Original Article

Screening several novel biomarkers for predicting the survival of patients with lung adenocarcinoma

Mingzhu Li, Shenyu Wang

Department of Integrated Traditional Chinese and Western Medicine, Cancer Hospital of China Medical University, Liaoning Cancer Hospital & Institute, No.44 Xiaoheyan Road, Dadong District, Shenyang 110042, Liaoning Province, PR China

Received January 31, 2019; Accepted May 9, 2019; Epub August 15, 2019; Published August 30, 2019

Abstract: Background: Increasing evidence has confirmed that long non-coding RNAs (lncRNAs) could regulate protein levels of genes as well as the cellular biological behavior in the competing endogenous RNA (ceRNA) network as miRNA sponges. Our study aimed to explore the role and regulatory mechanisms of lncRNA-mediated ceRNA network in lung adenocarcinomas (LUAD), as well as their potential to predict the prognosis of LUAD. Methods: A total of 568 samples with LUAD were downloaded from The Cancer Genome Atlas (TCGA). Differentially expressed lncRNAs (DElncRNAs), miRNAs (DEmiRNAs) and mRNAs (DEmRNAs) were identified with EdgeR package in R according to \(|\log \text{fold change}| > 2\) and a corrected \(p\)-value < 0.05. DElncRNA-DEmiRNA pairs were predicted by the online data base of miRcode, while DEMiRNA-DEmiRNA pairs were predicted through “miRDB”, “miRTarBase” and “TargetScan”. And the ceRNA network was visualized through Cytoscape 3.6. After calculating the degrees of lncRNA nodes, we constructed the ceRNA sub-network. Finally, overall survival analysis was utilized to assess prognostic performance of differentially expressed genes in sub-network. Results: 1503 DElncRNAs, 118 DEMiRNAs and 2501 DEmRNAs were identified associated with LUAD. And the ceRNA network was constructed containing 120 lncRNA nodes, 23 miRNA nodes and 39 mRNA nodes. 5 lncRNAs including MEG3, AP002478.1, LINC00461, HOTTIP and MUC2 (degree > 12) were selected to build the sub-network. And two lncRNAs (AP002478.1 and MUC2) and seven mRNAs (E2F7, KPNA2, CEP55, HOXA10, CCNE1, CHEK1 and CLSPN) in the sub-network had poor prognosis according to overall survival analysis. Conclusion: MUC2 and AP002478.1 could become potential prognostic biomarkers for patients with LUAD. Based on the ceRNA hypothesis, we predicted that AP002478.1-induced AP002478.1-hsa-mir-144-HOXA10/KPNA2 and MUC2-hsa-mir-195-CCNE1/CEP55/CHEK1/CLSPN/E2F7/HOXA10 pathways are associated with LUAD and will be investigated for further experimental confirmation.

Keywords: Lung adenocarcinoma, MUC2, AP002478.1, prognosis, biomarkers, competitive endogenous RNA

Introduction

Lung cancer remains one of the most common malignant tumor types due to cancer-related habits such as smoking [1, 2]. Compared with other lung cancer subtypes, lung adenocarcinoma (LUAD) has a higher incidence and mortality [3, 4]. Moreover, LUAD has become more prevalent in Asian countries, especially in females and non-smokers [5]. Therefore, it is critical for early diagnosis or the development of novel treatment targets in order to prolong the survival time of LUAD patients. However, excessively frequent low-dose CT screening can lead to associated harm for high-risk individuals [6]. And the application of molecular biomarkers contributes to the early detection of lung cancer and improves the screening ability of patients. At the same time, individuals can avoid the hazards associated with low-dose CT screening by using molecular biomarkers [7]. Therefore, it is necessary to explore biomarkers with high specificity and sensitivity for clinical detection.

Long non-coding RNAs (> 200 nucleotides) play a critical role in regulating gene expression including translation, epigenetic modifications, splicing and mRNA stability [8-10]. Abnormal lncRNAs are closely related to the occurrence of cancer. In the past few years, many studies have focused on finding abnormal lncRNAs rel-
Screening novel biomarkers of lung adenocarcinoma

relevant to LUAD. However, most of these studies pay close attention to the function of a single lncRNA in LUAD. And there are few studies on the mechanisms of lncRNAs interacting with other RNAs in LUAD. Recent studies have found that lncRNA is involved in the regulation of target genes in tumors as a competitive endogenous RNA (ceRNA) sponging up microRNAs (miRNAs) [11, 12]. Specifically, lncRNAs could competitively bind to shared miRNA response elements (MREs) through miRNA in 3'-untranslated regions (UTRs) [13]. And miRNAs can cause gene silencing by binding to mRNAs. Therefore, the pattern of mutual regulation among lncRNA, miRNA and its downstream target genes is closely related to the occurrence and development of LUAD. Therefore, our study aimed to explore the function of lncRNA as a ceRNA in LUAD.

In our study, we collected RNA profile data from The Cancer Genome Atlas (TCGA). We constructed a lncRNA-mediated ceRNA network and sub-network in LUAD. We explored the biological process of the mRNAs in a lncRNA-mediated ceRNA sub-network by Cytoscape plug-in BinGO. At the same time, prognosis value of differentially expressed RNAs (DEGs) was evaluated according to the overall survival analysis. The present study results illustrate the function of lncRNAs via the lncRNA-miRNA-mRNA ceRNA network and sub-network in LUAD and provide novel lncRNAs as potential prognostic biomarkers for clinicians.

Materials and methods

Sample collection and difference analysis

RNA sequencing (RNA-Seq) data of LUAD patients were downloaded from TCGA (https://cancergenome.nih.gov/). A total of 568 samples with LUAD were included in our study. The lncRNA and mRNA profile data included 58 normal samples and 510 primary samples. miRNA profile data included 45 normal samples and 510 primary LUAD samples. The exclusion criteria were as follows: (1) histological diagnosis negating LUAD; (2) presence of a malignancy other than LUAD; (3) lack of complete clinical information. The gene expression profiles of 510 primary LUAD samples and 58 normal samples, and miRNA data of 510 primary LUAD samples and 45 normal samples were retrieved from TCGA data portal. Therefore, approval by the ethics committee was not needed.

Gene differential expression analysis

The differentially expressed lncRNAs (DElncRNAs), miRNAs (DEMiRNAs) and mRNAs (DEmRNAs) were identified between LUAD and normal samples for further analysis using EdgeR package in R 3.4.3 [14]. The differentially expressed genes (DEGs) between LUAD and normal samples met the following selection criteria: |log fold change (FC)| > 2 and a corrected p-value < 0.05. Volcano plots were plotted utilizing the ggplot2 package, and the heat map was visualized by use of the heatmap package in R 3.4.3 [15].

The ceRNA network construction

First, the DElncRNA-DEMmiRNA pairs were predicted by the online database of miRcode (http://www.mircode.org/) [16]. Then the target DEMiRNAs of DEMiRNAs were predicted through “miRDB” (http://www.mirdb.org/), “miRTarBase” (http://miRTarBase.mbc.nctu.edu.tw/) and “TargetScan” (http://www.targetscan.org/) [17-19]. Finally, the DElncRNA-DEMmiRNA-DEmRNA network was constructed by assembling the DEMiRNAs that were regulated by DElncRNAs and DEMiRNAs simultaneously. The ceRNA network was visualized through Cytoscape 3.6 [20].

Key lncRNA-miRNA-mRNA sub-network

In order to predict the role of hub lncRNAs in ceRNA networks, the DElncRNA-DEMmiRNA-DEmRNA sub-network was built. After calculating the degree of lncRNAs in the ceRNA network, lncRNAs with degree > 12 were selected to construct the ceRNA sub-network by Cytoscape plug-in MCODE.

Functional enrichment analysis

To better predict the function of abnormally expressed DEGs in the ceRNA sub-network, biological processes of Gene Ontology (GO) for DEmRNAs were built utilizing Cytoscape plug-in BinGO. GO terms were based on the threshold of p-value < 0.01.

Statistical analysis

The overall survival analysis was performed by survival and qvalue packages in R 3.4.3. And the log-rank test was used to evaluate the dif-
Screening novel biomarkers of lung adenocarcinoma

Figure 1. 2501 DEmRNAs, 1503 DEIncRNAs and 118 DEmiRNAs in volcano plots (A-C) and heat maps (D-F) between LUAD samples and normal samples. (Red represents up-regulated DEGs and green stands for down-regulated DEGs).

A p-value < 0.05 was considered statistically significant.
Table 1. Predicted LUAD specific miRNA-mRNA pairs

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>mRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsa-mir-143</td>
<td>COL1A1, COL5A2</td>
</tr>
<tr>
<td>Hsa-mir-144</td>
<td>FGF2, HOXA10, KCNQ5, KPNA2, TBX18</td>
</tr>
<tr>
<td>Hsa-mir-182</td>
<td>BDNF, NPTX1</td>
</tr>
<tr>
<td>Hsa-mir-183</td>
<td>CCNB1</td>
</tr>
<tr>
<td>Hsa-mir-195</td>
<td>CFX2, CCNE1, CDC25A, CEP55, CHEK1, CLSPN, E2F7, FGF2, HOXA10, KIF23, OSCAR, PSAT1, RET, RS1, SALL1, TGFBR3, TMEM100</td>
</tr>
<tr>
<td>Hsa-mir-200a</td>
<td>CEN2, DLC1, ELAVL2</td>
</tr>
<tr>
<td>Hsa-mir-205</td>
<td>LRRK2</td>
</tr>
<tr>
<td>Hsa-mir-210</td>
<td>SERTM1</td>
</tr>
<tr>
<td>Hsa-mir-216b</td>
<td>MCM4</td>
</tr>
<tr>
<td>Hsa-mir-31</td>
<td>HOXC13, SELE, ELAVL2, KPNA2, MIXL1, PFKP, SLC7A11, TMEM100</td>
</tr>
<tr>
<td>Hsa-mir-372</td>
<td>ELAVL2, KPNA2, MIXL1, PFKP, SLC7A11, TMEM100</td>
</tr>
<tr>
<td>Hsa-mir-373</td>
<td>ELAVL2, KPNA2, MIXL1, PBK, PFKP, SLC7A11, TMEM100</td>
</tr>
<tr>
<td>Hsa-mir-96</td>
<td>PROK2, SLC1A1</td>
</tr>
</tbody>
</table>

Figure 2. The LUAD specific ceRNA network. (Round stands for lncRNAs, triangular stands for mRNAs and rectangular stands for miRNAs. Red represents up-regulated genes and blue represents down-regulated genes).
Results

DEmRNAs, DEIncRNAs and DEMiRNAs

2501 DEmRNAs, 1503 DEIncRNAs and 118 DEMiRNAs were identified between LUAD and normal samples with \(|\log FC| > 2\) and FDR adjusted \(p\)-value < 0.05. The volcano plots and heat maps showed the DEGs results in Figure 1A-F, respectively. And these DEGs were listed in the raw data. Of these, 1958 DEmRNAs, 1296 DEIncRNAs and 98 DEMiRNAs were up-regulated.

Figure 1 shows 2501 DEmRNAs, 1503 DEIncRNAs and 118 DEMiRNAs in volcano plots (Figure 1A-C) and heat maps (Figure 1D-F) between LUAD samples and normal samples (Red represents up-regulated DEGs and green stands for down-regulated DEGs).

Construction of the ceRNA network

DEIncRNA-DEmiRNA pairs were predicted through “miRcode”. Twenty-three DEmiRNAs putatively targeted 120 DEIncRNAs. Then we took advantage of “miRDB”, “miRTarBase” and “TargetScan” to match the DEMiRNAs and DEmiRNAs. The results listed 13 DEMiRNAs targeted 39 DEmiRNAs in LUAD (Table 1). Based on IncRNA-miRNA and mRNA-miRNA pairs above, we constructed a IncRNA-miRNA-mRNA network containing 120 IncRNA nodes (102 up, 18 down), 23 miRNA nodes (19 up and 4 down), and 39 mRNA nodes (27 up, 12 down) (Figure 2).

Table 1 shows the predicted LUAD specific miRNA-mRNA pairs. Figure 2 shows the LUAD specific ceRNA network. (Round stands for IncRNAs, triangular stands for mRNAs and rectangular stands for miRNAs. Red represents up-regulated genes and blue represents down-regulated genes).

Key IncRNA-miRNA-mRNA sub-network and functional prediction of ceRNA sub-network

We calculated the degrees of all IncRNA nodes in the ceRNA network and the degree distribution of nodes was performed in Figure 3. We selected IncRNA nodes by degree > 12, and 5 qualified IncRNAs (MEG3, AP002478.1, LINCO0461, HOTTIP, MUC2) were extracted to construct the ceRNA sub-network (Figure 4). There were 5 IncRNAs (Table 2), 39 mRNAs (Table 3) and 22 miRNAs (Table 4) in the ceRNA sub-network. To explore the role of the abnormal ceRNA sub-network, we made biological processes of GO enrichment analysis for all 39 DEmiRNAs in the sub-network (Figure 5). Table 5 listed the top five GO analysis terms in the biological process of DEmiRNAs related with LUAD.

Figure 3 shows the degree distribution of IncRNA nodes in the ceRNA network. Figure 4 shows the key IncRNA-miRNA-mRNA in the sub-network. (Round stands for IncRNAs, triangular stands for mRNAs and rectangular stands for miRNAs. Red represents up-regulated genes and blue represents down-regulated genes).

Figure 5 illustrates the GO analysis terms in the biological process of DEmiRNAs in the ceRNA sub-network related with LUAD. (Yellow nodes: nodes with corrected \(P\)-value < 0.05). Table 2 lists the DEIncRNAs in the ceRNA sub-network. Table 3 lists the DEmiRNAs in the ceRNA sub-network. Table 4 lists the DEmiRNAs in the ceRNA sub-network. Table 5 lists the top five GO analysis terms in the biological process of DEmiRNAs related with LUAD.

Prognostic overall survival assessment of DEIncRNAs, DEMiRNAs and DEmRNAs in the ceRNA sub-network

To assess the prognostic value of DEIncRNAs, DEMiRNAs and DEmiRNAs in the ceRNA sub-network, the overall survival analysis was performed using log-rank test (Figure 6). The results revealed that two IncRNAs including AP-
Seven mRNAs including E2F7, KPNA2, CEP55, HOXA10, CCNE1, CHEK1 and CLSPN of high expression were negatively associated with overall survival time.

**Figure 6** shows Kaplan-Meier survival curves of the 2 lncRNAs and 7 mRNAs correlated with overall survival in LUAD.

**Discussion**

LUAD is currently the most common histological subtype of lung cancer [21]. Although sur-
Screening novel biomarkers of lung adenocarcinoma

In recent years, growing evidence has emphasized the lncRNA-induced ceRNA network in cancer development and progression. LncRNAs could competitively bind to miRNA, while miRNAs can lead to gene silencing by binding to mRNAs. However, the mechanisms of the ceRNA network in LUAD still remain unclear. In our study, we identified 2501 DEmRNAs, 1503 DElncRNAs and 118 DEmiRNAs related to LUAD from TCGA database. Based on DElncRNA-DEmiRNA pairs and DEmRNA-DEmiRNA pairs, we constructed a lncRNA-miRNA-mRNA network. Furthermore, under the calculation of the degree of lncRNA nodes, we selected 5 lncRNAs with degrees more than 12 to construct sub-network including MEG3, AP002478.1, LINCO0461, HOTTIP and MUC2. To explore the function of DEmRNAs in the sub-network, we made GO enrichment analysis. The results revealed that these DEmRNAs are mainly involved in anatomical structure morphogenesis, anatomical structure formation involved in morphogenesis, tissue development, organ...
Screening novel biomarkers of lung adenocarcinoma

Figure 5. GO analysis terms in the biological process of DEmRNAs in the ceRNA sub-network related with LUAD. (Yellow nodes: nodes with corrected $P$-value $< 0.05$).
Table 5. The top five GO analysis terms in the biological process of DEmRNAs related with LUAD

<table>
<thead>
<tr>
<th>GO-ID</th>
<th>GO terms</th>
<th>Genes</th>
<th>Corrected p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>9653</td>
<td>anatomical structure morphogenesis</td>
<td>RET</td>
<td>BDNF</td>
</tr>
<tr>
<td>48646</td>
<td>anatomical structure formation involved in morphogenesis</td>
<td>RET</td>
<td>COL1A1</td>
</tr>
<tr>
<td>9888</td>
<td>tissue development</td>
<td>RET</td>
<td>COL1A1</td>
</tr>
<tr>
<td>9887</td>
<td>organ morphogenesis</td>
<td>RET</td>
<td>COL1A1</td>
</tr>
<tr>
<td>48729</td>
<td>tissue morphogenesis</td>
<td>RET</td>
<td>COL1A1</td>
</tr>
</tbody>
</table>
morphogenesis and tissue morphogenesis and so on. We assessed the prognostic value of DEmRNAs, DEIncRNAs and DEMiRNAs in the sub-network via the overall survival analysis. As for 5 hub lncRNA in the sub-network, AP002478.1 and MUC2 of high expression have negative impact on survival time of patients as potential prognostic biomarkers, while MEG3, LINC00461 and HOTTIP have no effect on overall survival. Recently, a study found that MEG3 may interact with miR-106 regulating MAPK9 in a LUAD specific ceRNA network acting in MAPK signaling pathways [23]. However, the role of LINC00461 in LUAD is unclear, while we only found that overexpressed LINC00461 is a hub gene in the sub-network. Only one literature has reported that LINC00461 has a carcinogenic effect on glioma [24]. With regard to HOTTIP, overexpressed HOTTIP can promote the progression of LUAD [25].

In the ceRNA sub-network, we found that upregulated IncRNAs AP002478.1 and MUC2 are both closely associated with poor prognosis. A previous study has reported that AP002478.1 of high expression could be negatively associated with the overall survival time of patients with Helicobacter pylori (+) gastric cancer [26]. In our study, we found a negative correlation between AP002478.1 and hsa-mir-144, which

Figure 6. Kaplan-Meier survival curves of 2 IncRNAs and 7 mRNAs correlated with overall survival in LUAD.
suggests that AP002478.1 of high expression could inhibit the expression level of hsa-mir-144 by sponging hsa-mir-144. It has been reported that hsa-miR-144-3p could suppress LUAD cell proliferation through the IL-1β/miR-144-3p/WT1D pathway and highly expressed miR-144-3p has a better prognosis for LUAD patients [27]. As for the targeted mRNAs of hsa-mir-144, up-regulated KPNA2 and HOXA10 have poor prognosis for LUAD patients. HOXA10 gene expression could be associated with the pathogenesis of LUAD as an oncogenic target gene [28]. And ELK1-induced up-regulation of HOXA10-AS induces LUAD progression via elevating Wnt/β-catenin pathway [29]. Our results also revealed that HOXA10 was up-regulated in LUAD and up-regulation of HOXA10 significantly decreased the survival time of patients. Overexpressed KPNA2 is involved in the LUAD cell invasion [30]. According to the ceRNA sub-network, we concluded there may exist a AP002478.1-induced AP002478.1-hsa-mir-144-HOXA10/KPNA2 pathway related with LUAD.

As for MUC2, up-regulated MUC2 is the target gene of down-regulated hsa-mir-195. The targeted four mRNAs genes of hsa-mir-195 including CCNE1, CEP55, CHEK1, CLSPN, E2F7 and HOXA10 are elevated in the ceRNA network. At the same time, overexpressed mRNAs are related with poor prognosis for patients with LUAD. Non-small-cell lung cancer (NSCLC) occupies approximately 85% of lung cancer, of which LUAD accounts for 60% [31]. Highly expressed hsa-mir-195 inhibits proliferation, migration and invasion of NSCLC cells [32]. Hsa-mir-195 has down-regulated expression in the developing lung-like LUAD subtype, whereas the introduction of miR-195 into the lung cancer cell lines could evoke the changes of mRNA and LUAD subtypes [33]. CCNE1 is obviously overexpressed in LUAD, and G1-phase is arrested for LUAD cells through inducing the down-regulation of CCNE1 [34, 35]. CHEK1 has been identified as the hub gene in the p53 signaling pathway, which is related to LUAD tumorigenesis. CHEK1 could become the potential drug target of daunorubicin, mycophenolic acid, and pyrvinium [36]. At the same time, IncRNA SNHG6 could regulate E2F7 expression by sponging miR-26a-5p in LUAD [37]. Therefore, we inferred that MUC2 as a ceRNA could induce the abnormal expression of CCNE1, CEP55, CHEK1, CLSPN, E2F7 and HOXA10 by sponging hsa-mir-195.

Therefore, IncRNAs AP002478.1 and MUC2 are overexpressed in LUAD and considered as hub genes in the ceRNA sub-network. And overexpressed AP002478.1 and MUC2 are closely related with poor prognosis of patients with LUAD. Hence, AP002478.1 and MUC2 could become potential pathogenic genes, which could have prognostic value. We discovered that AP002478.1-induced AP002478.1-hsa-mir-144-3p-HOXA10/KPNA2 and MUC2-induced MUC2-hsa-mir-195-CCNE1/CEP55/CHEK1/CLSPN/E2F7/HOXA10 pathways are related with LUAD. However, the conclusion still needs further experimental confirmation. Our laboratory will be dedicated to this research.

**Conclusion**

In our study, we identified differentially expressed IncRNAs, miRNAs and mRNAs in LUAD from TCGA. Then a LUAD specific ceRNA network was constructed based on DElncRNA-DEmiRNA and DEmiRNA-DEmRNA pairs. Five IncRNAs including MEG3, AP002478.1, LINCO0461, HOTTIP, MUC2 (degree > 12) were selected to build a ceRNA sub-network after calculating the degree of all IncRNA nodes. According to the overall survival analysis, MUC2 and AP002478.1 could be associated with poor prognosis as potential biomarkers for patients with LUAD. Based on ceRNA hypothesis, we predicted that AP002478.1-induced AP002478.1-hsa-mir-144-3p-HOXA10/KPNA2 and MUC2-induced MUC2-hsa-mir-195-CCNE1/CEP55/CHEK1/CLSPN/E2F7/HOXA10 pathways are associated with LUAD and are open for further experimental confirmation.

**Disclosure of conflict of interest**

None.

**Abbreviations**

IncRNAs, long non-coding RNAs; TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinomas; DElncRNAs, differentially expressed IncRNAs; DEMiRNAs, differentially expressed miRNAs; DEmRNAs, differentially expressed mRNAs; ceRNA, competitive endogenous RNA; miRNAs, microRNAs; DEGs, differentially expressed RNAs.
Screening novel biomarkers of lung adenocarcinoma

Address correspondence to: Shenyu Wang, Department of Integrated Traditional Chinese and Western Medicine, Cancer Hospital of China Medical University, Liaoning Cancer Hospital & Institute, No.44 Xiaoheyuan Road, Dadong District, Shenyang 110042, Liaoning Province, PR China. E-mail: 148-003642@qq.com

References


Screening novel biomarkers of lung adenocarcinoma


