Expression profiles of long noncoding RNAs in lung adenocarcinoma

Bing Zhao
Hang Xu
Xiaoye Ai
Yasheng Adalat
Ying Tong
Jiazhennan Zhang
Shune Yang

1Department of Day Oncology, Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi, People’s Republic of China; 2Intensive Care Unit, The Sixth Affiliated Hospital, Xinjiang Medical University, Urumqi, Xinjiang, People’s Republic of China; 3Department of Breast Cancer and Lymphoma, Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi, People’s Republic of China

Purpose: This study aimed to analyze expression profiles of long noncoding RNAs (lncRNAs) in lung adenocarcinoma.

Methods: lncRNA microarray technology was employed to detect lncRNA profiles of 3 pairs of lung adenocarcinoma tissues and adjacent tissues.

Results: We found 134 upregulated lncRNAs and 460 downregulated lncRNAs in lung adenocarcinoma tissues compared to adjacent tissues. Among them, LINC00152, LINC00691, and LINC00578 showed the most significant changes of upregulation, while LINC00668, LINC00710, and LINC00607 showed the most significant changes of downregulation. Fluorescent quantitative polymerase chain reaction (PCR) analysis of tissue samples from an additional 90 patients with lung adenocarcinoma showed significantly increased levels of LINC00152, LINC00691, and LINC00578 and decreased levels of LINC00668, LINC00710, and LINC00607 in lung adenocarcinoma tissues. In addition, LINC00578 was closely associated with the existence of metastasis of lung adenocarcinoma, but the other 5 lncRNAs showed no significant correlation with clinicopathologic characteristics such as age, gender, tumor stage, and the existence of metastasis. Further follow-up study showed that LINC00578 expression was closely associated with the survival of patients with lung adenocarcinoma.

Conclusion: We revealed the expression profiles of IncRNAs in lung adenocarcinoma and identified LINC00578 as a promising biomarker and therapeutic target for lung adenocarcinoma.

Keywords: long noncoding RNA, microarray, lung adenocarcinoma, survival

Introduction

Lung cancer is one of the malignant tumors with the biggest threat to human health and life. In the last 50 years, there has been significant increase in the morbidity and mortality of lung cancer worldwide. The morbidity and mortality of male lung cancer rank at the top of the list all malignant tumors, while the morbidity and mortality of female lung cancer rank the second. About 80%-90% of patients are at advanced stage and have distant metastasis when lung cancer is detected, and thus are not suitable for radical surgeries. Among pathological types, lung adenocarcinoma accounts for 40% of all non-small-cell lung cancers, and it metastasizes at an early stage, has poor prognosis, and has low treatment efficacy. Therefore, it is particularly important to elucidate molecular mechanisms for the metastasis of lung adenocarcinoma in order to develop novel approaches to early diagnosis and effective treatment of lung adenocarcinoma.

Long noncoding RNAs (lncRNAs) are noncoding RNAs >200 nucleotides long. lncRNAs play important role in dosage compensation, epigenetic control, cell cycle control, and cell differentiation. The abnormal expression or function of lncRNAs is closely related to the occurrence of human diseases such as cancer and neurodegenerative disease. However, current literature on the role of lncRNAs in malignant lung adenocarcinoma...
is rare. This study aimed to compare lncRNA profiles between lung adenocarcinoma tissues and paired adjacent tissues using lncRNA microarray technology and identify lncRNAs associated with the occurrence of lung adenocarcinoma as novel biomarkers and therapeutic targets for lung adenocarcinoma.

Materials and methods

Patients and tissues
All patients gave written informed consent and the study was approved by Ethics Committee of Tumor Hospital Affiliated to Xinjiang Medical University. Lung adenocarcinoma tissues and normal adjacent tissues were dissected from 3 patients who underwent lung adenocarcinoma surgery from July 2012 to June 2015 for microarray analysis, and lung adenocarcinoma was pathologically verified. Patients with advanced lung adenocarcinoma with the complication of liver metastasis, brain metastasis, bone metastasis, and abdominal cavity metastasis were excluded because these conditions affected their survival and interfered with the analysis of the correlation of lncRNAs with clinicopathologic characteristics. Patients who had previously received radical lung adenocarcinoma surgery, radiotherapy, and chemotherapy were excluded. For verification study, an additional 90 patients with lung adenocarcinoma were recruited and their data are listed in Table 1.

Microarray analysis
Microarray of lncRNAs was performed by Beijing CapitalBio Technology Co., Ltd (Beijing, People’s Republic of China). lncRNAs from all canonical databases, NCBI Refseq, UCSC Known Gene, NRED, and RNAdb, were covered with a 12×135 k lncRNA microarray manufactured by Arraystar (Rockville, MD, USA). The probes of microarray were made of 60 mer long oligonucleotides. More than 1 probe was designed for each sequence to increase the reliability of signals.

Total RNA was extracted from tissue samples using Trizol (Invitrogen, Carlsbad, CA, USA) at the ratio of 50 mg sample/mL Trizol. The quality of RNA was analyzed by formaldehyde-denatured gel electrophoresis. Reverse transcription was conducted and ds-cDNA was synthesized using a ds-cDNA synthesis kit (Invitrogen). ds-cDNA was labeled using a monochromatic DNA labeling kit (NimbleGen, Madison, WI, USA). The efficiency of fluorescent labeling was tested using a spectrophotometer. Later, the labeled ds-cDNAs were sequenced and hybridized using a hybridization system (NimbleGen). The fluorescence intensity of microarray was scanned using Axon GenePix Scanner 4000B (Axon Instruments, Union City, CA, USA) and analyzed with NimbleScan software (Ver 2.5) (NimbleGen). GO and pathway analysis was performed using DAVID bioinformatic

Table 1 Clinical data of 90 patients

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total</th>
<th>Stage I</th>
<th>Stage II</th>
<th>Stages III + IV</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>90</td>
<td>24</td>
<td>21</td>
<td>45</td>
<td>0.237</td>
</tr>
<tr>
<td>Median age (year)</td>
<td></td>
<td>62</td>
<td>63</td>
<td>63</td>
<td>0.315</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>M 51</td>
<td>12</td>
<td>15</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F 39</td>
<td>12</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.168</td>
</tr>
<tr>
<td>Left lung</td>
<td>41</td>
<td>12</td>
<td>9</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Right lung</td>
<td>49</td>
<td>12</td>
<td>12</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Diameter of tumor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.373</td>
</tr>
<tr>
<td>≥5 cm</td>
<td>21</td>
<td>0</td>
<td>6</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>&lt;5 cm</td>
<td>69</td>
<td>24</td>
<td>15</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Degree of differentiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.412</td>
</tr>
<tr>
<td>High + medium</td>
<td>50</td>
<td>17</td>
<td>7</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>40</td>
<td>7</td>
<td>14</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>EGFR state</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.097</td>
</tr>
<tr>
<td>M+</td>
<td>62</td>
<td>18</td>
<td>12</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>M–</td>
<td>28</td>
<td>6</td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Smoking history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.401</td>
</tr>
<tr>
<td>Yes</td>
<td>36</td>
<td>6</td>
<td>9</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>54</td>
<td>18</td>
<td>12</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>CEA</td>
<td>3.95</td>
<td>2.58</td>
<td>4.24</td>
<td>5.37</td>
<td>0.049</td>
</tr>
<tr>
<td>CYFRA21-1</td>
<td>2.30</td>
<td>1.90</td>
<td>2.35</td>
<td>2.50</td>
<td>0.237</td>
</tr>
<tr>
<td>CA125</td>
<td>9.10</td>
<td>5.30</td>
<td>9.50</td>
<td>9.10</td>
<td>0.567</td>
</tr>
</tbody>
</table>

Abbreviations: CA125, cancer antigen 125; CEA, carcino-embryonic antigen; CYFRA21-1, cytokeratins 21-1; EGFR, epidermal growth factor receptor.
resources 6.7 (http://david.abcc.ncifcrf.gov) and KEGG Pathway Database (http://www.genome.jp/kegg/pathway.html) to predict IncRNA function and the annotation of coexpressed genes.

Construction of coexpression network

Coexpression analysis of the coding–noncoding genes that included the differentially expressed IncRNAs and nearby coding genes (distance <300 kb) was conducted in R environment using RedeR package as described previously.11 Pearson’s correlation coefficient was not 0.94. The P-values for enrichment between differentially expressed IncRNAs and transcription factors (TFs) was calculated using the hypergeometric distribution method, and the recommended P-value cut-off was 0.05.

Real-time polymerase chain reaction (PCR)

Tumor tissues and adjacent tissues were excised from 90 lung adenocarcinoma patients and resuspended in Trizol (Invitrogen) at the ratio of 50 mg sample/mLTrizol. Tissues were homogenized and total RNA was extracted following the manufacturer’s protocols. Reverse transcription was conducted using Quant Script RT kit (Takara, Kyoto, Japan) according to the user’s manual. Real-time PCR was performed for LINC00578, LINC00668, LINC00710, LINC00607, LINC00152, LINC00691, and GAPDH using SYBR Green Real-time PCR Masker Mix kit (Takara) on 7900HT fluorescent quantitative PCR machine (ABI). Primers are listed in Table 2. GAPDH was used as the internal control, and the fold change was calculated.

Statistical analysis

Statistical analyses were accomplished using SAS software 9.3 (SAS Institute Inc., Cary, NC, USA), with all data expressed as means ± SD. The correlation between IncRNA and clinical and pathological characteristics was analyzed using Pearson χ² test. The Student’s t-test was used to evaluate the significance of difference between groups. P<0.05 was considered as significant.

Results

Microarray analysis

Microarray analysis of 3 pairs of qualified samples from lung adenocarcinoma tissues and adjacent tissues showed that a total of 6,960 IncRNAs were detected. After clustering analysis and comparison, 594 IncRNAs were differentially expressed. Among them, 134 IncRNAs were upregulated and 460 IncRNAs were downregulated in tumor tissues compared to adjacent tissues. Notably, 12 IncRNAs had levels increased by >5 times and 121 IncRNAs had levels decreased by >5 times in tumor tissues than in adjacent tissues (Tables 3 and 4).

<table>
<thead>
<tr>
<th>Target ID</th>
<th>Primers</th>
<th>Product length (bp)</th>
<th>Anneal temperature (°C)</th>
</tr>
</thead>
</table>
| LINC00152 | Forward: CATTGGAATGGAGGGAAAT  
Reverse: CTTCCAGGAACCTGCTGT | 120 | 60 |
| LINC00691 | Forward: AGGGATTCTGGGCTACGAGT  
Reverse: ATGTCGCTCTTCTGCCCCTTG | 90 | 60 |
| LINC00578 | Forward: GAAGACCGAAAGTTGCGAG  
Reverse: CACCACCAACACAAAGAT | 80 | 60 |
| LINC00668 | Forward: GCAATCTCCTGCTCTGAGT  
Reverse: CATTCTCTGCACTTCCTAT | 110 | 60 |
| LINC00710 | Forward: AAAGATGCCTCCTCCCTGATG  
Reverse: CAATGGCCTCCCTGCTTTC | 115 | 60 |
| LINC00607 | Forward: ATGCCCTCTGCTGCAAACCT  
Reverse: TAATGGTGGTGGTGGAACC | 180 | 60 |
| GAPDH   | Forward: CAGAGGCTAATGCTGATG  
Reverse: GAAGCTGGGCTTCATT | 100 | 60 |

Table 2 Primers used for real-time polymerase chain reaction (PCR)

### Table 3 Top 10 upregulated IncRNAs with fold change >5 in tumor tissues

<table>
<thead>
<tr>
<th>Number</th>
<th>IncRNA name</th>
<th>Fold change</th>
<th>Chrom</th>
<th>Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ENST00000415479.1</td>
<td>27.92906</td>
<td>chr2</td>
<td>Up</td>
</tr>
<tr>
<td>2</td>
<td>CARIntergenic103FLJ31066</td>
<td>15.07073</td>
<td>None</td>
<td>Up</td>
</tr>
<tr>
<td>3</td>
<td>ENST00000578206.1</td>
<td>10.11382</td>
<td>chr17</td>
<td>Up</td>
</tr>
<tr>
<td>4</td>
<td>HIT000310840</td>
<td>9.037025</td>
<td>chr16</td>
<td>Up</td>
</tr>
<tr>
<td>5</td>
<td>TCONS_00025758</td>
<td>7.797149</td>
<td>chr17</td>
<td>Up</td>
</tr>
<tr>
<td>6</td>
<td>ENST00000583521.1</td>
<td>6.906059</td>
<td>chr19</td>
<td>Up</td>
</tr>
<tr>
<td>7</td>
<td>ENST00000436484.1</td>
<td>6.899019</td>
<td>chr1</td>
<td>Up</td>
</tr>
<tr>
<td>8</td>
<td>ENST00000513899.1</td>
<td>6.744434</td>
<td>chr5</td>
<td>Up</td>
</tr>
<tr>
<td>9</td>
<td>ENST00000569808.1</td>
<td>6.71074</td>
<td>chr15</td>
<td>Up</td>
</tr>
<tr>
<td>10</td>
<td>TCONS_00029766</td>
<td>6.192585</td>
<td>chr22</td>
<td>Up</td>
</tr>
</tbody>
</table>

Abbreviations: Chrom, chromosome; IncRNA, long noncoding RNA.
To verify microarray results, we focused on 6 lncRNAs which showed the most significant changes of upregulation or downregulation: LINC00152, LINC00691, LINC00578, LINC00668, LINC00710, and LINC00607, and employed quantitative PCR to detect their expression in 90 pairs of lung adenocarcinoma tissues and adjacent tissues. The expression of LINC00152, LINC00691, and LINC00578 was increased significantly, while the expression of LINC00668, LINC00710, and LINC00607 was decreased significantly in lung adenocarcinoma tissues compared to adjacent tissues (Figure 1). Thus, both microarray and PCR analysis showed the upregulated expression of LINC00152, LINC00691, and LINC00578 and the downregulated expression of LINC00668, LINC00710, and LINC00607.

Pearson correlation analysis showed that LINC00578 was closely associated with the metastasis of lung adenocarcinoma \((r=0.92)\). However, there was no significant correlation of the clinicopathologic characteristics such as age, gender, tumor stage, and metastasis with the expression levels of the other 5 lncRNAs \((r<0.864)\) (Table 5).

### Bioinformatics analysis

Using GO cluster analysis, we found that lncRNAs upregulated in lung adenocarcinoma tissues were involved in a variety of malignant phenotypes related to tumorigenesis, such as cell cycle control and mitotic control (Figure 2). In addition, we found several signaling pathways in which upregulated lncRNA are involved (Figure 3).

### Construction of the coexpression network

We constructed a coexpression network composed of 641 nodes and 571 connections between 109 differentially expressed lncRNAs and 532 coding genes. Notably, most of these potential transregulatory lncRNAs could participate in pathways regulated by 3 transcription factors (TFs): NF-κB, SPI1, and BATF. In the core network of 1,092 lncRNA–TF pairs, NF-κB participates in 222 pairs, SPI1 participates in 439 pairs, and BATF participates in 431 pairs (Figure 4). When the top 300 coexpression pairs in \(P\)-value order were clustered in lncRNA–TF–mRNA network, NF-κB was identified as the transcription regulator that plays a central role in signaling pathways (Figure 5).

### Relationship between LINC00578 expression and patient survival

Of the 90 patients with lung adenocarcinoma, the median follow-up was 34 months (range, 4–60 months), with 52 patients alive and 38 cancer-related deaths at the last clinical follow-up. The median disease-free survival (DFS) and overall survival (OS) were 28 and 32 months, respectively. The 1-, 3-, and 5-year survival rates for the entire cohort of patients were 85.6%, 62.2%, and 13.3%, respectively. By Kaplan–Meier analysis, LINC00578 expression was closely associated with DFS and OS. For the whole cohort, median DFS was longer for patients with low LINC00578 expression than for patients with high LINC00578 expression (34 vs 22 months; \(P=0.008\); Figure 6A). Median OS was also longer for patients with low LINC00578 expression than for
patients with high LINC00578 expression (37 vs 25 months; $P=0.031$; Figure 6B).

**Discussion**

With the development of genome sequencing technology, it was found that >90% of human genomic DNA can be transcribed into RNAs, of which only 2% encode proteins. LncRNAs account for 80% of all noncoding RNAs. Different from micro-RNAs (miRNAs), lincRNAs refer to noncoding RNAs with length >200 nt, and they have a specific and complex secondary structure which can provide many loci to bind proteins and interact with DNA and RNA specifically.
Figure 4 LncRNA–TF core network consisting of the top 100 pairs of lncRNAs and TFs with the most relevance. 

Notes: The blue squares represent lncRNAs, and yellow triangles represent TFs. The edges between them mean that the lncRNAs are potentially regulated by the TFs.

Abbreviations: lncRNAs, long noncoding RNAs; TF, transcript factor.

Figure 5 LncRNA–TF–mRNA core network consisting of the top 300 pairs of lncRNA, TF, and mRNA with the most relevance.

Note: The red squares represent lncRNAs, yellow squares represent TF, and blue squares represent mRNAs.

Abbreviations: lncRNAs, long noncoding RNAs; TF, transcript factor.
and dynamically through complementary base pairing, thus forming a complex and delicate gene expression control network.\textsuperscript{14–16}

Recent studies have shown that lncRNAs play an important role in tumorigenesis, such as the regulation of invasion and metastasis of malignant tumors.\textsuperscript{17–20} In particular, a novel lncRNA MUC5B-AS1 was identified to be upregulated in lung adenocarcinoma tissues and promote metastasis of lung adenocarcinoma through regulating MUC5B expression.\textsuperscript{21} A recent study reported that LINCO00707 promoted lung adenocarcinoma cell proliferation and migration through regulating Cdc42 expression.\textsuperscript{22} Another recent study reported that overexpression of LINCO0319 indicated poor prognosis of lung adenocarcinoma patients, and LINCO0319 contributed to lung adenocarcinoma progression by modulating miR-450b-5p/\textit{EZH2} pathway.\textsuperscript{23} However, only few studies have analyzed lncRNA profiles in cancers on a large scale, such as high-throughput RNA-seq sequencing.\textsuperscript{24} Currently, the molecular mechanism by which lncRNAs participate in lung adenocarcinoma remains elusive. Therefore, it is of great significance to explore lncRNAs involved in lung adenocarcinoma to provide a new strategy for treating lung adenocarcinoma.

In this study, we collected 3 pairs of lung adenocarcinoma tissues and adjacent tissues and analyzed the differences in lncRNA profiles. The results showed that 594 lncRNAs changed significantly (fold change >2), of which 134 were upregulated and 460 were downregulated. In particular, the levels of 12 lncRNAs increased by >5 times and the levels of 121 lncRNAs decreased by >5 times, indicating that these lncRNAs are involved in the tumorigenesis of lung adenocarcinoma.

lncRNAs regulate the expression and modification of coding genes in a variety of ways, including epigenetic control, transcriptional control, post-transcriptional control, and cytoplasm transport.\textsuperscript{25–27} To further validate the microarray results, we performed fluorescent quantitative PCR on 6 selected lncRNAs in samples from additional 90 patients with lung adenocarcinoma. The results confirmed the upregulated and downregulated lncRNAs. In addition, we found that LINCO0578 was closely associated with the metastasis of lung adenocarcinoma, but there was no significant correlation of clinicopathologic characteristics with the other 5 lncRNAs. Further follow-up study showed that LINCO0578 expression was closely associated with DFS and OS of patients with lung adenocarcinoma. Collectively, these data indicate that LINCO0578 is a potential biomarker and therapeutic target for lung adenocarcinoma.

In summary, we explored the expression profiles of lncRNAs in lung adenocarcinoma and identified LINCO0578 as a promising biomarker and therapeutic target. Further in-depth studies are needed to elucidate the mechanism of the differentially expressed lncRNAs in lung adenocarcinoma in order to provide novel approaches for the clinical treatment of lung adenocarcinoma.

\section*{Acknowledgment}

This study was supported by Natural Science Foundation of Xinjiang Uyghur autonomous region (No 2016D01C368).

\section*{Disclosure}

The authors report no conflicts of interest in this work.

\section*{References}


