Identifying survival-associated ceRNA clusters in cholangiocarcinoma

MING WAN1*, FU-MIN ZHANG1,2,*, ZHENG-LONG LI1, PENG-CHENG KANG1, YI-MIN WANG1, ZHI-DONG WANG1, XIANG-YU ZHONG1, CHUN-LONG LI1, HAO WANG1, SHI-YONG ZHAO1 and YUN-FU CU1

1Department of Hepatopancreatobiliary Surgery, Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150086; 2Department of General Surgery, Daqing Oilfield General Hospital, Daqing, Heilongjiang 163000, P.R. China

Received February 6, 2016; Accepted March 17, 2016

DOI: 10.3892/or.2016.4934

Abstract. Competing endogenous RNAs (ceRNAs) represent a novel layer regulations of long non-coding RNAs (lncRNAs) and genes that play important roles in cancer pathogenesis by binding microRNAs (miRNAs). However, the competition mechanism of ceRNAs in cholangiocarcinoma (CHOL) is not fully understood. In this study, we constructed a dysregulated ceRNA competitive network (CCEN) to globally characterize the competing difference between CHOL and normal tissues. Then, we integrated affinity propagation and Kaplan-Meier (K-M) methods to identify functional clusters associated with survival. A total of 7 key ceRNA clusters were identified. Further functional annotation analyses found that Cluster23 and Cluster32 involved cell based functions, and the loss of ceRNA competitive relations in clusters may contribute to CHOL, by disturbing important biological processes, such as ‘Pathway in cancer’, MAPK and Neurotrophin signaling pathway. This study provides further insights into understanding the competitive mechanism of ceRNAs in CHOL.

Introduction

Cholangiocarcinoma (CHOL), which predominantly arises from the epithelial bile duct, is an aggressive malignant tumor of the biliary tract system (1,2). In clinical treatment, due to detection difficult at an early stage, most CHOLs are diagnosed at an advanced stage, which leads to high mortality, shorter survival and poor prognosis (3,4). Therefore, identifying potential molecular biomarkers is beneficial for CHOL diagnosis, patient prognosis and targeting treatments (5,6), and can promote understanding of the pathological mechanism of CHOL. Long non-coding RNAs (IncRNAs), as transcripts with more than 200 nucleotides in length, play key roles in transcriptionally and post-transcriptionally regulating gene expression (7-9). RNAs have been found involved in multiple cancers by influencing a wide range of functions, such as cell proliferation, cell apoptosis and cell invasion (10-13). Recent studies reported that IncRNAs may be potential diagnostic and prognostic biomarkers for CHOL patients (14,15). Though the roles of IncRNAs acting as biomarkers have been detected, the potential mechanisms are not fully understood in CHOL.

Recent studies revealed that IncRNAs with microRNA (miRNA) responsive elements (MRE) can act as competing endogenous RNAs (ceRNAs) to compete with mRNAs for binding miRNAs, and thus affect the expression level of genes (16-18). The abnormal regulation of ceRNAs relate to many cancers, such as breast, colorectal, gastric and lung cancer (19-23). For example, CXCR as a ceRNA promotes metastasis, proliferation and survival of MCF-7 cells by controlling miRNA activities (24). ROR-Nanog can function as a ceRNA pair and compete for miR-145 predicting poor clinical outcome of pancreatic cancer patients (25). To detect the competing roles of ceRNAs in disease pathology processes, Salmena et al focused on constructing a large-scale ceRNA regulatory network (26). Broad ceRNA interaction network was also identified by considering the ceRNA pairs who significantly shared common miRNAs, and revealed the switch roles of ceRNAs between breast cancer and normal tissues (27). Li et al also demonstrated that the dysregulated interactions in ceRNA network are responsible for high-risk cancers (28).

In the present study, to identify the potential ceRNA interactions contributing to patient survival, we constructed a global ceRNA regulatory network that dysregulated between CHOL and normal tissues. A total of 7 key ceRNA clusters were identified. The Cluster23 involved cell cycle, cell
division, cell proliferation and cell death processes, and mediated the regulation of ‘Pathway in cancer’. Another key cluster: Cluster32, was found involved in cell proliferation, apoptosis and programmed cell death processes, and associated with Neurotrophin signaling, focal adhesion and MAPK signaling pathway. This study provides further insights into understanding the competitive mechanism of ceRNAs in affecting the survival of CHOL patients, and may have important clinical significance for screening diagnostic markers.

Materials and methods

Materials

LncRNA and miRNA sequences. We obtained the gene annotation of lncRNAs from GENCODE v7 GTF file (29) and the sequences of lncRNAs from Ensembl database (GRCH37) (30). A total of 9,532 lncRNAs with length >200 nt were considered. The sequences of mature miRNAs were identified from miRBase (31) and 2,857 miRNAs were finally considered.

Verification of miRNA-target interactions. We obtained the experimentally verified miRNA-gene interactions from mirTarBase (V4.5) (32) and TarBase (V6.0) (33). A total of 3,566 miRNA-gene relation pairs were identified. The sequences of mature miRNAs were identified from miRBase (31) and 2,857 miRNAs were finally considered.

Identification of potential miRNA-lncRNA-gene relations. There are four steps to identify potential ceRNA interactions: i) we predicted miRNA-lncRNA interactions based on sequence matching algorithms, such as miRanda (36), RNAhybrid (V2.1.1) (37), PITA (38) and TargetScan (V6.0) (39), ii) to extract functional interactions, we analyzed if lncRNA sequences had the binding sites of miRNAs, by integrating the genome coordinates of CLIP-seq peaks (40), then 39,762 miRNA-lncRNA interactions were obtained, iii) by using 3,566 miRNA-gene interactions and 39,762 miRNA-lncRNA interactions, we identified 526,173 potential miRNA-lncRNA-gene (MLG) relations.

Methods

Identification of phenotype-associated ceRNA interactions. We obtained the phenotype-associated MLG relations
that involved 4,912 differential genes and 193 differential lncRNAs from 526,173 potential MLG relations. Then hypergeometric method was used to test the enrichment significance of miRNAs between one differential gene and one differential lncRNA. One ceRNA pair was considered as a candidate with \( \text{PG,}_L < 0.05 \). The enrichment significance can be calculated by using the following formula:

\[
P_{G,L} = 1 - \sum_{j=0}^{k-1} \binom{n_j}{j} / \binom{n}{j}
\]

In the above formula, \( m \) represents the total number of mature miRNAs in miRBase, \( n \) is the number of miRNAs targeting gene \( G \), \( i \) is the number of miRNAs targeting lncRNA \( L \), \( k \) is the number of miRNAs shared by \( G \) and \( L \). We identified 932 candidate ceRNA pairs that associated with the phenotype, containing 427 genes and 34 lncRNAs.

**Evaluation of significantly dysregulated ceRNA interactions.** We considered one ceRNA interaction was dysregulated by estimating the difference score (\( DS \)) of Pearson's correlation coefficients (\( PCC \)) between cancer samples and normal samples. The formula is as follows:

\[
DS_{G,L} = PCC_{cancer}(G, L) - PCC_{normal}(G, L)
\]

The \( PCC_{cancer}(G, L) \) and \( PCC_{normal}(G, L) \) was the \( PCC \) between gene \( G \) and lncRNA \( L \) in cancer samples and normal samples, respectively. The \( DS_{G,L} \) was the differential score of \( PCC \) between gene \( G \) and lncRNA \( L \). The higher \( DS_{G,L} \) and lower \( DS_{G,L} \), respectively, indicated increasing positive and negative interaction in cancer samples. We called it ‘gain interaction’ when \( DS_{G,L} > 0.5 \) and ‘loss interaction’ when \( DS_{G,L} < -0.5 \).

To evaluate the statistical significance of \( DS_{G,L} \), we randomly disturbed the normal and cancer labels of all samples while keeping the sample size of normal and cancer unchanged. We performed this process 1,000 times and re-estimated the random \( DS_{G,L} \) scores. The statistical significance can be formulated as follows:

\[
P\text{-value} = \frac{N(\text{random} \ DS_{G,L} > \text{true} \ DS_{G,L})}{1,000}
\]

One ceRNA pair was significantly dysregulated with \( P\text{-value} < 0.05 \). We identified 504 ceRNA pairs with 33 lncRNAs and 298 genes, which significantly dysregulated between two different phenotypes. Finally, 107 ceRNA pairs

<table>
<thead>
<tr>
<th>Cluster ID</th>
<th>LncRNA</th>
<th>Gene</th>
<th>Log-rank (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cluster23</td>
<td>RP4-564F22.2, HCG18, LOC678655, RP11-4369.2, CTD-2561B21.9, C12orf47, CTC-228N11.3, AC092296.1, RP11-403I13.5, H19, MEG3</td>
<td>UBE2I, IL8, HMGN1, PLP2, FSCN1, DAB2IP, CCNE1, FAR1, CDK9, PKM, MEIS2, C11orf9, HRSP12, CD44, PGF, BCL2, TGFβ2</td>
<td>4.63E-05</td>
</tr>
<tr>
<td>cluster24</td>
<td>LOC678655, H19, RP11-622K12.1, AC092296.1, C12orf47, RP11-875O11.1, AC108488.3, RP11-403I13.5, RP11-444D3.1, MEG3</td>
<td>RAC1, RHOC, MYLIP, UBE2S, PLAT, JAG1, MAP3K10, SULF1, FZD7, E2F5, SOX9, SOX4, ACVR1, CYP1A1</td>
<td>4.63E-05</td>
</tr>
<tr>
<td>cluster42</td>
<td>HCG18, MEG3, H19, TUG1, RP11-2C24.4, LINCO0152, C12orf47, RP11-444D3.1, AC092296.1, ZNFX1-AS1, RP11-622K12.1, PTEN, CTC-504A5.1, CTD-2561B21.9, RP4-564F22.2, RP11-4369.2</td>
<td>LIF, CHD4, UBE2I, C12orf18, CYP7A1, CCND3, ADAM17, HNF4A, POGK, PLEC, GALNT7, CCNE1, SRGAP1, ARID4B, ARL2, CDH1, RHOC, COL1A2, PEA15, TPM1, PDLM7</td>
<td>4.63E-05</td>
</tr>
<tr>
<td>cluster43</td>
<td>LOC678655, LINCO0152, C12orf47, RP11-2C24.4, AC092296.1, ZNFX1-AS1, CTC-504A5.1, RP11-622K12.1, RP4-564F22.2, RP11-444D3.1, TUG1, SNSHG6, PTEN</td>
<td>MFSID10, MYH9, MRE11A, ZNF384, SLC12A2, COL4A1, SKAP2, TCEAL1, EGLN3, MTA2, ARID4B, TPM2, RASA1, FGRF3, PLAT, IGF2BP1, SLC7A1, ABCG2, CCND3, S100B</td>
<td>0.00236</td>
</tr>
<tr>
<td>cluster32</td>
<td>RP11-622K12.1, MEG3, RP11-2C24.4, RP4-564F22.2, C12orf47, PTEN, H19, AC09288.3, RP11-4369.2, LOC678655, HCG18, RP11-91K9.1, RP11-279F6.1, AC092296.1, RP11-444D3.1</td>
<td>BCL2, RAC1, DSTRYK, CDCP1, SETD4, TP53, HBXIP, HRNPDL, SOX9, MUC1, ARID4B, TMED3, SOX4, CHEK1, SPARC, STMN1, PEA15, DMD, TNC, IL8</td>
<td>0.0117</td>
</tr>
<tr>
<td>cluster40</td>
<td>AC108488.3, HCG18, RP11-875O11.1, LINCO0152, TUG1, LOC678655, VT1, CTD-2561B21.9, RP11-444D3.1, C12orf47, AC092296.1, RP11-4369.2, RP11-2C24.4, ALDH1L1-AS2, H19, MEG3, RP11-91K9.1</td>
<td>ARHGEF3, PPARA, ANXA11, MMP15, ALDOA, KRT7, DMTF1, PKMYT1, ISYNA1, ANK1B1, API1M2, MYO10, JAG1, EHD2, THRA, PEA15, GALNT7, RARG, TSPAN15, BCL2, CCNE1, SULF1, DVL2, SOD1, CCND2, PDLM7, TMEM109, COL15A1, PTPN13, CFL2</td>
<td>0.0251</td>
</tr>
</tbody>
</table>

Table I. The clusters associated with survival.
with gain interaction and 396 ceRNA pairs with loss interactions were obtained.

Cluster functional ceRNA modules. We used an R package ‘apcluster’ (41) to cluster functional ceRNA modules. The ‘negDistMat’ function firstly created a squared negative distance matrix with $D_{S_{G,L}}$ scores of 504 ceRNA pairs as parameters. The ‘apcluster’ function was secondly used, with distance matrix as input, to perform affinity propagation clustering process. This method produced clusters by iteratively maximizing their similarities. We obtained a total of 48 ceRNA clusters and only considered 41 clusters with nodes size >5.

Survival analysis. We performed Kaplan-Meier (K-M) method to estimate cumulative survival rates and log-rank test to evaluate the difference in overall survival between different group patients. A P-value of <0.05 was considered to indicate statistical significance.

Results

The construction of CCEN. The dysregulated ceRNA network in CHOL (CCEN) was constructed, based on IncRNA-gene interactions with significant $D_{S_{G,L}}$ scores. This network contained 294 genes, 33 IncRNAs and 504 edges (Fig. 1). The gain and loss interaction modes of IncRNAs and genes can be shown in CCEN, which characterized that two nodes linked to common edge had increasing positive and negative mode in CHOL, respectively. A total of 108 ‘gain interaction’ and 396 ‘loss interaction’ were obtained (data not shown).
Survival-related ceRNA clusters. The ‘Affinity Propagation’ method was used to cluster functional ceRNA modules from CCEN. We obtained 49 clusters and only considered 32 clusters with node size >5 and edge size >5. By performing ‘K-M’ survival analysis, 7 key clusters can significantly distinguish survival rates between two different groups (Table I).

Function analysis of key ceRNA clusters. We performed GO and pathway functional annotation for each key ceRNA cluster, by using functional annotation tool of dA VId (42). Genes within one cluster were used. The annotation results are shown (data not shown).

Discussion

LncRNAs have been shown to play key roles in regulating gene expression by competing for miRNAs, and the ceRNA relations formed can contribute to the development and progression of multiple cancers (17,43). To explore the competitive relationship of IncRNAs and genes, we firstly established the potential miRNA-IncRNA-gene set. The experimentally verified miRNA-mRNA interactions werw collected from mirTarBase (V4.5) and TarBase (V6.0). However, detecting the interactions between miRNAs and IncRNAs is a great challenge. Several studies identified physical miRNA-IncRNA interactions by evaluating the match extent and conserved seed regions (44,45). The functional miRNA-IncRNA interactions were predicted by identifying the SNPs that affected the binding of miRNAs and IncRNAs (46). In addition, the experimentally verified and predicted miRNA-IncRNA interactions were recently identified in DIANA-LncBase and starBase (V2.0). Here, we applied the traditional miRNA target prediction methods by analyzing AGO-CLIP-seq data set and integrated the verified interactions to enhance the functional reliabilities of candidate miRNA-IncRNA interactions. The potential ceRNAs data set was finally obtained by integrating miRNA-mRNA and miRNA-IncRNA interactions. Because IncRNAs can compete with endogenous mRNAs for binding miRNAs, we further used hypergeometric method to identify the high-confidence ceRNAs.

We constructed a bilayer network named CCEN and dissected the characteristics of CCEN to reveal the potential competitive roles of ceRNAs. The CCEN conformed to power-law distribution with $R^2=0.768$ and slope=1.282 (Fig. 2). The average degree of genes in network was 1.69, whereas that of IncRNAs was up to 15.27 (Fig. 2). The average betweenness centrality of genes was 0.0039, whereas that of IncRNAs (0.0598) was about 15-fold higher than the genes (Fig. 2). The IncRNA PVT1 with a degree of 25 and betweenness centrality of 0.1119 were higher than each average level, have been demonstrated to be associated with many cancers. For example, PVT1 can frequently amplify or mutate to promote the pathophysiological process of ovarian, breast cancer, Burkitt's lymphomas and murine plasmacytomas (47-49).
The lncRNA ZNFX1-AS1 with degree of 21 and betweenness centrality of 0.0766, is a putative tumor suppressor downregulated in breast cancer and ductal carcinoma (50). It suggested that lncRNAs tended to compete with multiple genes and played crucial bridge roles for conveying information in cancers.

To obtain functional ceRNA modules that associated with patient survival, we applied affinity propagation cluster method to CCEN and identified 48 functional clusters. As the few nodes and edges in a single cluster were difficult to connect in function, we retained 32 clusters with both node size >5 and edge size >5. Then K-M survival analysis method was used, to estimate if the elements in one cluster can distinguish the survival rates between two different groups. The Cluster23 had 11 lncRNAs and 17 genes (Fig. 3A), which was one of the most significant clusters related to CHOL patient survival (P-value =4.63E-05, Fig. 3B). We extracted the subnetwork consisted of Cluster23 members in CCEN, most ceRNA pairs shared one or more competitors with others, and only 4 ceRNA pairs were specific in competition. In GO functional analysis (data not shown), Cluster23 was associated with many important biological processes, such as cell cycle,
cell division, cell proliferation, tissue and tube morphogenesis. Interestingly, we also found that all competitive relations except H19-HRSP12 of Cluster23 were ‘loss interaction’ in CHOL compared to normal tissues. It revealed that the lost of ceRNA competitive relations may promote the development and progression of CHOL, by disturbing the cycle, division or proliferation of cells. To examine the roles of Cluster23 in the pathway, we dissected the most significant pathway ‘Pathway in cancer’ (P-value =2.30E-04, data not shown). The Cluster23 members, Bcl-2, TGFB, VEGF and cyclin E, were all located in the central position of pathway and the upstream of other key pathways, such as sustained angiogenesis, proliferation and evading apoptosis process (Fig. 3C). Pugazhenthi et al demonstrated that H19-7 cells, with Bcl-2 overexpressed, can resist ROS-induced apoptosis (51). A recent study confirmed that MEG3 expression is low in breast tumors with high TGFB2 expression (52). Therefore, H19 and MEG3 negatively regulate Bcl-2 and TGFB2 expression in tumors, respectively, and may contribute to CHOL phenotype.

Another key survival cluster was Cluster32 (Fig. 4A), with 15 IncRNAs and 20 genes (P=0.012, Fig. 4B). In this cluster, all competitive relations were ‘loss interaction’, which may indicate the relationships between dysregulated ceRNA interactions and CHOL phenotype. We found that genes in this cluster involved cell proliferation, apoptosis and programmed cell death processes (data not shown). The result of further pathway enrichment analysis showed that Cluster32 was significantly associated with Neurotrophin signaling (P=0.011), focal adhesion (P=0.029) and MAPK signaling pathway (P=0.048) (data not shown). These pathways had been confirmed related

Figure 5. Analysis of survival-associated Cluster32. (A) The network structure of module Cluster32. (B) ‘K-M’ survival analysis of Cluster32 members. (C) The annotated pathway ‘Neurotrophin signaling pathway’.
to the development and progression of cancers. For example, the induce of MAPK signaling pathway activity promotes the metastasis of breast cancer cells and enhances cancer malignancy (53,54), inhibiting focal adhesion kinase can induce apoptosis of bladder and breast cancer cells (55,56), the activation of neurotrophin signaling promotes the growth of brain tumor-initiating cells and the metastasis of lung adenocarcinoma (57,58). Then we dissected the regulatory mechanism of ceRNA interactions in CHOL through analyzing Neurotrophin signaling pathway (Fig. 5). In this pathway, gene Rac1 was located in the initial position, IncRNA MEG3 targeting Rac1 was demonstrated to suppress migration and invasion of thyroid carcinoma (59). TP53 was a downstream gene of Rac1, the abnormal expression of which contributes to many cancers, such as large B-cell lymphoma (60), gastrointestinal cancer (61), and lymphocytic leukemia (62). The IncRNA C12orf47 competed with TP53 and influenced its expression, which may mediate the apoptosis of CHOL by disturbing the Neurotrophin signaling pathway. These results suggested that the Cluster32 we identified may control the survival of CHOL patients, by mediating ceRNA regulation in cell functions.

In conclusion, although the results of the present study require further experimental verification, the results provide further insights into understanding the loss of ceRNA interactions in affecting the survival of CHOL patients.

References


