obesity [4], non-alcoholic fatty liver disease [5], autoimmune hepatitis [6] and genetic risk factors [7] are risk factors for HCC development. The main treatment of early-stage HCC is surgery followed by chemotherapy or chemoradiation [8-10]. When the tumor recurs in HCC patients, most of the patients die from the heterogeneity of the tumor and lack of personalized treatment. Therefore, we explored HCC from its ceRNA mechanism to provide more information for clinical treatment and molecular mechanism research.

The ceRNA hypothesis was proposed by Pier Paolo Pandolfi’s research team at Harvard Medical School in 2011, to explain the intrinsic mechanisms of ncRNA [11]. CircRNA is a type of non-coding RNA, which plays a role in biologic regulation primarily by gene regulation [12]. CircRNA has abundant binding sites for miRNAs and thus acts by absorption of miRNAs like a sponge [13] and regulates other RNAs at the pre-transcription and post-transcription level by base-pairing [14]. It can also regulate the activity of proteins by interacting with them [15]. Moreover, evidence shows that circRNA is involved in apoptosis, autophagy and cell proliferation, suggesting that circRNA plays an important part in human disease [16-18].

Increasing evidence suggests that abnormal expression of circRNA is responsible for pathogenesis of multiple cancer types, including bladder cancer [19], lung cancer [20], breast cancer [21], liver cancer [22, 23], gastric carcinoma [24, 25] and colon cancer [26]. In liver
A ceRNA network in HCC

cancer, circRNAs have been studied by Liu, et al. who confirmed that circRNA-5692 competitively interacted with miR-328-5p to enhance the expression of DAB2IP, thereby inhibiting the growth of xenograft HCC tumors in vivo [27]. In addition, Yu et al. demonstrated circRNA-104718 was significantly upregulated in HCC, and indicated circRNA-104718 acts as a ceRNA combined with miR-218-5p to promote the progression of HCC in vivo and in vitro [28]. Furthermore, Dong et al. found that SCD-circRNA 2 was overexpressed and was related to poor clinical prognosis in HCC patients [29]. However, many function of circRNAs remain unknown, and remain to be explored in HCC.

In this study, we aimed to identify and validate possible circRNA biologic function in HCC. RNA-seq re-analyses were used to find the DEcircRNAs, DEmiRNAs, and DEmRNAs of HCC tissues and adjacent normal tissues. We conducted bioinformatic analysis by establishing a ceRNA network to select candidate circRNAs associated with HCC. Our results show that data mining and integration are effective methods for discovering new differentially expressed circRNA in HCC.

Materials and methods

Collection and processing of gene expression profile data

The GSE125469 and GSE128274 datasets were obtained from the Gene Expression Omnibus (GEO) database. Dataset GSE125469 includes 3 pairs of HCC and their adjacent paired tissues, and dataset GSE128274 contains 4 HCC patients’ fresh tumor tissues and paired adjacent non-tumor tissues. The raw sequence data were preprocessed and trimmed using Trimmomatic (version: 0.39) software. Clean reads were aligned against reference genome GRCh37 obtained from Ensembl using STAR (version: 2.7.3a). The resulting SAM files were further processed with HTSeq (version 0.12.4) to quantify read counts of protein-coding genes. CIRCexplorer2 was used to identify circRNA. The transcripts with reads greater than 3 were identified as candidate circRNAs. MiRNAs were quality trimmed using the above criterion and clean reads were subsequently aligned to mirbase to get the read counts.

Identification of DEcircRNAs, DEmRNAs and DEmiRNAs

The R package DESeq (version: 1.38.0) was employed to detect DEcircRNAs, DEmRNAs, and DEmiRNAs with thresholds of \( \log_{2} \text{FC} \geq 1 \) and \( P \)-value \( \leq 0.05 \). False discovery rate (FDR) was used to correct the statistical significance of all p-values. For the selected DEGs, we generated heatmaps using the heatmap plus packages in R software.

CeRNA network construction

In this study, we used the CircInteractome (https://circinteractome.nia.nih.gov/index.html) database to predict the miRNA binding sites of DEcircRNAs. Based on DEmiRNAs, we further screened target miRNAs. Then, the interactions between miRNA and mRNA were predicted by starBase v2.0 database (https://www.lncrnablog.com/tag/starbase-v2-0/). The Cancer-Specific CircRNA Database (http://gb.whu.edu.cn/CSCD/) was used to predict miRNA binding sites (MREs), RNA binding protein (RBP), and open reading frame (ORF). According to the circRNA-miRNA and mRNA-miRNA interactions, the ceRNA network of circRNA-miRNA-mRNA was structured by Cytoscape 3.8.0 software (https://cytoscape.org/).

Functional and pathway enrichment analysis

In order to determine the underlying biologic function of the DEmiRNAs in the ceRNA network during HCC tumorigenesis, functional enrichment analysis among Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was performed using the ClusterProfiler package of R software. \( P \)-value < 0.05 was considered significant, unless specifically indicated.

Statistical analysis

Receiver operating characteristics (ROC) analysis reflected a comprehensive index of sensitivity and specificity continuous variables. The true positive rate was taken as the ordinate, and the false positive rate was taken as the abscissa. The area under the curve (AUC) was used to determine the diagnostic performance. AUC was an indicator of diagnostic efficacy, and its range of value from 0.5 and 1.0. The value closer the AUC is to 1.0, the better the diagnos-
A ceRNA network in HCC

The analysis of ROC was performed by Statistical Product and Service Solutions (SPSS, version 24.0), and P-value ≤ 0.05 was considered significant.

**Results**

**DEcircRNAs, DEmRNAs, and DEmiRNAs in HCC**

A total of 95 and 249 the DEcircRNAs were identified from GSE125469 and GSE128274 datasets, respectively (Figure 1A, 1B). Venn analysis showed that 17 DEcircRNAs were screened out in the two datasets (Figure 1C). Among them, 15 genes presented identical expression trends in the two datasets, consisting of 12 upregulated genes and 3 downregulated genes in HCC tissue compared to normal liver tissue (Figure 1D). 3990 and 4456 DEm-RNAs were selected from GSE125469 and GSE128274 datasets respectively (Figure 2A, 2B). 2200 DEmRNAs were screened out in the two datasets (Figure 2C). There were 2084 DEmRNAs that showed the same expression trend in the two datasets, consisting of 997 upregulated DEmRNAs and 1087 downregulated DEmRNAs in HCC tissues (Figure 2E). To further establish a circRNA-miRNA-mRNA ceRNA network, we also analyzed miRNA expression profiles of HCC in a GSE128274 dataset. As a result, 65 DEmiRNAs were identified that contained 43 upregulated and 22 downregulated (Figure 2D).

**Construction of the ceRNA Network in HCC**

To better understand the role of DEcircRNAs in HCC and to further elucidate the interaction among the DEcircRNAs, DEmRNAs, and DEMiRNAs, the ceRNA network of circRNA-miRNA-mRNA was constructed (Figure 3A). In the ceRNA network including two subnetworks, there were 2 DEcircRNAs, 2 DEMiRNAs, and 98 DEmRNAs. The result proved that hsa_circ_0007813 regulated 80 DEmRNAs high expressions by competitively combined with hsa_miR_139-5p, and hsa_circ_0089372 downregulate 18 DEmRNAs expressions by competitive bonding with hsa_miR_346. The basic characteristics of the hsa_circ_0007813 and hsa_circ_0089372 were exhibited in Table 1. The basic structural patterns included MRE (miRNA response element), RBP (RNA binding protein) and ORF (open reading frame) of the two circRNAs, displayed in Figure 3B, 3C. Then, we performed the expression of the hsa_circRNA_0007813 (Figures S1A and S2A) and hsa_circ_0089372 (Figures S1B and S2B) in datasets GSE125469 and GSE128274, respectively. We found has_circ_0007813 was up-regulated in HCC, but the has_circ_0089372 was down-regulated. Also, we analyzed the ROC of the two circRNAs in GSE125469 and GSE128274 datasets respectively. Hsa_circRNA_0007813 showed an AUC of 0.667 and 0.875, P < 0.05 (Figures S1C and S2C). Has_circ_0089372 showed an AUC of 1.00 and 1.00, P < 0.05 (Figures S1D and S2D).

**Functional and pathway enrichment analyses**

GO and KEGG analyses were carried out to detect the functional characteristics of DEmRNAs in the two subnetworks, and these upregulated DEmRNAs were mainly enriched. The top 10 highly enriched GO terms of biological process (BP), molecular function (MF), and cellular component (CC) are shown in Figure 4A-C. The most enriched GO terms that “organelle fission”, “nuclear division”, “chromosome segregation” and “mitotic nuclear division” were significantly enriched in BP; in terms of molecular function (MF), DEmRNAs were mostly enriched in “protein C-terminus binding”, “protein-lysine N-methyltransferase activity”, “histone methyltransferase activity” and “histone-lysine N-methyltransferase activity”. Among the CC terms, the most DEmRNAs were enriched in “spindle”, “chromosomal region”, “condensed chromosome” and “chromosome, centromeric region” (P < 0.01). In addition, KEGG pathway analysis revealed that these DEmRNAs were significantly enriched in the “cell cycle”, “microRNAs in cancer”, and “cellular senescence” as shown in Figure 4D. Moreover, the upregulated DEmRNAs were associated with the “cell cycle”. A detailed map of the DEmRNAs for the “cell cycle” signaling pathway is exhibited in Figure 5.

**Discussion**

In the past decades, the functions of an abundance of circRNAs has been revealed using the upgrading of sequencing technology, the development of bioinformatics, and molecular biology [12, 30]. Circular RNA can regulate gene expression during transcription or post-transcription, and plays an important role in many
Figure 1. Identification of DEcircRNAs in HCC and normal tissue. A, B. Volcano plots represent DEcircRNAs in tumor and normal tissue. C. Venn analysis for the intersections between GSE125469 and GSE128274 datasets. D. Common DEcircRNAs in the two datasets.
A ceRNA network in HCC

Figure 2. Identification of DEmRNAs and DEmiRNAs in HCC and normal tissue. A, B. Volcano plots for DEmRNAs in HCC and normal tissue. C. Venn diagram represents DEmRNAs compared between two datasets GSE125469 and GSE128274. D. Volcano plot represent DEmiRNAs. E. The common DEmRNAs in GSE125469 and GSE128274 datasets.
A ceRNA network in HCC

human diseases. In addition, circRNAs have the potential to diagnose and prognose malignancies that serve as biomarkers [31-33]. In HCC, an increasing number of circRNAs, such as hsa_circ_0001649 [34], hsa_circ_0003998 [35], and cSMARCA5 [36], have been reported to play critical roles in development of tumor and guiding clinical diagnosis and treatment. In this study, we integrated circRNA, miRNA, and mRNA data comparing HCC tissues and adjacent tissues from the GEO database and established the ceRNA regulatory network.

CeRNA plays a crucial role in the occurrence and development of cancer by forming a complex ceRNA network [37, 38]. In the current study, we first gathered two RNA-seq high-throughput datasets from GEO database and screened out 15 DEcircRNAs, 65 DEMiRNAs, and 2084 mRNAs with R package DESeq in HCC. Then, we constructed a ceRNA network of circRNA-miRNA-mRNA by these DEGs, we found two key circRNAs that hsa_circ_0007813 and hsa_circ_0089372 may act as ceRNAs respectively to capture hsa_miR_139-5p and hsa_miR_346, and subsequently regulate the 98 genes expression. Prior study indicated that miR-139-5p plays a pivotal role in lung cancer by targeting oncogenic c-Met, inhibiting cell proliferation, metastasis, and promoting apoptosis [39]. Furthermore, lncRNA NEAT1 can regulate the expression of TGF-β1 by competitively binding hsa-mir-139-5p in HCC [40]. Meanwhile, there were also some previous studies on hsa_miR_346 in cancers. For example, in the MHCC-97H and Hep3B cells, DGCR5 can inhibit the progress of HCC by adsorbing miR-346 [41]. In breast cancer, miR-346 may function as a carcinogenic miRNA, targeting SRCIN1 to mediate chemotherapy sensitivity of docetaxel [42]. Nevertheless, there are no rel-

Figure 3. The view of circRNA-miRNA-mRNA network in HCC and circRNA structure chart. A. The network consists of 2 circRNAs, 2 miRNAs, and 98 mRNAs. Ellipse represents mRNA, rhombus represents circRNA and triangle denotes miRNA. The nodes highlighted in orange and green represent upregulation and downregulation, respectively. B, C. Structural patterns of hsa_circ_0089372 and hsa_circ_0007813.
Table 1. Essential characteristics of the two differently expressed circRNAs

<table>
<thead>
<tr>
<th>circRNA ID</th>
<th>Position</th>
<th>strand</th>
<th>genomic length</th>
<th>best transcript</th>
<th>gene symbol</th>
<th>Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa_circ_0007813</td>
<td>chr10: 12123470-12162266</td>
<td>+</td>
<td>38796</td>
<td>NM_018706</td>
<td>DHTKD1</td>
<td>up</td>
</tr>
<tr>
<td>hsa_circ_0089372</td>
<td>chr9: 136302868-136303486</td>
<td>+</td>
<td>618</td>
<td>NM_139025</td>
<td>ADAMTS13</td>
<td>down</td>
</tr>
</tbody>
</table>

Figure 4. GO and KEGG enrichment analyses of DEmRNAs in the ceRNA network. GO terms included (A) biological process, (B) molecular function, (C) cellular component. (D) The enriched KEGG pathways. The vertical axis represents various GO terms, circle-dot represents the number of genes enriched, color represents the P-value.
Figure 5. DEmRNAs mapped with cell cycle signaling pathways. DEmRNAs involved in cell cycle signaling pathways are labeled in red. All the genes are upregulated.
event studies of hsa_circ_0007813 and hsa_circ_0089372.

To better understand the key circRNAs participating in the ceRNA network, we performed the GO and KEGG pathway analyses of DEmRNAs in the ceRNA network. The results of BP and CC displayed that the DEmRNAs were significantly enriched in activity of chromosome and organelle division, while the results of MF were involved in protein methyltransferase activity. Daniela et al. [43] indicated that tumor cells alter mitochondrial dynamics to prevent apoptosis, and adjust their bioenergetic and biosynthetic needs to support tumor proliferation, migration, and therapeutic resistance. Mengnuo et al. [44] found that methyltransferase-like 3 (METTL3) is frequently overregulated in human HCC and participates in the progress of HCC. Subsequently, the results of KEGG pathway analysis showed that DEmRNAs were significantly enriched in signaling pathways related to the cell cycle. The complexity of the cell cycle regulatory mechanism and the frequency of component disorders reflect the importance of unplanned division of malignant phenotypes [45]. In conclusion, functional enrichment analyses revealed that DEmRNAs could be involved in the development of HCC.

To sum up, we successfully applied bioinformatics analysis of plentiful samples in the GEO database to identify cancer-specific circRNAs in HCC. Our study could provide a new cue for diagnosis and treatment of HCC patients. However, this study has some shortcomings. First of all, the prognostic value of DEcircRNAs is unclear. In the next study, we will verify our findings and further explore the functions of these DEGs since our findings have not been verified by experiment.

Conclusions

In the current work, we constructed a ceRNA network of DEcircRNAs by bioinformatic analysis and identified two HCC-related circRNAs. In addition, the DEmRNAs regulated by two HCC-related circRNAs were analyzed by the GO function enrichment and KEGG pathway, which revealed that these genes related to the occurrence and progression of HCC. Finally, our article provided a new perspective and strategy in predicting and diagnosing HCC.

Disclosure of conflict of interest

None.

Address correspondence to: Rong Xie, Department of General Surgery, Jinshan Branch of The Sixth People’s Hospital of Shanghai, No. 147, Jian Kang Road, Jinshan District, Shanghai 201599, China. Tel: +86-13761131083; E-mail: drianyu@163.com

References


A ceRNA network in HCC


A ceRNA network in HCC

Figure S1. Expression level of hsa_circ_0007813 and hsa_circ_0089372 in dataset GSE125469 (A, B). The red bars represent the control group, and the blue bars represent the HCC group. ROC analysis of hsa_circ_0007813 and hsa_circ_0089372 for HCC in GSE125469 (C, D). *P < 0.05, ***P < 0.01.

Figure S1. Expression level of hsa_circ_0007813 and hsa_circ_0089372 in dataset GSE125469 (A, B). The red bars represent the control group, and the blue bars represent the HCC group. ROC analysis of hsa_circ_0007813 and hsa_circ_0089372 for HCC in GSE125469 (C, D). *P < 0.05, ***P < 0.01.
Figure S2. Expression level of has_circ_0007813 and has_circ_0089372 in dataset GSE128274 (A, B). The red bars represent the HCC group, and the blue bars represent the adjacent tissue group. ROC analysis of has_circ_0007813 and has_circ_0089372 for HCC in GSE128274 (C, D). *P < 0.05.