Identification Invasion-Related Long Non-Coding RNAs in Lung Adenocarcinoma and Analysis of Competitive Endogenous RNA Regulatory Networks

Yuze Mao1,*, Fangyu Cai2,*, Tengjiao Jiang1, Xiaofeng Zhu1

1Department of Cardio-Thoracic Surgery, First Affiliated Hospital of Jiamusi University, Jiamusi, Heilongjiang, 154000, People’s Republic of China; 2Department of Thoracic Surgery, Beidahuang Industry Group General Hospital, Harbin, Heilongjiang, 150088, People’s Republic of China

*These authors contributed equally to this work

Correspondence: Xiaofeng Zhu, Department of Cardio-Thoracic Surgery, First Affiliated Hospital of Jiamusi University, Jiamusi, Heilongjiang, 154000, People’s Republic of China, Tel +86-13845456700, Email jiamusizxf@163.com

Background: Cell invasion plays a vital role in cancer development and progression. Aberrant expression of long non-coding RNAs (lncRNAs) is also critical in carcinogenesis. However, the prognostic value of invasion-related lncRNAs in lung adenocarcinoma (LUAD) remains unknown.

Methods: Differentially expressed mRNAs (DEmRNAs), lncRNAs (DElncRNAs), and microRNAs (DEmiRNAs) were between LUAD and control samples. Pearson correlation analyses were performed to screen for invasion-related DElncRNAs (DEIRLs). Univariate and multivariate Cox regression algorithms were applied to identify key genes and construct the risk score model, which was evaluated using receiver operating characteristic (ROC) curves. Gene set enrichment analysis (GSEA) was used to explore the underlying pathways of the risk model. Moreover, an invasion-related competitive endogenous RNA (ceRNA) regulatory network was constructed. Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) was performed to detect the expression of prognostic lncRNAs in the LUAD and control samples.

Results: A total of 45 DElncRNAs were identified as DEIRLs. RP3-525N10.2, LINC00857, EP300-AS1, PDZRN3-AS1, and RP5-1102E8.3 were potential prognostic lncRNAs, the expression of which was verified by RT-qPCR in LUAD samples. Both the risk score model and nomogram used the prognostic lncRNAs. ROC curves showed the risk score model had moderate accuracy and the nomogram had high accuracy in predicting patient prognosis. GSEA results indicated that the risk score model was associated with many biological processes and pathways relevant to cell proliferation. A ceRNA regulatory network was constructed in which PDZRN3–miR-96-5p–CPEB1, EP300–AS1-miR-93-5p–CORO2B, and RP3–525N10.2-miR-130a-5p–GHR may be key invasion-related regulatory pathways in LUAD.

Conclusion: Our study identified five novel invasion-related prognostic lncRNAs (RP3-525N10.2, LINC00857, EP300-AS1, PDZRN3-AS1, and RP5-1102E8.3) and established an accurate model for predicting the prognosis of patients with LUAD. These findings enrich our understanding of the relationships between cell invasion, lncRNAs, and LUAD and may provide novel treatment directions.

Keywords: lung adenocarcinoma, invasion, LncRNA, prognosis, CeRNA

Introduction

According to the World Health Organization, cancer is a leading cause of death worldwide.1 According to the International Agency for Research on Cancer Global Burden of Cancer Report 2020,2 an estimated 2.2 million new cancer cases and 1.8 million deaths, lung cancer is the second most commonly diagnosed and the leading cause of cancer death, representing approximately one in ten cancers diagnosed and one in five deaths. Non-small cell lung cancer
accounts for approximately 85% of all lung cancers and is the leading subtype in morbidity and mortality. The annual incidence of lung adenocarcinoma (LUAD) is increasing. The major LUAD subclassifications are invasive, invasive non-mucinous, invasive mucinous, colloid, fetal, and enteric-type adenocarcinoma. Although surgical resection, chemotherapy, radiotherapy, immunotherapy, and targeted therapies have been used for treating LUAD for decades, patients still have a poor prognosis due to local recurrence or distant metastases. Therefore, there is an urgent need to identify new predictors of LUAD prognosis.

Metastasis is the process by which cancer cells from a primary lesion invade a distal organ or location and involves cell detachment and adhesion, cell invasion, and response to chemokine stimuli. Among these, invasion of cancer cells is one of the most important causes of metastasis. In recent years, markers of invasion-prone tumors have been identified. Liu et al constructed a prognostic stratification system based on invasion-related genes (IRGs) in colon cancer that could accurately predict patient outcomes. Zheng et al developed prognostic risk models based on differentially expressed IRGs in different molecular subtypes of LUAD. Therefore, studying invasion-related markers to ensure earlier treatment can potentially improve the prognosis and survival of cancer patients.

Long non-coding RNA (lncRNAs) are a class of 200 bp–100 kb non-coding RNA with unique sequences. Many non-coding RNAs are aberrantly expressed in tumors with high tissue specificity and may serve as prognostic biomarkers. For example, the lncRNA Neat1, a key molecule in cancer research, is elevated in breast, prostate, stomach, colorectal, liver, and ovarian alveolar cancer. IncRNA prognostic markers have also been reported in LUAD, such as immune- and autophagy-related lncRNAs. However, the prognostic value of invasion-associated lncRNAs in LUAD has not yet been investigated.

This study aimed to investigate invasion-related lncRNAs (IRLs) in LUAD. We screened IRLs, constructed a prognostic risk model, and investigated regulatory mechanisms in LUAD. We also verified their expression in patient samples. We identified that lncRNA–miRNA–mRNA relationships might play critical roles in the prognosis of LUAD. This study revealed IRLs and their role in LUAD. Our findings may help predict the prognosis of LUAD and distinguish the chemotherapeutic sensitivity of patients in low- and high-risk groups.

Materials and Methods

Data Acquisition

The lncRNA and mRNA expression data of 521 LUAD and 58 adjacent control samples and the microRNA (miRNA) expression data of 515 LUAD and 45 adjacent control samples were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/). The GSE31210 dataset containing 226 LUAD and 20 adjacent control samples and GSE50081 dataset containing overall survival (OS) information of 127 patients with LUAD, were downloaded from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) database. We used the clinical data of 497 patients with LUAD from TCGA-LUAD cohort to construct a prognostic model and nomogram. Moreover, 689 IRGs were obtained from the Molecular Signatures Database (MSigDB) (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp) and the cancer single-cell state atlas (CancerSEA) (http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp) database after de-duplicating. We randomly collected surgical specimens from ten groups of patients with LUAD who had signed informed consent to detect the expressions of prognostic lncRNAs in LUAD and control samples. Research was performed in accordance with the Declaration of Helsinki (as revised in 2013) and approval was obtained from the Ethics Committee of the First Affiliated Hospital of Jiamusi University (NO.20222001).

Differential Expression Analysis

The “DESeq2” R package was used to screen differentially expressed mRNAs (DEmRNAs), differentially expressed lncRNAs (DElncRNAs), and differentially expressed miRNAs (DEmiRNAs) between LUAD and control samples with |log2Fold change (FC)| >1 and adjusted p-value < 0.05.
Identification of Invasion-Related DElncRNAs

DEmRNAs and DElncRNAs were identified by overlapping TCGA-LUAD and GSE31210 datasets. The DEmRNAs were overlapped with IRGs to obtain DEIRGs. Next, Pearson correlation between DElncRNAs and DEIRGs were calculated to screen invasion-related DElncRNAs (DEIRLs) using |r >0.5| and p-value < 0.05. Overlapping DEIRLs between TCGA and GSE31210 were used for subsequent analyses.

Construction, Evaluation, and Validation of Risk Score Model

At a ratio of 7:3, 497 patients with LUAD in TCGA database were randomly segregated into either training or internal validation sets. Subsequently, univariate and multivariate Cox regression analyses were performed to filter OS-related DEIRLs for risk score model construction in the training set (p < 0.05). The risk score was calculated using the following formula:

\[ h_0(t) \times \exp(\beta_1X_1 + \beta_2X_2 + \ldots + \beta_nX_n) \]

where \( h_0(t) \) is the baseline hazard function, \( \beta \) is the regression coefficient, and \( X \) is the normalized expression value of each prognostic IRL. Risk model establishment and validation were conducted according to Wang et al.\(^{18} \) In brief, patients with LUAD in the TCGA training set were separated into low- and high-risk groups based on the optimal cut-off calculated by the surv_cutpoint function in the “survminer” R package. Kaplan–Meier (KM) analysis was used to analyze the OS of the low- and high-risk groups. Then, to evaluate the performance of the risk model, ROC curves were plotted using the R package “survivalROC.” The risk model was subsequently evaluated using the internal and external validation sets. Subsequently, the independent prognostic factors for patients with LUAD were identified by univariate and multivariate Cox regression analyses for plotting a nomogram, which was used to predict the survival of patients with LUAD at 1, 3, and 5 years. Finally, calibration and decision curves were constructed to assess nomogram performance.

Construction of ceRNA Regulatory Network

The miRanda database (http://cbio.mskcc.org/miRNA2003/miRanda.html) was used to screen for DEmiRNAs that interact with prognostic lncRNAs. The miRwalk (walk.umm.uni-heidelberg.de) and TargetScan (http://www.targetscan.org/) databases were used to predict the DEmRNAs targeted by the screened DEmiRNAs. A ceRNA regulatory network was constructed and visualized using the Cytoscape software.

RT-qPCR Analysis

We used reverse transcription-quantitative polymerase chain reaction (RT-qPCR) to detect the expression of the five potential prognostic lncRNAs in LUAD and normal lung tissues, according to Gan et al.\(^{19} \) Total RNA from LUAD samples (n = 10) and adjacent tumor control samples (n = 10) was extracted using Nuclezol LS RNA Isolation Reagent (ABP Biosciences Inc., China). After determining the concentration and purity of RNA, qualified RNA was used for reverse transcription using a SureScript-First-strand-cDNA-synthesis-kit (GeneCopoeia, USA). RT-qPCR was performed on a CFX96 Real-time PCR System (Bio-Rad, USA) using BlazeTag SYBR Green qPCR Mix 2.0 (GeneCopoeia, USA) under the following thermal cycling conditions: 40 cycles at 95 °C for 30s, 95 °C for 10s, 60 °C for 20s, and 72 °C for 30s. The 2^−ΔΔCt method was used to calculate gene fold expression. The primer sequences used in this study are listed in Table 1.

Statistical Analysis

Principal component analysis (PCA) was performed for effective dimension reduction and pattern recognition of high-dimensional data in the risk model. Gene set enrichment analysis (GSEA) was performed for functional annotation of the low- and high-risk groups. The half-maximal inhibitory concentration of chemotherapy drugs was evaluated by the “pRProphetic” R package to predict the chemotherapeutic sensitivity of patients in low- and high-risk groups. Comparisons between two groups and among multiple groups were performed using the Wilcoxon and Kruskal–Wallis tests, respectively. All data were analyzed using R (version 4.0.0), and statistical significance was set at p < 0.05.
Results

Identification of 45 DEIRLs in LUAD

A total of 5353 DEmRNAs were identified between LUAD and control samples in the TCGA-LUAD cohort, including 3346 upregulated and 2007 downregulated genes (Figure 1A). The top 100 DEmRNAs are displayed in a heatmap (Figure 1B). A total of 3560 DElncRNAs were identified between LUAD and control samples in the TCGA-LUAD cohort, including 2489 upregulated and 1071 downregulated genes (Figure 1C). The top 100 DElncRNAs are depicted as a heatmap (Figure 1D). In the GSE31210 dataset, 1603 DEmRNAs and 104 DElncRNAs were identified between LUAD and control samples (Figure 1E), and the top 100 differentially expressed genes (DEmRNAs and DElncRNAs) are presented in a heatmap (Figure 1F). Then, 1362 common DEmRNAs were obtained by overlapping upregulated and downregulated DEmRNAs from TCGA-LUAD and GSE31210 datasets (Figure 1G). Next, 85 DEIRGs were screened by intersecting the common DEmRNAs with 689 IRGs (Figure 1H). Meanwhile, 58 common DElncRNAs were selected by intersecting upregulated and downregulated DElncRNAs from the TCGA-LUAD cohort and GSE31210 dataset (Figure 1I).

Next, we calculated the correlations between DEIRGs and common DElncRNAs in TCGA and GSE31210 datasets using Pearson correlation analysis. The correlation network of DEIRGs and common DElncRNAs from the TCGA cohort comprised 125 nodes and 987 edges (Figure 2A), and a correlation network from the GSE31210 dataset was composed of 136 nodes and 1827 edges (Figure 2B). Lastly, 45 common DEIRLs were obtained using a Venn algorithm for downstream analysis (Figure 2C).

Construction and Validation of the Risk Score Signature Based on Prognostic DEIRLs in LUAD

We explored the prognostic value of the 45 DEIRLs by performing univariate Cox regression in the training set. We observed that RP3-525N10.2, LINC00857, EP300-AS1, SBF2-AS1, PDZRN3-AS1, DEPDC1-AS1, and RP5-1102E8.3 were related to the OS of patients with LUAD (Figure 3A). Further multivariate Cox regression analysis revealed that RP3-525N10.2, LINC00857, EP300-AS1, PDZRN3-AS1, and RP5-1102E8.3 were potential prognostic lncRNAs in patients (Figure 3B). Moreover, RT-qPCR results revealed that LINC00857 expression was significantly elevated, and the expressions of RP3-525N10.2, EP300-AS1, PDZRN3-AS1, and RP5-1102E8.3 were significantly reduced in LUAD samples compared to those in controls (Figure 3C), which was consistent with the sequencing results. Subsequently, the risk score of each patient was calculated in the training, testing, GSE31210, and GSE50081 datasets. Next, the relationship between risk score and clinical features was explored in the TCGA-LUAD cohort. We observed significant differences in risk scores between groups stratified by gender, N staging, tumor staging, T staging, and smoking history (Figure 3D).

Table 1 RT-qPCR Primer Sequences Used in the Current Study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINC00857 F</td>
<td>GAAAAAGACACCAAAACTCGG</td>
</tr>
<tr>
<td>LINC00857 R</td>
<td>CTCTACACTCAACCCACGC</td>
</tr>
<tr>
<td>EP300-AS1 F</td>
<td>GTGACGGCCTGTGAGAGGAGG</td>
</tr>
<tr>
<td>EP300-AS1 R</td>
<td>TGGGAGGCTGAGATGGATG</td>
</tr>
<tr>
<td>PDZRN3-AS1 F</td>
<td>ATGTTCTACACCCCTCAGG</td>
</tr>
<tr>
<td>PDZRN3-AS1 R</td>
<td>TGGCTCTCCCGTCTCCTG</td>
</tr>
<tr>
<td>RPS-1102E8.3 F</td>
<td>TGGTAAATGACATAGGTGTTGAATCCTT</td>
</tr>
<tr>
<td>RPS-1102E8.3 R</td>
<td>GACCATTTCAAAGTAATCTGATTGATA</td>
</tr>
<tr>
<td>RP3-525N10.2 F</td>
<td>TAAAGCATGCTGGGATAAACA</td>
</tr>
<tr>
<td>RP3-525N10.2 R</td>
<td>CAAAAGGACACAAAACAAATG</td>
</tr>
<tr>
<td>GAPDH F</td>
<td>CCCATCACCATCCTCCAGG</td>
</tr>
<tr>
<td>GAPDH R</td>
<td>CATCAGCCACAGTTTCCC</td>
</tr>
</tbody>
</table>

https://doi.org/10.2147/IJGM.S407266

DovePress
International Journal of General Medicine 2023:16

Mao et al

Dove Press
Powered by TCPDF (www.tcpdf.org)

Powered by TCPDF (www.tcpdf.org)

Powered by TCPDF (www.tcpdf.org)
Using the optimal cut-off value, we divided the patients into low- and high-risk groups (Supplementary Figure 1). The risk score and survival status distributions of patients with LUAD in the TCGA training set are presented in Figure 3E. KM analysis indicated that the high-risk group had worse clinical outcomes (Figure 3F). The expression levels of RP3-525N10.2, LINC00857, EP300-AS1, PDZRN3-AS1, and RP5-1102E8.3 are presented in Figure 3G. Moreover, the ROC curve indicated that the risk score model had moderate accuracy in predicting the 1-, 3-, and 5-year survival in the TCGA training set with an area under the curve (AUC) value > 0.6 (Figure 3H). Next, we validated the risk model in the TCGA testing set and the external GSE31210 and GSE50081 datasets. Consistent results were observed in the internal (Supplementary Figure 2A–D) and external (Supplementary Figure 2E–H and Supplementary Figure 2I–L) datasets, suggesting risk model reliability.
Figure 2 Identification of DEIRLs in LUAD. (A) The correlation network of DEIRGs and DElncRNAs in TCGA; (B) The correlation network of DEIRGs and DElncRNAs in GSE31210; (C) Venn plot of common DEIRLs in TCGA-LUAD and GSE31210.

Abbreviations: lncRNA, long non-coding RNA; IRL, invasion-related lncRNA; DE, differentially expressed; LUAD, lung adenocarcinoma.
Figure 3 Identified prognostic DEIRLs and constructed a risk score model in LUAD from the training set. (A) Univariate COX analysis of DEIRLs in forest plot; (B) Identification of risk signature by multivariate COX analysis in forest plot; (C) The expression levels of IRLs signature in tumor and control samples by RT-qPCR; (D) Relationship between risk score and clinical features; (E) The risk curve and survival status of patients in TCGA-LUAD; (F) Kaplan-Meier curves of survival in low- and high-risk; (G) Heatmap of the expressions of LINC00857, RP5-1102E8.3, RP3525N10.2 in low- and high-risk score; (H) ROC curves of the risk score model for predicting the survival of 1-, 3- and 5-year.

Abbreviations: lncRNA, long non-coding RNA; IRL, invasion-related lncRNA; DE, differentially expressed; LUAD, lung adenocarcinoma; ROC, receiver operating characteristic. *P<0.05, **P<0.001, ***P<0.0001.
Development of a Nomogram for Predicting the Survival of LUAD Patients

Next, we explored whether the risk score was an independent prognostic factor using univariate and multivariate Cox regression analyses. Univariate Cox regression analysis revealed that the risk score was significantly associated with prognosis in both TCGA and GSE50081 datasets, and multivariate Cox regression analysis further confirmed that result (Table 2). Thus, we included risk score, N staging, and T staging to construct a nomogram to predict the 1-, 3-, and 5-year survival of patients with LUAD (Figure 4A). We noted good concordance between the observed and predicted 1-, 3-, and 5-year OS (Figure 4B). Moreover, the decision curves suggest that the nomogram has a better predictive value compared to a single risk score, T staging, and N staging (Figure 4C–E).

Characterization of Low- and High-Risk Group

PCA was performed to compare the difference between low- and high-risk groups based on the risk model of the five prognostic IRLs (Figure 5A), 45 common DEIRLs (Figure 5B), and whole genome expression profiles (Figure 5C) from TCGA. The results suggested that the low-risk and high-risk groups were distinctly distributed (Figure 5A) prior to the other two patterns (Figures 5B and C). These results suggest that the risk model divided patients with LUAD into two groups and that the invasion propensity in the high-risk group differs from the low-risk group.

We performed GSEA to identify the biological processes and pathways related to this risk model. The top ten enriched biological processes were mainly associated with cell proliferation, including mitochondrial translation, mitochondrial gene expression, chromosome segregation, meiotic cell cycle process, DNA-dependent DNA replication, meiotic cell cycle, DNA replication, cell cycle G2/M phase transition, microtubule-based movement, and

<table>
<thead>
<tr>
<th>Table 2 Independent Prognostic Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>TCGA cohort</td>
</tr>
<tr>
<td>Risk score</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Stage</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>GSE50081 dataset</td>
</tr>
<tr>
<td>Risk score</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Stage</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
</tbody>
</table>

Note: Statistical significance is shown in bold.
cilium movement (Figure 5D and E). Similarly, cell cycle-related pathways, such as cell cycle and DNA replication, were also enriched. In addition, the risk score model was also relevant to proteasome, ribosome, spliceosome, steroid hormone biosynthesis, Parkinson’s disease, oxidative phosphorylation, asthma, and systemic lupus erythematosus (Figure 5F and G).

Additionally, we compared the sensitivity of patients in the low- and high-risk groups to drugs used in the treatment of LUAD, including docetaxel, gefitinib, paclitaxel, rapamycin, gemcitabine, and cisplatin. We observed that patients in the high-risk group were more sensitive to docetaxel, paclitaxel, and gemcitabine than those in the low-risk group (Figure 6).

Construction of Invasion-Related ceRNA Network in LUAD
A total of 418 DEmiRNAs were identified between LUAD and control samples in TCGA, including 244 upregulated and 174 downregulated genes in the tumor samples (Figure 7A). The expression of the top 100 DEmiRNAs is depicted in a heatmap (Figure 7B). Using miRanda, we predicted 39 prognostic IncRNA–DEmiRNA pairs. Next, 506 DEmiRNA–DEmRNA pairs were predicted using miRWalk and TargetScan databases. Accordingly, we constructed a ceRNA network comprised 333 nodes and 545 edges (Figure 7C). In the ceRNA network, we observed that the DEmRNAs CPEB1, CORO2B, and GHR were IRGs. Thus, we extracted PDZRN3–miR-96-5p–CPEB1, EP300–AS1–miR-93-5p–CORO2B, and RP3–525N10.2–miR-130a-5p–GHR from the network (Figure 7D), which may play a role in critical invasion-related ceRNA regulatory relationships in LUAD. Figure 8A shows that the expression of CPEB1, CORO2B, and GHR was significantly lower in LUAD samples than in normal samples, which was contrary to the expression trend of their corresponding IncRNAs. The ROC curves of three IncRNAs and three DEmiRNAs were plotted to explore the
relevance of their expression to the survival of patients with LUAD. The results demonstrated that the survival rate of patients in the high- and low-expression groups of GHR, PDZRN3-AS1, and RP3–525N10.2 was notably different, and the prognosis was better in the high-expression group (Figure 8B).

Discussion
In recent years, lung cancer incidence and mortality rates have been among the highest in the world and continue to increase. Non-small cell lung cancer is the most predominant subtype, and delayed diagnosis leads to patients with advanced disease, especially LUAD. To discover new biomarkers for diagnosis, the role of lncRNAs in lung cancer has
been extensively investigated.\textsuperscript{21–23} In recent years, lung biopsies, electronic bronchoscopy, and mediastinum testing have been used. However, the screening process takes several weeks, and the condition of some patients deteriorates while they wait for hospital appointments.\textsuperscript{24} If specific and accurate biomarkers for diagnosis and prognosis are identified and applied, the survival and prognosis of patients with cancer will be improved through the use of corresponding therapy. Thus, we investigated the prognostic role of invasive markers in patients with LUAD.

In this study, we generated five potential prognostic lncRNAs for LUAD using univariate and multivariate Cox regression analyses. Based on these five lncRNAs, we constructed a risk model and nomogram that could accurately predict the prognosis of patients with LUAD. Lan et al\textsuperscript{25} reported that LINC00857 is a novel lncRNA signature for predicting the prognosis of patients with early-stage LUAD, and LINC00857 overexpression promotes cancer cell proliferation and invasion.\textsuperscript{26} Another study reported that LINC00857 promotes LUAD progression by targeting the miR-1179/SPAG5 axis, thus regulating cell proliferation, glycolysis, and apoptosis.\textsuperscript{27} In addition, LINC00857 is associated with invasion in colorectal and gastric cancer.\textsuperscript{28,29} No studies have reported the prognostic roles of RP3-525N10.2, EP300-AS1, PDZRN3-AS1, and RP5-1102E8.3 in cancer. Thus, further research is required to explore the mechanisms by which these lncRNAs promote invasion in LUAD and future research should focus broadly on the molecular pathways by which these lncRNAs act.

To identify the mechanism of these five lncRNAs in LUAD prognosis, we performed functional analysis using GSEA. This analysis indicated that these lncRNAs were significantly enriched in cell proliferation-related biological processes and cell cycle-related pathways, both of which play important roles in LUAD development. Several studies have also suggested that cell cycle and proliferation are regulated by lncRNAs. LINC00857 knockdown causes cell cycle arrest at G1/S via cyclin E1 (CCNE1) regulation, which decreases lung tumor cell growth.\textsuperscript{26} Yin et al observed that YT521-B homology domain family 2 (YTHDF2) deficiency inhibited lung cancer cell proliferation and migration.\textsuperscript{30} Yang et al reported that PTPRG-AS1 accelerated the cell cycle in LUAD cells.\textsuperscript{31} These findings indicate that the prognostic lncRNAs identified in the current study may be essential in LUAD pathogenesis and development by increasing the cell cycle rate and proliferation.
risk groups to drugs used in the treatment of LUAD and found that patients in the high-risk group were more sensitive to docetaxel, paclitaxel and gemcitabine than those in the low-risk group. Investigating these genes may provide new clues for LUAD prognosis.

Finally, we constructed an invasion-related ceRNA regulatory network and extracted three key sets of ceRNA regulatory relationships (CPEB1–hsa-miR-96-5p–PDZRN3-AS1, CORO2B–hsa-miR-93-5p–EP300-AS1, and GHR–hsa-miR-130a-5p–RP3-525N10.2). In miRNA-mRNA regulatory relationships, CPEB1, CORO2B, and GHR are differentially expressed IRGs. Nagaoka et al reported that reduced CPEB1 levels promote breast cancer cell metastasis to the
lungs, whereas ectopic expression of CPEB1 inhibits this process. CORO2B is targeted by hsa-miR-93-5p and is involved in cigarette smoke-induced malignant transformation in BEAS-2B cells. In addition, GHR stimulation can promote cell invasion and metastasis, whereas its deficiency decreases fatal neoplasm incidence. In this study, we discovered that lncRNAs associated with LUAD metastatic propensity in the ceRNA network are indirectly associated with mRNA signaling. The data suggests that these lncRNA–miRNA–mRNA relationships may play critical roles in the prognosis of LUAD.

We analyzed the differences in the expression of CPEB1, CORO2B, and GHR in normal and tumor samples through bioinformatics methods and assessed the differences in survival between the high- and low-expression groups. Liu et al observed that MiR-96-5p overexpression promotes the proliferation, migration, and invasion of LUAD cells; Wang et al observed that upregulated CPEB1 reversed the promoting effect of BMSC-Exos on lung cancer cell invasion, migration, and lung metastasis; Zhao et al observed that the expression of EP300-AS1 was downregulated in LUAD, which may play a tumor-suppressive role in LUAD. Pan et al reported that ALK rearrangements occurred in approximately 5% of non-small cell lung cancer, and GHR was identified as a novel ALK fusion partner in patients with non-small cell lung cancer. These findings further validate that PDZRN3-miR-96-5p-CPEB1, EP300-AS1-miR-93-5p-CORO2B, and RP3-525 N10.2 miR-130a-5p-GHR may be involved in a key regulatory pathway associated with LUAD invasion. In the future, studies of these pathways will further play positive roles.

Our study had some limitations. First, the reliability of this model needs to be iteratively improved in long-term clinical applications with larger sample sizes. Moreover, our gene signature was suited to the patients with surgical or biopsy specimens because this gene signature was based on gene expression data. Lastly, PDZRN3-miR-96-5p-CPEB1, EP300-AS1-miR-93-5p-CORO2B, and RP3-525 N10.2 miR-130a-5p-GHR may be involved in a regulatory pathway related to LUAD invasion, but we did not functionally validate this. This result will be verified in later studies using basic experiments. Nevertheless, despite these limitations, our study established an invasion-related prognostic lncRNAs signature that could be applied as an independent prognostic factor for LUAD and to predict response to prognosis. We will conduct in-depth research to validate our prognostic signature and strengthen its effectiveness and clinical practicability.

**Conclusion**

We identified five invasion-related prognostic lncRNAs (RP3-525N10.2, LINC00857, EP300-AS1, PDZRN3-AS1, and RP5-1102E8.3), established a risk model and nomogram that can accurately predict the survival of patients with LUAD, and preliminarily explored the molecular mechanism of prognostic lncRNAs in regulating LUAD. Thus, our study provides new insights into the prognostic indicators of LUAD.
Data Sharing Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Approval and Consent to Participate
The study was approved by the Ethics Committee of the First Affiliated Hospital of Jiamusi University. All participants provided signed informed consent in accordance with the Declaration of Helsinki.

Acknowledgments
We appreciate the support of all participators involved in this study.

Disclosure
The authors declare that they have no competing interests in this work.

References


