ORIGINAL ARTICLE

Identification of a RNA-Seq Based Prognostic Signature with Seven Immune-Related IncRNAs for Lung Adenocarcinoma

Jianbin You1,*, Wenting Fang2,*, Qiurong Zhao1, Liqing Chen1, Liyanguan Chen1, Falin Chen1

* Authors contributed equally

1 Department of Clinical Laboratory, Fujian Provincial Hospital, Shengli Clinical Medical College of Fujian Medical University, Fuzhou, China
2 Department of Clinical Laboratory, Fujian Maternity and Child Health Hospital, Affiliated Hospital of Fujian Medical University, Fuzhou, China

SUMMARY

Background: Lung adenocarcinoma (LUAD) is still a worldwide challenge. Accumulated evidence demonstrates that the superiority of immune-related long noncoding RNAs (lncRNAs) are closely connected with tumorigenesis and prognosis of cancer. However, no detailed studies have been conducted to present a reliable signature for predicting prognosis in LUAD patients from the perspective of tumor immunology. The aim of this study was to construct a risk score model based on the signature of the group of seven immune-related lncRNAs to predict the prognosis of patients with LUAD.

Methods: We performed a genome-wide analysis of expression profiles in 522 LUAD patients from The Cancer Genome Atlas (TCGA) project to explore the prognostic ability of immune-related lncRNAs. By using Kaplan-Meier analysis, univariate/multivariate Cox regression, receiver operating characteristic curve (ROC), and principal components analysis (PCA), a risk score model was constructed based on the signature of the group of seven immune-related lncRNAs to predict the prognosis of patients with LUAD.

Results: Using survival analysis and Cox regression model, we identified a set of seven lncRNAs (LINC00941, FAM83A-AS1, AC026355.1, AC068338.3, AC010980.2, AL365181.2, and AC079949.2) demonstrating an ability to stratify patients into high and low risk groups with significantly different survival outcomes. Moreover, the signature was identified as an independent prognostic factor and significantly associated with the overall survival (OS) of LUAD. The area under curve (AUC) of a ROC curve for the signature of the group of seven immune-related lncRNAs in predicting OS was 0.757. In addition, low-risk and high-risk groups displayed different immune statuses based on PCA.

Conclusions: This study suggested a promising seven prognostic immune-related lncRNAs risk scoring system and may provide new information for immunological treatment in LUAD.


KEY WORDS

long non-coding RNAs, lung adenocarcinoma, signature, prognosis, immune

INTRODUCTION

Lung cancer is still a worldwide challenge, which includes non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) [1]. The NSCLC group is composed of squamous cell carcinoma (SCC), adenocarcinoma (ADC), and large cell carcinoma (LCC). Lung adenocarcinoma (LUAD) is one of the main histopatholo-
gical categories of NSCLC, accounting for more than 50% of cases with NSCLC patients [2,3]. Although considerable progress has been made in the diagnosis and treatment of LUAD over the past several years, the overall survival (OS) remains poor [4,5]. Thus, it can be seen that knowledge of specific prognostic factors in LUAD patients is crucial to increase treatment efficacy and improve life expectancy.

Updated comprehensive genomic studies confirm that long non-coding RNAs (lncRNAs) are defined as a specific group of non-protein-coding transcripts measuring > 200 nucleotides in length, which are located between protein-coding genes without the ability to code proteins [6]. A growing body of research evidence has indicated that lncRNAs are significantly differentially expressed in tumors compared with normal tissues. Aberrant expression of lncRNAs has been demonstrated to have essential roles in oncogenic or tumor suppressor gene regulation in multivariate types of cancer [7,8]. Furthermore, lncRNAs serve emerging crucial roles in various stages of cancer immunity, including antigen recognition, immune activation, and tumor clearance [9-11]. For example, Inc-RNA RUNXOR is significantly associated with the immunosuppression of MDSCs in lung cancer [12]. Lnc-EGFR promotes T-regulatory cell differentiation and is a potential therapeutic target for hepatocellular carcinoma [13]. To date, many studies have identified that the aberrantly expressed lncRNAs have prognostic functions in a variety of tumor types. However, the general role of immune-related lncRNA signature in the prognostication of LUAD has not been investigated thoroughly.

Herein, we performed a genome-wide analysis of lncRNA expression profiles in 522 LUAD from The Cancer Genome Atlas (TCGA) project to explore the prognostic value of immune-related lncRNAs and identified seven significant prognostic immune-related lncRNAs as independent, remarkable prognostic factors. By using Kaplan-Meier analysis, univariate Cox regression, multivariate Cox regression, ROC and PCA analyses, we developed and further researched the risk score model based the signature of the group of seven immune-related lncRNAs to predict the prognostic value of patients with LUAD. Based on the above analysis, our study demonstrated that the seven immune-related lncRNA signatures not only were independent prognostic factors but also disclosed a better predictive performance in predicting survival of patients with LUAD.

MATERIALS AND METHODS

Patient data sets

In our study, the intact RNA sequencing (RNA-seq) database for 522 LUAD cancer patients along with their clinical follow-up information were extracted from the public TCGA data portal (https://cancergenome.nih.gov/) [14]. The clinical information includes the age, gender, stage, and TNM grade of LUAD patients. The information of all patients was documented before November 15, 2019. The data were retrieved from TCGA, which is a community resource project providing available data for community research. It was not necessary for the current study to receive approval from the local research ethics committee, as it adhered to the TCGA publication principles and data use policies. The RNA-Seq data of LUAD covered 60,483 mRNAs containing 14,142 lncRNAs, which have been labeled in NCBI (https://www.ncbi.nlm.nih.gov/) or Ensembl (http://asia.ensembl.org/).

The acquisition of prognostic immune-related lncRNA profiles

Two immune-related gene sets, including IMMUNE_RESPONSE and IMMUNE_SYSTEM_PROCESS, were retrieved from the Molecular Signatures Database (MSigDB) immune library [15]. The immune-related lncRNAs were identified by constructing immune-lncRNA co-expression networks. The relationship of abundantly expressed immune-related lncRNAs with OS time of patients was evaluated by univariate Cox regression analysis using survival R package [16]. The immune-related lncRNAs with statistically significant (p < 0.001) were considered as candidate lncRNAs. To enhance the dependability and feasibility of prognostic immune-related lncRNA signatures, we further screened candidate lncRNAs with the multivariate cox regression model. Moreover, prognostic immune-related lncRNAs were divided into detrimental factor and protective factor in the light of their coefficient.

Risk score model construction

A risk score model was built for predicting prognosis of LUAD patients by including the expression level of screened prognostic lncRNAs, weighted by their estimated regression coefficients generated from multivariate Cox regression model [17]. In the following study, the median risk score was used as the cutoff to sort all samples into low risk and high risk groups. The comparison of OS between high risk and low risk groups was assessed by Kaplan-Meier curve and compared by log-rank test.

Statistical analysis

Stratified analysis, based on univariate and multivariate Cox regression analyses, was conducted to evaluate and verify the effectiveness of the independent prognostic risk factors containing age, gender, stage, TNM grade, and risk score. The survival receiver operating characteristic (ROC) package in R software was used to plot ROC curve and calculate the area under the curve (AUC) value of each independent risk factors. In addition, we also analyzed the association between the expression of the 7 immune-related lncRNAs and the individual clinical risk factors including age, gender, stage, and TNM grade. To study the immune status of patients in two different risk groups, principal component analysis (PCA) was used to perform the degree of isolation of
RESULTS

Identification of prognostic immune-related IncRNAs from TCGA data sets

In order to explore the prognostic immune-related IncRNAs, univariate Cox proportional hazards regression analysis was performed on the expression data of all IncRNAs in the 522 LUAD of TCGA project. A total of 17 immune-related IncRNAs were identified to be prominently correlated with the LUAD patients’ OS (adjusted p-value < 0.001) and were joined into the list of candidates for further singling out (Figure 1). Multivariate Cox proportional hazards regression was adopted to evaluate the independent prognostic values of candidate prognostic immune-related IncRNAs, and we identified seven significant prognostic immune-related IncRNAs as independent remarkable prognostic factors (Table 1). Among the seven prognostic immune-related IncRNAs, two IncRNAs (AC026355.1 and AC068338.3) tended to be sheltered factors due to high expression in connection with longer survival (coefficient < 0); nevertheless, the remaining five IncRNAs (LINC00941, FAM83A-AS1, AC010980.2, AL365181.2 and AC079949.2) were inclined to be prognostic risky factors with their high expression affiliated to shorter survival (coefficient > 0).

Construction of a lncRNA-based risk score model

We constructed a risk score model by combining the expression of seven immune-related IncRNAs and the respective coefficients of these IncRNAs obtained from multiple Cox regression analysis. The risk-score formula was as follows: risk score = (0.305 x the expression level of LINC00941) + (0.143 x the expression level of FAM83A-AS1) + (-0.358 x the expression level of AC026355.1) + (-0.517 x the expression level of AC068338.3) + (0.434 x the expression level of AC010980.2) + (0.156 x the expression level of AL365181.2) + (0.364 x the expression level of AC079949.2). We computed the risk scores of 522 LUAD patients with the above formula. We classify patients into the high risk group (n = 238) or low risk group (n = 239) according to the median risk score at the cutoff point. Survival analysis suggested a significant difference in OS time between the predicted two risk groups (Figure 2, p < 0.001, log-rank test). Besides, we analyzed the distribution of the risk scores for all patients (Figure 3A). The survival status of patients was marked on the dot plot (Figure 3B) and the expression level of seven prognostic immune-related IncRNAs were revealed on the heatmap (Figure 3C).

Independent risk factor analysis for the prognosis

We carried out univariate and multivariate Cox regression analyses to assess and validate the independent prognostic risk factors, which include several clinical parameters and the seven immune-related lncRNA signatures. The univariate Cox regression analyses had manifested that four factors, including stage (HR = 3.148, p < 0.001), T grade (HR = 2.712, p < 0.001), N grade (HR = 2.823, p < 0.001) and the seven immune-related lncRNA signatures (HR = 1.555, p < 0.001), were associated with OS for the LUAD patients. In addition, results of multivariate Cox regression analysis showed that N grade (HR = 1.879, p = 0.006) and the seven immune-related lncRNA signatures (HR = 1.501, p < 0.001) were certified to be independent prognostic predictors (Table 2). In order to elaborate on the correctness of the aforementioned results, we utilized the ROC curve analysis to measure the AUC of all independent prognostic risk factors. Results showed that the signature of the group of seven immune-related IncRNAs had the highest AUC value (AUC = 0.757) compared with the remaining risk factors (Figure 4), suggesting better reliability of the combined seven immune-related lncRNA signature in predicting prognosis.

Correlation analysis between the expression of lncRNA and the clinical factors

We evaluated the expression of the seven immune-related IncRNAs in different clinical factors. The results revealed that the expression of several IncRNAs at different stages of various clinical factors also have their own special characteristics. The expression of FAM83A-AS1 in the low-age (age < 60.5) group were higher than that in the high-age (age > 60.5) group (Figure 5A). Moreover, the male group had more expression of AC079949.2 than the female group, while less expression of AC010980.2 (Figure 5B). Furthermore, patients in the advanced stage group (stage III - IV) showed increased the expression of AL365181.2, AC079949.2, and FAM83A-AS1 in contrast to those in the early stage group (stage I - II), whereas the expression of AC068338.3 was decreased (Figure 5C). In accordance with the outcomes of T grade, elevated expression of AL365181.2, AC079949.2, and FAM83A-AS1 in contrast to those in the early stage group (stage I - II), whereas the expression of AC068338.3 was up-regulated (Figure 5D). The expression of FAM83A-AS1, AC010980.2, AC079949.2, and AL365181.2 had varying degrees of down-regulation in early N grade group (N 0), whereas the expression of AC068338.3 was up-regulated (Figure 5E).

Analysis of immune status for low and high risk groups

PCA was applied to investigate the different separation patterns between low and high risk groups based on the expression of four sets, namely, the seven immune-related IncRNAs set, the 17 immune-related IncRNAs set, the immune gene set, and whole gene set. PCA of whole genome expression data set demonstrated that two groups had an unclear separation in immune status of patients (Figure 6A). The consequence of analysis
Table 1. Seven immune-related lncRNAs significantly associated with prognosis of LUAD patients by using the multivariable Cox proportional hazards regression analysis.

<table>
<thead>
<tr>
<th>lncRNA’s name</th>
<th>coef</th>
<th>HR</th>
<th>HR.95L</th>
<th>HR.95H</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINC00941</td>
<td>0.30504374</td>
<td>1.356698771</td>
<td>1.109507086</td>
<td>1.658963316</td>
<td>0.00295331</td>
</tr>
<tr>
<td>FAM83A-AS1</td>
<td>0.14266317</td>
<td>1.153341257</td>
<td>0.997498072</td>
<td>1.333532457</td>
<td>0.054086608</td>
</tr>
<tr>
<td>AC026355.1</td>
<td>-0.358461857</td>
<td>0.698750277</td>
<td>0.536147593</td>
<td>0.910667056</td>
<td>0.007992701</td>
</tr>
<tr>
<td>AC068338.3</td>
<td>-0.516973895</td>
<td>0.596322354</td>
<td>0.356124907</td>
<td>0.998527043</td>
<td>0.049348488</td>
</tr>
<tr>
<td>AC010980.2</td>
<td>0.433841698</td>
<td>1.543174561</td>
<td>1.162983929</td>
<td>2.04763167</td>
<td>0.002645273</td>
</tr>
<tr>
<td>AL365181.2</td>
<td>0.156161741</td>
<td>1.169015266</td>
<td>1.008595826</td>
<td>1.354949779</td>
<td>0.038114878</td>
</tr>
<tr>
<td>AC079949.2</td>
<td>0.364189117</td>
<td>1.439346394</td>
<td>1.133394499</td>
<td>1.827887857</td>
<td>0.002817753</td>
</tr>
</tbody>
</table>

Abbreviations: lncRNA - long noncoding RNA, LUAD - lung adenocarcinoma, coef - coefficient, HR - hazard ratio.

Table 2. Factors with prognostic significance in univariate Cox regression analysis were selected for further multivariate analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate</th>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>p-value</td>
</tr>
<tr>
<td>Age, year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 60.5 vs. &lt; 60.5</td>
<td>1.023</td>
<td>0.910</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male vs. Female</td>
<td>0.986</td>
<td>0.942</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage III - IV vs. Stage I - II</td>
<td>3.148</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>T grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3 - T4 vs. T1 - T2</td>
<td>2.712</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>M grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1 vs. M0</td>
<td>1.736</td>
<td>0.072</td>
</tr>
<tr>
<td>N grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1 - N3 vs. N0</td>
<td>2.823</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Risk score</td>
<td>1.555</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Abbreviations: HR - hazard ratio.

displayed that seven immune-related lncRNAs tend to subdivide LUAD patients into two sides, indicating that there were obvious differences in immune status of patients between high risk and low risk groups (Figure 6B).

DISCUSSION

LUAD is one of the most common causes of cancer death. Late diagnosis and poor prognosis make it difficult to treat the advanced metastatic disease [18]. Therefore, it is essential to explore potential prognostic indicators beyond traditional clinical risk factors for LUAD patients. In our study, we acquired the immune-related lncRNAs associated with prognosis by sorting the data from the TCGA project. Then, we found that the risk score model could be utilized as a good prognostic marker for LUAD patients by constructing a prognostic signature using the expressions of seven immune-related lncRNAs.

A considerable amount of literature has been published that lncRNAs aberrantly expressed in tumor tissues have close correlation with the occurrence, progression,
and recurrence of manifold tumors [19-21]. Recently, researchers have found that immune-related lncRNAs observably regulate the immunoreaction of cancers and take part in immune response, gene activation, and tumor clearance, etc. [11,22,23]. In addition, to coordinate the immune system of patients, lnc-MALAT1 may enhance the effectiveness of chemotherapy for osteosarcoma, according to a new study [24]. Furthermore, lnc-AGER-1 inhibits the development of lung cancer by up-regulating the innate immune pattern-recognition recep-

Figure 1. A total of 17 immune-related lncRNAs identified from univariate Cox regression analysis.
Figure 2. Kaplan Meier survival curves of OS among LUAD patients from different groups stratified by the signature in the LUAD patients.

OS - overall survival, LUAD - lung adenocarcinoma.
Figure 3. Immune-related lncRNA risk score analysis of LUAD patients in TCGA.

(A) The low and high score group for the immune-related lncRNA signatures in LUAD patients. (B) The survival status and duration of LUAD cases. (C) Heatmap of the expression of the group of seven key lncRNAs in LUAD. The color from blue to red shows an increasing trend from low levels to high levels.

Figure 4. ROC curves indicate the prognostic value of the seven independent prognostic factors.

ROC - receiver operating characteristic.

Figure 5. Expression of seven immune-related lncRNAs in different age, gender, stages, and TNM grades.
mRNA stability of FAM83A and facilitate hepatocellular carcinoma progression [27]. FAM83A-AS1 also promotes the migration and invasion of lung cancer cells by targeting mir-150-5p and modifying MMP14 [28, 29]. Our study also manifested that the expression of FAM83A-AS1 increased significantly in the low-age (age < 60.5) group, advanced stage group (stage III-IV), and advanced N grade group (N1-N3). Recent reports have suggested that LINC00941 affected the development, prognosis, and therapeutic effect of various tumors [30-32]. In addition, our research demonstrated that the expressions of the seven immune-related IncRNAs in several clinical factors have their own significant differences. To sum up, these seven immune-related IncRNAs may also serve as prognostic factors in patients with LUAD individually. Thus, the seven immune-related IncRNAs may be promising prognostic marker candidates independently, and the risk score model constructed by the seven immune-related IncRNAs may also provide new information for immunological treatment in LUAD patients.

In recent years, immunotherapy has been regarded as a promising treatment, and we combined a latent new immune-related approach with prognostic target to predict the prognosis of LUAD. A limitation of our research is that it was retrospective and should be further confirmed by prospective studies, such as IHC, PCR, or flow cytometry. Meanwhile, in order to estimate the potential role of signatures in early diagnosis, the expression differences of the seven immune-related IncRNAs between normal and tumor tissues should be evaluated. Moreover, to better apply our research results to clinical diagnosis and treatment, more functional studies on the seven immune-related IncRNAs alone and in combination should be implemented. Also, further vitro experiments and animal studies should be performed to elucidate immune-related IncRNA potential mechanisms. Despite these limitations, our study draws the novel conclusion that the signature of the group of immune-related IncRNAs was significantly associated with the risk value, tumor status, and OS of LUAD patients.

In conclusion, we established a risk score model based on the seven immune-related IncRNAs (namely, LINC00941, FAM83A-AS1, AC026355.1, AC068338.3, AC010980.2, AL365181.2 and AC079949.2), which may have independent significance in predicting the prognosis of LUAD. In addition, our study provides a novel and feasible method to measure the prognosis of LUAD and presents enlightening insight into anti-tumor immunotherapy strategies.

Figure 6. The low risk and high risk groups displayed different immune statuses.

A. Principal components analysis between low and high risk groups on the basis of the immune-related IncRNAs set. B. Principal components analysis between low and high risk groups on the basis of the whole gene set.
Acknowledgment:
This work was supported by Startup Fund for scientific research, Fujian Medical University (Grant number: 2017XQ1130); Training Project of Young Talents in Health System of Fujian Province (Grant number: 2018-ZQN-08).

Declaration of Interest:
The authors declare that they have no conflict of interest.

References:


