Cancer-related long noncoding RNAs show aberrant expression profiles and competing endogenous RNA potential in esophageal adenocarcinoma

YANG YU*, XINGXING CHEN* and SHUNDONG CANG

Department of Oncology, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou University, Zhengzhou, Henan 450003, P.R. China

Received January 24, 2019; Accepted July 17, 2019

DOI: 10.3892/ol.2019.10808

Abstract. Long non-coding RNAs (lncRNAs) govern gene expression by competitively binding to microRNA response elements (MREs). Although they were initially considered as transcriptional noise, lncRNAs have attracted increased attention in oncology. Dysregulation of lncRNAs occurs in various types of human tumor, including esophageal adenocarcinoma (EAC). However, the functions of these cancer-associated lncRNAs and of their related competitive endogenous RNA (ceRNA) network in EAC remains unknown. To determine the relevant potential mechanisms, the present study analyzed the transcriptome sequencing data and clinical information of 79 patients with EAC, including 79 tumor samples and 11 normal samples, which were obtained from The Cancer Genome Atlas esophageal cancer project. The edgeR v3.25.0 software was used for differential gene expression analysis. The results exhibited 561 cancer-associated lncRNAs with a >2.0-fold change and a false discovery rate-adjusted P<0.01. Among these lncRNAs, 26 were significantly associated with patient overall survival. According to data from bioinformatics databases and differentially expressed RNAs, an lncRNA-regulated ceRNA network for EAC was constructed. The results demonstrated that the aberrantly expressed lncRNA-associated ceRNA network included 37 EAC cancer-associated lncRNAs, five miRNAs and 13 mRNAs. In conclusion, the present study identified novel lncRNAs as candidate prognostic biomarkers and revealed a potential regulatory network of gene expression in EAC.

Introduction

Esophageal adenocarcinoma (EAC) is a highly lethal malignancy that occurs mainly in the distal esophagus and gastroesophageal junction (1). EAC is rare in China; however, it represents the predominant type of esophageal cancer in North America and Europe. In these continents, the overall incidence of EAC has rapidly increased over the past three decades at a rate (5-10%) greater than that of any other major cancer and the incidence rate is higher in white males compared with that in white females (2-4). The reason for this increase is not entirely understood. Previous studies have reported that EAC differs from esophageal squamous cell carcinoma (ESCC) in terms of genetic and environmental risk factors such as tobacco use, alcohol, obesity and germline mutations (5,6). Systematic therapy for EAC typically includes endoscopic mucosal resection, surgical resection, chemoradiotherapy and neoadjuvant chemotherapy; however, the mortality rate remains high and the overall 5-year survival rate is 17% in the United States (7). Although endoscopy can accurately diagnose early-stage EAC, most patients are diagnosed with regional metastasis or distant metastasis, which are positively correlated with a considerable decline in the 5-year survival rate (8). There is therefore an urgency to identify novel potential diagnostic and prognostic biomarkers for EAC.

Long non-coding RNAs (lncRNAs) represent a new class of non-coding RNAs (ncRNAs) and are defined as transcripts >200 nucleotides in length (9). Unlike their shorter counterparts, including microRNAs (miRNAs), the roles and underlying mechanisms of lncRNAs in human disease remain largely unknown. Due to improvements in DNA sequencing techniques, numerous lncRNAs have been discovered. In addition, an increasing number of lncRNAs have been identified in human cancer, such as HAGL-1 and opposite strand IncRNA overexpression in gastric cancer (10). Previous studies have focused on the biological function and underlying molecular mechanism of lncRNAs in various types of cancer, including colorectal cancer, gastric cancer, hepatocellular carcinoma, renal cell carcinoma, prostate carcinoma and EAC (11-17). Although studies reported that lncRNAs can be involved in the development and progression of ESCC (6,14), only...
a few studies have determined the function of IncRNAs in EAC (5,18).

Emerging technologies have increased our ability to determine the functions of cancer-associated IncRNAs. Significant progress towards understanding the underlying molecular mechanism by which IncRNAs can regulate miRNA function has therefore been made. Salmena et al (19) proposed a competing endogenous RNA (ceRNA) language where protein coding genes, microRNAs and IncRNAs communicate with each other by competitively binding to shared miRNA response elements (MREs). Competing endogenous RNA networks comprise a new regulatory network of mRNAs and non-coding RNAs, which reveals a greatly expanded role for IncRNAs in human disease (20). This hypothesis has been experimentally validated. For example, Cesana et al (21) identified a muscle-specific IncRNA named linc-MD1, which regulates the expression of mastermind-like 1 and myocyte-specific enhancer factor 2C by serving as a ‘sponge’ for miR-133. Furthermore, Qu et al (22) demonstrated that IncARSRS mediates sunitinib resistance in renal cell carcinoma by competitively binding to miR-34/miR-449 to promote AXL receptor tyrosine kinase and c-MET expression. Exploration of RNA cross-talk offers therefore insights into cancer diagnosis and therapy. An IncRNA-miRNA-mRNA ceRNA network has therefore been constructed for various types of human cancer, in particular for ESCC (23-26); however, such a network has not yet been described for EAC.

In order to systematically describe EAC-associated pseudogenes and to construct a ceRNA network, the present study comprehensively analyzed RNA sequencing (RNA-Seq) transcript data that were obtained from The Cancer Genome Atlas (TCGA) esophageal cancer project (https://www.cancer.gov/types/esophageal). The database includes IncRNA, microRNA and mRNA data and clinical information from patients with EAC. The present study included 79 EAC tumor and 11 adjacent non-tumor esophagus tissue samples.

By using publicly available RNA-Seq data from TCGA, some EAC-associated IncRNAs, mRNAs and miRNAs were identified based on the ceRNA hypothesis. Furthermore, 561 differentially expressed IncRNAs (DElncRNAs), 1,289 differentially-expressed mRNAs (DEmRNAs) and 44 differentially-expressed miRNAs (DEmiRNAs) were identified. Subsequently, five dysregulated IncRNAs, 13 mRNAs and 32 miRNAs were identified and included in a constructed ceRNA network based on IncRNA-miRNA interactions predicted by miRcode v11 (www.mircode.org/). Potential prognostic biomarkers were then identified by exploring the influence of dysregulated RNAs on overall survival using the univariate Cox proportional hazards regression model and Kaplan-Meier curve analysis. The results from this comprehensive analysis provided the foundation for deeper understanding of the cancer-associated IncRNA functions in EAC and revealed potential prognostic biomarkers.

Materials and methods

Patients and samples. Data for 187 patients with esophageal cancer were obtained from the TCGA data portal (https://portal.gdc.cancer.gov/). The exclusion criteria were as follows: i) Patients with ESCC or undetermined pathological classification; and ii) samples without corresponding RNA-Seq and miRNA-Seq data. Overall, data from 79 patients with EAC were enrolled in the present study. This study followed the publication guidelines provided by TCGA (http://cancergenome.nih.gov/publications/publicationguidelines).

RNA-Seq data. RNA-Seq and miRNA-Seq data (level 3) were downloaded from 90 tissue samples of the TCGA database, including 79 EAC samples and 11 adjacent normal samples. The gene expression profiles generated from Illumina Hiseq platforms (Illumina, Inc.) were all publicly available data.

Analysis of DEmRNAs, DELncRNAs and DEmiRNAs. The raw count data were processed with edgeR v3.25.0 (Bioconductor), which is a package based on the R language (v3.5.0) (27)

Table I. Clinical characteristics of patients with esophageal adenocarcinoma.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No. cases</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis, years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;60</td>
<td>26</td>
<td>32.9</td>
</tr>
<tr>
<td>≥60</td>
<td>53</td>
<td>67.1</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>11</td>
<td>13.9</td>
</tr>
<tr>
<td>Female</td>
<td>68</td>
<td>86.1</td>
</tr>
<tr>
<td>Metastasis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>57</td>
<td>72.2</td>
</tr>
<tr>
<td>M1</td>
<td>10</td>
<td>12.7</td>
</tr>
<tr>
<td>MX</td>
<td>10</td>
<td>12.7</td>
</tr>
<tr>
<td>Lymph node status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>22</td>
<td>27.8</td>
</tr>
<tr>
<td>N1</td>
<td>44</td>
<td>55.7</td>
</tr>
<tr>
<td>N2</td>
<td>5</td>
<td>6.3</td>
</tr>
<tr>
<td>N3</td>
<td>5</td>
<td>6.3</td>
</tr>
<tr>
<td>NX</td>
<td>3</td>
<td>3.8</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>9</td>
<td>11.4</td>
</tr>
<tr>
<td>II</td>
<td>22</td>
<td>27.8</td>
</tr>
<tr>
<td>III</td>
<td>26</td>
<td>32.9</td>
</tr>
<tr>
<td>IV</td>
<td>5</td>
<td>6.3</td>
</tr>
<tr>
<td>T stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>1</td>
<td>1.3</td>
</tr>
<tr>
<td>T1</td>
<td>19</td>
<td>24.1</td>
</tr>
<tr>
<td>T2</td>
<td>10</td>
<td>12.7</td>
</tr>
<tr>
<td>T3</td>
<td>46</td>
<td>58.2</td>
</tr>
<tr>
<td>T4</td>
<td>1</td>
<td>1.3</td>
</tr>
<tr>
<td>TX</td>
<td>2</td>
<td>2.6</td>
</tr>
<tr>
<td>Histological grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>1</td>
<td>1.3</td>
</tr>
<tr>
<td>G2</td>
<td>28</td>
<td>35.4</td>
</tr>
<tr>
<td>G3</td>
<td>24</td>
<td>30.4</td>
</tr>
<tr>
<td>GX</td>
<td>26</td>
<td>32.9</td>
</tr>
</tbody>
</table>
for differential gene expression analysis. For all P-values, a false discovery rate (FDR) was applied to correct the statistical significance of multiple testing. Genes with >2.0 fold change (FC) and FDR-adjusted P<0.01 were considered significant. The volcano plot and heat map were designed to visualize the results using ggplots v3.0, which is a package based on the R language.

Association between DElncRNAs and patient prognosis. All patients were classified into high or low lncRNA-expression groups according to the median. Kaplan-Meier and log-rank methods were used to test differences between the two groups. P<0.05 was considered to indicate a statistically significant difference.

Construction of the ceRNA network. Three calculations were performed to construct the ceRNA network as follows: i) Cancer-associated lncRNA filtration, where lncRNAs with FC>2.0 (either up- or downregulated) and FDR-adjusted P<0.01 were considered as cancer-associated lncRNAs [to improve data reliability, cancer-associated lncRNAs that were not annotated by GENCODE (http://www.gencodegenes.org/) were excluded]; ii) lncRNA-miRNA interactions were predicted by miRcode (http://www.mircode.org/) and starBase (http://starbase.sysu.edu.cn/); and iii) target mRNAs of DEmiRNAs were predicted using the three bioinformatics databases miRDB (http://mirdb.org/), miRTarBase (http://miRTarBase.mbc.nctu.edu.tw/php/index.php) and TargetScan (http://www.targetscan.org). Gene Oncology (GO) was analyzed using Database for Annotation, Visualization, and Integrated Discovery bioinformatics tools (DAVID; v6.8; https://david.ncifcrf.gov). In order to improve the consistency of the bioinformatics analysis, the target genes were retained.

A network graph was constructed and visualized using Cytoscape v3.5.1 (https://cytoscape.org/).

Results

Patient characteristics. The detailed clinical information and pathological characteristics of the patients included in the present study, including sex, age at diagnosis, metastasis status, lymph node status and tumor-node-metastasis stage, are presented in Table I. The median age for all patients was 69 years (range, 27-86 years). The median overall survival was 14.29 months (range 0.36-83.18 months).

Identification of DElncRNAs. lncRNAs with FDR-adjusted P<0.01 and FC >2.0 were considered to be differentially expressed. A total of 561 DElncRNAs were identified, of which 217 were upregulated and 344 were downregulated (Table SI). A volcano plot was therefore constructed (Fig. 1A) to visually describe the FDRs and FCs. In addition, a heat map was designed (Fig. 1B) to highlight the top 100 significant DElncRNAs according to the FDR-adjusted P-values.

A network graph was constructed and visualized using Cytoscape v3.5.1 (https://cytoscape.org/).
LINC00163, LINC00906, LINC01695, SLCO4A1-AS1 and UG0898H09, were positively correlated with OS (Fig. 2B).

**Identification of DEmRNAs.** The RNA expression levels in 79 EAC tumor samples and 11 normal samples were analyzed. With a cut-off value of FDR-adjusted $P<0.01$ and FC $>2.0$, 367 upregulated and 922 downregulated mRNAs were identified (Table SII). A volcano plot was therefore constructed to visualize the results (Fig. 3A). The top 100 significant DEmRNAs were then highlighted by plotting FDR-adjusted $P$-values in a heat map (Fig. 3C).

To analyze the DEmRNA functions, enrichment analysis based on enriched functional GO modules was performed. The results demonstrated that the DEmRNAs were significantly enriched in the ‘chemokine-mediated signaling pathway’ (GO: 0070098), ‘plasma membrane’ (GO: 0005886) and ‘calcium ion binding’ (GO: 0005509) GO terms under ‘biological process’, ‘cellular component’ and ‘molecular function’, respectively (Fig. 3B).

**Identification of DEmiRNAs.** In order to design a ceRNA network for EAC, the miRNA expression profiles between tumor samples and normal samples were compared. Subsequently, 44 DEmiRNAs, including 28 upregulated and 16 downregulated were identified (Fig. 4A and C). The miRNAs that were targeted by the 44 DEmiRNAs from miRDB,
miRTarBase and TargetScan were then screened. To improve data reliability, mRNAs that were not included in the set of 1,289 DEmRNAs were excluded. Eventually, 13 DEmiRNAs remained in the ceRNA network (Fig. 4B).

In order to identify the DEmRNAs that may have potential prognostic ability, the expression profiles of the 13 DEmiRNAs included in the ceRNA network were analyzed using Kaplan-Meier curve. The results demonstrated that the

Figure 2. Continued. (B) A total of 14 lncRNAs were positively correlated with overall survival. lncRNA, long non-coding RNA; OS, overall survival.
expression profiles of three DEmRNAs were positively correlated with OS (P<0.05). Two of these DEmRNAs, angiopoietin 2 and interleukin 11 (IL11) were negatively correlated with OS, whereas neurotrophic receptor tyrosine kinase 2 (NTRK2) was positively correlated with OS (Fig. 4D).

To identify the lncRNA-miRNA interactions in EAC, the potential MREs in DElncRNAs were screened using miRcode. For miRNA-mRNA interactions, miRDB, miRTarBase, and TargetScan were used to identify the DEmRNAs targeted by DEmiRNAs. The results are listed in Tables II and III.

d**ceRNA network construction.** In order to improve knowledge on DElncRNA function in EAC, a dysregulated lncRNA-miRNA-mRNA ceRNA network based on the aforementioned data (presented in Tables II and III) was constructed. The results demonstrated that in the ceRNA network, interaction of five DEmiRNAs with 37 DElncRNAs was predicted, according to the results retrieved from miRcode. The ceRNA network is presented in Fig 5.

**Discussion**

lncRNAs represent a crucial category of non-coding genes in the transcriptome that act as pivotal regulators of cell physiology and pathology in human cancer by mediating gene expression through multiple mechanisms (28). The dysregulation of lncRNA expression is involved in the pathogenesis of various types of solid tumor (29,30). Numerous novel biological functions of lncRNAs have been reported in cancer, including transcriptional regulation, competing endogenous RNA (ceRNA) network construction, and epigenetic modifications.
functions have been attributed to lncRNAs, which have become the focal point of many studies (10,17). The ceRNA hypothesis describes regulatory networks among protein-coding mRNAs and non-coding RNAs, including miRNAs and lncRNAs at the post-transcription level. According to this hypothesis, changes in the expression of one or multiple miRNA targets can alter the number of unbound miRNAs and lead to observable changes in miRNA activity. The various transcripts

Figure 4. DEmiRNAs in esophageal adenocarcinoma. (A) Volcano plot of log2FC vs. log10(FDR) for differentially expressed miRNAs. Red dots represent significantly upregulated miRNAs and green dots represent significantly downregulated miRNAs. (B) Venn analysis of overlapping genes between statistically significant mRNAs and target genes of DEmiRNAs. (C) Significant DEmiRNAs were visualized by heat map. Each column represents one sample and each row represents an individual miRNA. (D) Kaplan-Meier survival curves of three target genes in overall survival prediction (P<0.05). DEmiRNAs, differentially expressed miRNAs; FC, fold change; FDR, false discovery rate; ANGPT2, angiopoietin 2; IL-11, interleukin-11; NTRK2, neurotrophic receptor tyrosine kinase 2; miRNA, microRNA.
from the transcriptome communicate with one another by competitively binding to shared MREs (20). ceRNA networks in human cancer include cancer-associated IncRNAs, microRNAs and mRNAs. A previous study demonstrated a miRNA-IncRNA-mRNA interaction in ESCC (31). However, the ceRNA network in EAC remains poorly understood.

At present, since IncRNAs are able to regulate miRNA functions by competitively binding to shared MREs in mRNA, they are considered as diagnostic and prognostic biomarkers. Numerous well-studied IncRNAs have been identified as potential targets or powerful predictors in various types of cancer, including LINC00668, H19 and UCA1 (32-34). However, studies on EAC remain rare. Based on the RNA-Seq data and clinical data from 79 patients with EAC, the present study demonstrated that 26 cancer-associated IncRNAs may affect the OS of patients with EAC. In particular, the results from this study reported that two DEIncRNAs, CYP1B1-AS1 and HOTAIR, were not only identified as part of the ceRNA network, but were also positively and negatively correlated with OS, respectively, which suggested that these two lncRNAs may serve as essential oncogenes and as prognostic markers in EAC.

HOTAIR is a highly studied IncRNA. Previous studies demonstrated that it serves a role in the development and progression of various types of solid tumor, including renal cell carcinoma, colorectal cancer, breast cancer, gastric cancer, non-small cell lung cancer, cervical cancer and ovarian epithelial carcinoma (35-45). In addition, Ren et al (46) demonstrated that HOTAIR can control the cell cycle by acting as a competing endogenous ‘sponge’ to downregulate miR-1 and upregulate cyclin D1 in ESCC. The present study reported that HOTAIR expression was upregulated in EAC tumor tissues. In addition, patients with highly expressed HOTAIR had worse survival outcomes. HOTAIR may therefore compete with miR-301b and miR-204 to regulate chordin like 1, NTRK2, IL11, neuronal pentraxin 1, homeobox C8 and solute carrier family 22 member 6 expression. Although these mRNAs have been identified as aberrantly expressed, their roles have not been fully investigated in EAC.

In addition to HOTAIR, LINC00163 and SLCO4A1-AS1 have also been reported to be associated with cancer prognosis. Guo et al (47) demonstrated that the LINC00163 level is significantly decreased in lung cancer tissues and cell lines following bioinformatics and reverse transcription-quantitative PCR analyses. LINC00163 expression was lower in metastatic tissues compared with non-metastatic tissues, and a higher LINC00163 expression in patients with lung cancer could predict a better prognosis. Yang et al (48) reported that SLCO4A1-AS1 expression was more upregulated in bladder cancer tissues compared with that in adjacent normal tissues, and that SLCO4A1-AS1 overexpression is associated with poor
prognosis and tumor metastasis. Yu et al (49) demonstrated that a high SLCO4A1-AS1 expression level is associated with bladder cancer progression and that SLCO4A1-AS1 promotes malignant phenotypes of bladder cancer cells via the miRNA-335-5p/OCT4 axis.

To confirm the accuracy of the ceRNA network prediction presented in this study, interactions among lncRNAs, miRNAs and mRNAs in EAC were measured. Only cancer-associated lncRNAs and miRNAs with >2.0 FC and FDR <0.01 were selected. These non-coding genes were then annotated by GENCODE. Interactions among lncRNAs, miRNAs and mRNAs were predicted by experimentally conformed algorithms or by using miRDB, miRcode, miRTarBase and TargetScan databases. In the present study, cancer-associated lncRNAs in EAC were identified based on the RNA-Seq data of 79 EAC tissues and 11 normal tissues. Subsequently, cancer-associated miRNAs and mRNAs were identified. Eventually, interactions between lncRNAs, miRNAs and mRNAs was identified by constructing an lncRNA-miRNA-mRNA ceRNA network. A total of 37 DElncRNAs, five miRNAs and 13 mRNAs were selected to construct this newly-identified ceRNA-mediated gene regulatory network. This network included numerous active lncRNA-miRNA-mRNA interactions that may be used as prognostic biomarkers in EAC.

In conclusion, the present study identified some cancer-associated lncRNAs and revealed their potential use in prognosis prediction. In particular, some cancer-associated lncRNAs may serve as ceRNAs. The ceRNA network that was built in the present study may help understanding the mechanisms involved in the development and progression of EAC.

Acknowledgements
Not applicable.

Funding
No funding was received.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
SC and YY conceived and designed the study. XC acquired the data. YY analyzed and interpreted the data. YY and XC wrote and revised the manuscript. SC supervised the study. All authors read and approved the final manuscript.
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


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.