Identification of a long non-coding RNA-mediated competitive endogenous RNA network in hepatocellular carcinoma

HUI HE¹*, DI CHEN²*, SHIMENG CUI³*, HAILONG PIAO², HAIBO TANG¹, XUN WANG¹, PENG YE⁴ and SHI JIN¹

¹Department of Laparoscopic Surgery, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning 116000; ²CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023; ³Department of Hepatobiliary Surgery, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning 116000; ⁴Department of Urological Surgery, Liaoning Cancer Hospital and Institute, Shenyang, Liaoning 110042, P.R. China

Received December 2, 2018; Accepted March 26, 2019

DOI: 10.3892/or.2019.7181

Correspondence to: Professor Shi Jin, Department of Laparoscopic Surgery, The First Affiliated Hospital of Dalian Medical University, 193 Lianhe Road, Shahekou, Dalian, Liaoning 116000, P.R. China
E-mail: jinshi_1969@hotmail.com

*Contributed equally

Key words: IncRNA, ceRNA network, HCC, OIP5-AS1, topological analysis

Abstract. The present study was designed to identify the endogenous RNA regulatory networks involved in hepatocellular carcinoma (HCC) by bioinformatic analysis. Both miRNA interaction network-based correlation analysis and expression-based Spearman correlation coefficients were utilized to identify potential mRNA-IncRNA interactions. Then, a competitive endogenous (ce)RNA network was constructed from these interactions, and network topology and Gene Ontology enrichment analyses were conducted to mine potential functions of ceRNAs. In HCC samples, a ceRNA network was constructed. It was composed of 35,657 edges connecting 113 IncRNAs and 6,136 mRNAs which were differentially expressed in HCC and normal liver tissues. Meanwhile, a number of significantly positively correlated mRNA and lncRNA pairs in this ceRNA network were found to be consistently positively correlated in another independent dataset. To be noted, further analyses on the potential roles of ceRNAs demonstrated that various IncRNAs such as LINC00657, TUG1 and SNHG1 may play key roles in HCC by regulating protein phosphorylation or cell cycle pathways or influencing mRNAs. From the perspective that lncRNAs can function as ceRNAs, this study revealed that the interaction between IncRNAs, miRNAs and mRNAs may provide new insight for the diagnosis and treatment in the tumorigenesis of hepatocellular carcinoma.

Introduction

The liver is the metabolic repository and the largest internal organ in the human body. Metabolic or nutritional disorders can often lead to the tumorigenesis of hepatocellular carcinoma (HCC) (1,2). HCC remains one of the most deadly cancers in the world, particularly in China (3). MicroRNAs (miRNAs) regulate post-transcriptional gene expression and participate in the biological processes of a variety of diseases, including HCC (4). Recently, long non-coding RNAs (IncRNAs) are attracting widespread attention (5). These RNAs contain more than 200 nucleotides and do not encode any protein. With the rapid development of high-throughput sequencing technology, IncRNAs have been found to be involved in a large number of biological processes (6). In addition, recent studies have found that IncRNAs can form competitive endogenous RNAs (ceRNAs) through sponge adsorption of miRNAs to regulate mRNA expression. This plays an important role in the pathological development of tumors (7-12). For example, IncRNA HULC plays an important regulatory role in lung cancer through ceRNA (13). Currently, some related ceRNA databases have been developed (11,14,15), but research on ceRNAs is still relatively scarce in regards to their implications in HCC.

In the ceRNA network, there are multiple miRNA binding sites on each ceRNA, and the number of the same miRNA binding sites also differs; thus, there are multiple targets for each miRNA (16). This makes the ceRNA network a complex large-scale post-transcriptional regulatory network. Changing one node or edge in the network can affect the entire network (17,18). These promote ceRNA interactions among different signaling pathways (19). ceRNA network members are varied and complicated: mRNA, IncRNA and circRNA transcripts could interact with miRNAs involved in network regulation (20). Further exploration of the ceRNA network
will facilitate a better understanding of post-transcriptional regulation. In addition, the imbalance of the ceRNA network may lead to the occurrence of various diseases (21). Further study concerning the mechanisms of ceRNAs may expand our understanding of the pathogenesis of diseases.

In the present study, we constructed a ceRNA network and identified a number of the significant lncRNAs in HCC in a computational way. According to the edgeR algorithm, 7,334 mRNAs and 138 lncRNAs were detected as being differentially expressed between HCC and normal liver tissues. Based on these mRNAs and lncRNAs, we constructed a ceRNA network with 35,637 edges connecting 113 lncRNAs and 6,136 mRNAs by integrating both the miRNA interaction and expression-based mRNA-lncRNA correlations. Various connections in the network were also validated by another independent dataset. Further network and functional enrichment analyses may identify various hub lncRNAs in the network that may play key roles in HCC.

**Materials and methods**

**Gene expression profiles and clinical information.** RNASeq based gene expression data which included the expression of 20,531 genes in 372 HCC tissues and 50 normal liver tissues, were downloaded from TCGA (https://www.cancer.gov/tcga). The edgeR algorithm was used to identify differentially expressed genes between HCC and normal liver samples. The lncRNA expression profiles and lncRNA transcript sequences were separated from the lncRNA annotation, which was downloaded from the GENCODE database (v19) (22). In addition to the TCGA data, another independent dataset to verify the accuracy of the constructed ceRNA network was used. A public data set GSE62232 (23) was downloaded from the GEO database. It contained expression profile data from 81 HCC patients detected by the Affymetrix Human Genome U133 Plus 2.0 Array platform (Thermo Fisher Scientific, Inc., Waltham, MA, USA).

**Probe annotation.** The probe annotation sequences supported by Affymetrix (http://www.affymetrix.com/) were compared with human long non-coding transcript sequences and human coding transcript sequences from the GENCODE database (http://www.gencodegenes.org/) using the BLASTn tool separately. The sequence comparison results are filtered as follows: i) Removal of non-coding transcripts and transcript-encoding probes that were compared simultaneously; and ii) removal of probes that were compared with multiple transcripts.

**miRNA target information.** A total of 410,384 and 423,975 miRNA-mRNA interactions were downloaded from the miTarBase (http://mi tarbase.mbc.nctu.edu.tw/php/index.php) and StarBase (http://starbase.sysu.edu.cn/) databases, respectively. A total of 713,391 interactions remained after removing the duplicate interactions.

A total of 62,838 and 10,213 miRNA-lncRNA interactions were downloaded from the lncBase (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbase2%2FFindexperimental) and StarBase databases, respectively. A total of 68,773 interactions remained after removing the duplicate data. Among these, 18,972 interactions finally remained after GENCODE database filtration.

**ceRNA network construction**

**Acquisition of candidate ceRNA relationship based on miRNA interactions.** The potential ceRNA relationship for lncRNA-mRNA met two basic requirements. Firstly, the number of shared target miRNAs between lncRNA and mRNA was >3. Secondly, the hypergeometric test false discover rate (FDR) on the significance of the shared miRNAs between one mRNA and one lncRNA was <0.01. The hypergeometric test formula is as follows:

$$P = 1 - \sum_{t=0}^{r-1} \frac{\binom{m-t}{r} \binom{n-t}{m-n}}{\binom{m}{n}}$$

where, $m$, is the number of all miRNAs, $t$, represents the number of miRNAs that interact with mRNA, $n$, stands for the number of miRNAs that interact with lncRNA, and $r$, indicates the number of miRNAs that are shared by miRNAs and lncRNAs.

**Further filtration by expression-based correlations.** The spearman correlation coefficient between the lncRNA and mRNA was calculated for each potential lncRNA-mRNA pair. A correlation coefficient $>0$ and the corresponding FDR $<0.01$ were retained to construct the ceRNA network only.

**Validation of the ceRNA network by independent dataset.** Using probe re-annotation, microarray profiling data were obtained and the constructed ceRNA network was validated.

**Network analysis.** First, Cytoscape 3.1.1 (https://cytoscape.org/) was utilized to display the constructed ceRNA network, where the built-in Network Analyzer tool was used to analyze the topological properties of the network and calculate the degree, betweenness and closeness of the nodes in the network.

Second, the hub nodes in the ceRNA network were identified based on the above 3 node centrality measurements, and then the DAVID database (24) was applied for Gene Ontology (GO) function enrichment and functional annotations of these hub nodes. The R package 3.4.1 (https://www.r-project.org/) ‘go Profiles’ was used for result visualization.

**Results**

**ceRNA network construction.** Gene expression profiling data of 372 HCC samples and 50 normal liver tissues were downloaded from the TCGA database. Based on the edgeR algorithm, 7,919 differentially expressed genes were identified. Among them, 7,334 mRNAs and 138 lncRNAs were also annotated with at least one miRNA partner based on the collected miRNA-mRNA and miRNA-lncRNA interactions, and the basic nodes of the ceRNA network were utilized.

Based on these basic nodes, 370,256 miRNA-mRNA interactions and 5,504 miRNA-lncRNA interactions from the miRNA interaction information were extracted, of which 133,300 mRNA-lncRNA interactions shared $>3$ miRNA targets. Based on hypergeometric enrichment analysis, 10,204 mRNA-lncRNA interactions were with significantly shared (FDR <0.01) miRNAs, and they were taken as potential pairs for ceRNA network construction. Considering the competitive binding with miRNA, the mRNA and lncRNA
expression should show a positive correlation. Using Spearman rank correlation, 35,637 mRNA-lncRNA pairs had significantly positive correlations among the 10,2031 potential interactions, including 6,136 mRNAs and 113 lncRNA; these nodes and their interactions constituted the HCC ceRNA network (Fig. 1; Table SI).

Verification analysis of the constructed HCC ceRNA network.

The annotation data for lncRNA and mRNA based on the re-annotation of ‘BLASTn’ were obtained due to the lack of annotation data for the lncRNA chip in GEO database. First, the public data set GSE62232 which contained gene expression profiles for 81 HCC patients was downloaded from the GEO database. Furthermore, using the BLASTn algorithm, the platform probe sequences were aligned to non-coding and protein-coding transcripts respectively (see Materials and methods section). A total of 16,074 mRNA-probes and 6,273 lncRNA-probes were obtained by the previous filter condition introduced in Materials and methods. For mRNA/lncRNA which was matched by multiple probes, the average value of the detection probe was taken as the expression value, and finally the expression profile of 4,846 mRNA and 47 lncRNA were obtained. By mapping these mRNAs and lncRNAs to the ceRNA network we constructed, 13,419 mRNA-lncRNA pairs were obtained.

These mapped ceRNA pairs were also verified to display high correlations in expression level. The Spearman correlation coefficients between mRNA and lncRNA in these 13,419 ceRNAs pairs were calculated. Based on the results, 1,405 ceRNA relationships were significantly correlated (FDR <0.01), of which 1,378 ceRNAs showed a significant positive correlation (Fig. 2). Cumulative binominal distributions demonstrated that there was a significant positive correlation (P<1.0e-16) for non-randomized ceRNAs in the independent data set; this demonstrated the consistency of the obtained ceRNAs in different datasets.

Topological analysis of the network.

To further analyze the topological properties of the ceRNA networks, the degree, closeness and betweenness of the lncRNAs and mRNAs were calculated in the network. The distributions of lncRNAs and mRNAs are shown in Fig. 4A and B, and the average degree of a lncRNA node was significantly higher than that of a mRNA node. Using the mcode plugin in Cytoscape software, the ceRNA network was further mined to obtain 5 modules (Fig. 3), containing a total of 16 lncRNAs. The topological features of the nodes are ranked from large to small, and the genes in TOP10 in 3 dimensions are listed (Fig. 4C and D). Five lncRNAs were found in each dimension list and 3 lncRNAs were included in the mined ceRNA modules (LINC00657 in module 1, TUG1 in module 3, and SNHG1 in module 2).

Moreover, GO enrichment analysis was conducted on the mRNAs from the ceRNA network (David, FDR <0.05; Fig. 5; Table SII). For lncRNA LINC00657, significant biological pathways were found, including ‘protein phosphorylation’, ‘regulation of transcription’ and ‘epidermal growth factor receptor signaling pathway’. Significantly enriched biological pathways included ‘cell division’, ‘G2/M transition of mitotic cell cycle’ and ‘mitotic nuclear division’ for lncRNA TUG1 and SNHG1. These pathways have been
confirmed to be closely related to the occurrence and progression of HCC (25-30). In addition, Zhang et al reported that patients in the SNHG1 high-expression group had a worse recurrence-free survival (P=0.0073, log-rank test) and overall
survival (P=0.0068, log-rank test) than those in the SNHG1 low-expression group, indicating that SNHG1 may serve as a potential prognostic target (31).

In addition, the miR2Disease database (www.mir2disease.org/) was used to elucidate the relationship between miRNA expression and various diseases, including HCC (Fig. 6). Through the interaction between IncRNAs and miRNAs, an IncRNA-miRNA-disease network was constructed for the abovementioned 5 high-ranked IncRNAs. This showed that all 5 IncRNAs could regulate the occurrence of various diseases including HCC by affecting the expression of miRNAs.
Discussion

Competing endogenous RNAs (ceRNAs) may explain certain biological phenomena (autophagy and apoptosis, morphogenesis), and also can be involved in the inhibition of miRNA activity (32). As early as 2007, Ebert et al developed an miRNA sponge to inhibit the activity of specific miRNAs (33). Compared with the miRNA sponge, ceRNAs have the advantage of inhibiting many miRNAs and can play a role in network regulation by changing the type and number of miRNA binding sites (33). Tang et al developed a human ceRNA, known as a short tandem target mimic (STTM) that effectively inhibits many miRNAs (34,35). This suggests that ceRNA has good application prospects in the treatment of diseases.

The regulatory mechanisms of lncRNAs in cancer formation are diverse and complicated compared to miRNAs. Although a large number of studies have shown that lncRNAs play an important role in the development of tumors (36), the precise molecular mechanisms remain unclear. lncRNAs can regulate their downstream target genes through signaling, decoy, guide and scaffold action mode (37). A previous study found that lncRNAs can also serve as ceRNAs or miRNA sponge, through its microRNA response elements (MREs) competitive binding miRNAs, and inhibit the function and activity of miRNAs, thereby regulating the transcriptional level of miRNA target gene mRNA expression (38); and are involved in tumor proliferation, invasion, metastasis and angiogenesis and other biological behaviors (39). For example, IncRNA-NEAT1 was found to be upregulated in esophageal squamous cell carcinoma (ESCC), and was demonstrated to function as an endogenous sponge to downregulate miR-129, leading to its target mRNA CTBP2 depression. CTBP2 restoration overturned cellular proliferation and suppression of invasion regulated by NEAT1 depletion or miR-129 overexpression. This suggested that the lncRNA NEAT1/miR-129/CTBP2 axis regulates cell progression in ESCC (40). lncRNA -SNHG1 was found to be upregulated in osteosarcoma (OS) tissues and cell lines, and knockdown of SNHG1 was found to inhibit cell growth and metastasis in vitro and in vivo, and also showed better overall survival for OS patients. Additionally, SNHG1 increased the onco-gene NOB1 through sponging miR‑326 as a ceRNA, finally promoting cell growth, migration and invasion in OS. These findings uncovered that the SNHG1/miR‑326/NOB1 signaling axis plays a key role in OS progression also suggesting the potential application of SNHG1 and miR-326 as biomarkers in OS diagnosis and treatment (41).

The ceRNA hypothesis presents a new concept for the research of lncRNAs. With the development of genome-wide sequencing, especially gene chip and second-generation sequencing, more and more non-coding RNAs have been
shown to have special and important functions related to biological processes such as tumor formation, invasion and metastasis. IncRNAs expand the regulatory network through interactions with miRNAs to connect with other transcriptome members. In this study, LINC00657, TUG1, SNHG1, KCNQ1OT1 and OIP5-AS1 were ranked as important nodes in the ceRNA network by a series of topological features (Fig. 4C). A recent study by Liu et al found that knockdown of LINC00657 significantly inhibited the growth and progression of tumor cells and LINC00657 was found to be significantly upregulated in breast cancer, indicating that it exerts an oncogenic function in breast cancer (42). In another study, LINC00657 was found to be downregulated in HCC and a potential ceRNA regulatory network involving LINC00657 and mir-106a-5p was found to display action in the modulation of PTEN. These results may contribute to a better understanding of the role of LINC00657 and provide a new therapeutic target for HCC. TUG1 is a recently discovered proto-oncogene (43). Huang et al reported that TUG1 is upregulated in HCC and promotes cell growth by silencing the KLF2 gene (44). Zhang et al found that lncRNA-SNHG1 is a potential prognostic marker and therapeutic target (31), which was consistent with our findings. KCNQ1OT1 is an lncRNA gene located at the KCNQ1 locus and belongs to the ‘imprinted gene’ and only expresses the paternal allele. Its transcripts regulate the 15.5 centromere at p-terminal of chromosome 11 (45). Aberrant expression of imprinted genes elicits a variety of human diseases with complex mutations and phenotypic defects (46). The research concerning KCNQ1OT1 in regards to tumors is still in its infancy, yet a previous studies found that KCNQ1OT1 can promote the formation of HCC (47). OIP5-AS1 may have a ‘sponge’ function or serve as a ceRNA for RNA-binding protein HuR, which enhances cell proliferation. Competitive binding of HuR and miR-424 to OIP5-AS1 affects HuR binding to target mRNAs in HeLa cells, including those that encoded proliferative proteins (48). Another study showed that lncRNA OIP5-AS1 was downregulated and had a negative regulation effect on miR-410 expression in multiple myeloma MM tissues, thus promoting KLF10-mediated PTEN/AKT signaling in MM cells. Additionally, the OIP5-AS1-miR-410-KLF10/PTEN/AKT signaling axis was hypothesized to exert key functions in cell proliferation, cell cycle progression and apoptosis inhibition of MM and may represent a therapeutic target for MM patients (49). Currently, there is a limited number of cell function studies concerning OIP5-AS1 in the ceRNA network and we will carry out the above research.

The ceRNA hypothesis demonstrates a completely new mode of post-transcriptional regulation. Research has shown that complex organisms contain a high proportion of non-coding DNA. These non-coding DNA transcripts can participate in large-scale post-transcriptional regulation network, and become an integral part of life activities. However, how IncRNAs function by competing with ceRNAs has not been adequately investigated (47). Therefore, the establishment and application of bioinformatic research methods play an important role in the study of IncRNAs as a regulator of the ceRNA gene expression network. In the present study, based on the methods of bioinformatics, IncRNAs were obtained by different algorithms and the 5 IncRNAs that were found to function as ceRNAs in HCC were summarized, which provided an important basis of IncRNAs as the gene expression regulatory network of ceRNAs. Future research will investigate the biological function of these 5 IncRNAs, especially OIP5-AS1 in HCC, with the aim to reveal its ceRNA mechanism in the occurrence and development of HCC and provide a new class of molecules for tumor prediction and diagnosis. Overall, the ceRNA network was constructed and various significant IncRNAs in HCC were identified. The greatest utility of the ceRNA hypothesis may be to understand how ceRNA networks affect post-transcriptional regulation, and how mbalances in ceRNA networks may contribute to cancer. A large number of differentially expressed IncRNAs in HCC were identified which may be potential biomarkers for early diagnosis of HCC and targets for drug therapy, opening up new ideas for the clinical diagnosis and treatment of HCC.

Acknowledgements

Not applicable.

Funding

The present study was supported in part by the National Natural Science Foundation of China (NSFC 81702526).

Availability of data and materials

The datasets used during the present study are available from the corresponding author upon reasonable request.

Authors' contributions

HH and DC designed the study. HH, DC, HP, SC and HT collected and analyzed the data. HH, DC, XW, PY and SJ wrote the manuscript. HH and DC designed the study. HH, DC, HP , SC and HT wrote the manuscript and agree to be accountable for all aspects of the work are appropriately investigated and resolved.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


