A novel Chemokine-related LncRNA signature predicts the prognosis and immunotherapy response in lung adenocarcinoma patients

Kang Zhu  
Second Affiliated Hospital of Nanchang University

Linmin Xiong  
Second Affiliated Hospital of Nanchang University

Jianwen Xiong  
Second Affiliated Hospital of Nanchang University

Ying Zhu  
Nanchang University

Yonggang Liang  
Second Affiliated Hospital of Nanchang University

Dongliang Yu (✉️ 401441620027@email.ncu.edu.cn)  
Second Affiliated Hospital of Nanchang University

Research Article

Keywords: Chemokine, LncRNA, Prognostic, LUAD, immune

Posted Date: June 30th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3115047/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Chemokines and their receptors are widely reported to be closely associated with cancer progression, especially in the immune microenvironment. However, studies on chemokine-related IncRNAs (CRLs) in lung adenocarcinoma (LUAD) have not been reported. This study aimed to construct a prognostic model based on CRL signature to explore their relationship with prognosis and immune infiltration in LUAD.

Samples and methods:

We first obtained RNA-seq data and clinical information from The Cancer Genome Atlas (TCGA) database, then identified prognostic CRLs by co-expression analysis and univariate Cox analysis, and constructed a prognostic model based on CRLs to predict the prognosis of LUAD patients using multifactorial Cox analysis and the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm. Kaplan-Meier (K-M) survival curve analysis and receiver operating characteristic (ROC) curve analysis were used to assess the prognostic ability of the model. Finally, we also explored the relationship of the risk model with immune checkpoint gene expression, tumor mutation burden, immunotherapy scores, and drug sensitivity.

Results

We constructed a risk model based on seven CRLs (AL391261.1, AC034223.2, SH3BP5-AS1, LY86-AS1, AC104971.3, LINC01843, AL157388.1) that were significantly associated with prognosis. Patients with LUAD were divided into high-risk and low-risk groups, using the median value of the risk score as the cutoff. K-M survival analysis showed that the higher the risk score, the worse the prognosis. the area under the ROC curve (AUC) was 0.796, and multi-factor Cox analysis showed that the risk score was an independent risk factor affecting the prognosis of LUAD. In addition, our risk model played a key role in predicting immune checkpoint gene expression, tumor mutation burden, immunotherapy score, and drug sensitivity in LUAD patients.

Conclusion

We have identified a new CRL signature that has clinical value in predicting the prognosis of LUAD patients and provides a theoretical basis for the development of immunotherapy regimens for LUAD.

Introduction

Lung cancer is one of the malignant tumors with high morbidity and mortality rates worldwide, especially among men[1]. According to clinicopathological histotype, lung cancer can be divided into non-small cell
lung cancer and small cell lung cancer, of which non-small cell lung cancer accounts for about 80–85%, while lung adenocarcinoma (LUAD) is the most common pathological type of non-small cell lung cancer[2]. Patients with lung cancer do not have typical symptoms in the early stages and many patients are already at an advanced stage once they are diagnosed, thus missing the best time for treatment. There are many treatment options for lung cancer, with surgery remaining the treatment of choice[3]. Although immunotherapy has benefited lung cancer patients in recent years, immunotherapy is slow to take effect and the efficiency rate is not high, with clinical trial data at around 20%, and patients' survival is significantly longer after it takes effect[4, 5]. In addition, some patients have developed resistance to treatment, resulting in a still low 5-year survival rate for patients with advanced lung cancer[6]. The currently accepted predictor of the efficacy of immunotherapy is the level of PD-L1 expression, and patients who are PD-L1 positive have better results when receiving immunotherapy[7, 8]. Therefore, the identification of new biomarkers predicting prognosis and response to immunotherapy is of great scientific value for the determination of the entire disease process and individualized treatment.

Chemokines are soluble proteins that bind to G protein-coupled receptors with 7 transmembrane structural domains[9], initiating the dissociation of G protein subunits α and βγ, which activate the mitogen-activated protein kinase, phospholipase C, and phosphatidylinositol 3 kinase signaling pathways, ultimately driving cell polarisation, adhesion, and migration[10, 11]. The dysregulation of chemokines and chemokine receptors has been reported to be closely associated with tumor progression[12]. On the one hand, CC and CXC chemokines regulate tumor angiogenesis, which is essential for tumor growth and metastasis. CXCL12 and its receptor CXCR4 promote glioma growth and angiogenesis by stimulating VEGF production[13], CXCR7 promotes angiogenesis by activating the AKT and ERK pathways in colon cancer cells[14], and CCL28, a hypoxia-inducible chemokine promotes LUAD angiogenesis by targeting CCR3[15]. On the other hand, chemokines and their receptors can influence tumor progression by stimulating the accumulation of infiltrating immune cells. For example, CCL3 and CCL20 promote the recruitment of CD8 + T cells and the progression of LUCD[16], while CCL7, which is highly expressed in LUAD, promotes the aggregation of M2 macrophages and cDC1 cells, thereby inducing anti-tumor immunity in LUAD[17]. In addition, chemokine receptors, such as CCR7[18], CCR9[19], CXCR4[20], and CCL18[21], are promising prognostic biomarkers for LUAD. Therefore, chemokine-related genes play a key role in predicting the prognosis and immunotherapy of LUAD.

LncRNAs are an important class of transcripts that contain more than 200 nucleotides and have a low capacity to encode proteins[22]. However, there is growing evidence that IncRNAs often exhibit aberrant expression in cancer and may act as important regulators, thereby influencing the expression of genes downstream associated with cancer development[23]. In addition, IncRNAs are involved in the regulation of many biological processes in tumorigenesis. IncRNA PVT1 promotes invasion of colon cancer cells[24]. In addition, many IncRNAs have been shown to act as biomarkers of LUAD prognosis[25, 26], for example. IncRNAs have also been reported to regulate immune cell infiltration in tumors[27, 28]. This suggests that IncRNAs can predict prognosis and immune therapy response in patients with LUAD. To date, studies on chemokine-related genes in LUAD have been reported[29], but no studies on chemokine-related IncRNAs in prognosis and immunotherapy in LUAD have been conducted.
In this study, we constructed a predictive signature based on CRLs and developed a nomogram to predict OS in patients with LUAD. The potential mechanisms of CRLs in LUAD were then further explored by functional enrichment analysis, immune-related functions analysis, and drug sensitivity analysis to provide a theoretical basis for diagnosing and treating LUAD.

Results

Screening of CRLs in LUAD

The flow of this study is shown in Fig. 1. A total of 501 patients with RNA-Seq data were downloaded from the TCGA-LUAD database. 16,876 IncRNAs were extracted from the TCGA database. 64 genes associated with chemokine were collected from the literature (S1 Table). The association of IncRNAs with chemokine-related genes was assessed by Pearson correlation analysis, resulting in the identification of 180 CRLs (S2 Table). 460 patients with complete survival information were included and randomly assigned to the training and test groups in a 1:1 ratio. There were no statistically significant differences between the training and test groups (Table 1). In the training group, 77 CRLs were obtained to be significantly associated with total OS by univariate Cox regression analysis ($p < 0.01$, Fig. 2C). 12 CRLs were further identified as significant prognostic factors by LASSO regression analysis (Fig. 2A-B). A prognostic model was constructed using 7 IncRNAs obtained by multivariate Cox regression analysis, including AL391261.1, AC034223.2, SH3BP5-AS1, LY86-AS1, AC104971.3, LINC01843, AL157388.1. The correlation between CRLs and chemokine-related genes was visualized by a heat map (Fig. 2D). Among the 7 IncRNAs, patients with high gene expression of AL391261.1, AC034223.2, and LINC01843 suggested poor prognosis, patients with high gene expression of LY86-AS1, SH3BP5-AS1, AC104971.3, and AL157388.1 suggested better prognosis. In the predictive signature, the formula for the patient's risk score is risk score $= (0.84 \times \text{AL391261.1 expression}) + (0.27 \times \text{AC034223.2 expression}) + (0.2 \times \text{LINC01843 expression}) - (1.38 \times \text{LY86-AS1 expression}) - (0.22 \times \text{SH3BP5-AS1 expression}) - (0.43 \times \text{AC104971.3 expression}) - (0.57 \times \text{AL157388.1 expression})$ (Table 2).
Table 1
Distribution of clinical patients in the training and test groups for each clinical characteristic.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Type</th>
<th>Total</th>
<th>Test</th>
<th>Train</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>&lt;=65</td>
<td>219(47.61%)</td>
<td>118(51.3%)</td>
<td>101(43.91%)</td>
<td>0.1094</td>
</tr>
<tr>
<td></td>
<td>&gt; 65</td>
<td>231(50.22%)</td>
<td>106(46.09%)</td>
<td>125(54.35%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknow</td>
<td>10(2.17%)</td>
<td>6(2.61%)</td>
<td>4(1.74%)</td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>FEMALE</td>
<td>251(54.57%)</td>
<td>121(52.61%)</td>
<td>130(56.52%)</td>
<td>0.4538</td>
</tr>
<tr>
<td></td>
<td>MALE</td>
<td>209(45.43%)</td>
<td>109(47.39%)</td>
<td>100(43.48%)</td>
<td></td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td>Stage I</td>
<td>249(54.13%)</td>
<td>129(56.09%)</td>
<td>120(52.17%)</td>
<td>0.6976</td>
</tr>
<tr>
<td></td>
<td>Stage II</td>
<td>105(22.83%)</td>
<td>54(23.48%)</td>
<td>51(22.17%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage III</td>
<td>74(16.09%)</td>
<td>33(14.35%)</td>
<td>41(17.83%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage IV</td>
<td>24(5.22%)</td>
<td>11(4.78%)</td>
<td>13(5.65%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknow</td>
<td>8(1.74%)</td>
<td>3(1.3%)</td>
<td>5(2.17%)</td>
<td></td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>T1</td>
<td>156(33.91%)</td>
<td>82(35.65%)</td>
<td>74(32.17%)</td>
<td>0.8203</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>246(53.48%)</td>
<td>121(52.61%)</td>
<td>125(54.35%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>37(8.04%)</td>
<td>17(7.39%)</td>
<td>20(8.7%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>18(3.91%)</td>
<td>10(4.35%)</td>
<td>8(3.48%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknow</td>
<td>3(0.65%)</td>
<td>0(0%)</td>
<td>3(1.3%)</td>
<td></td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>M0</td>
<td>310(67.39%)</td>
<td>150(65.22%)</td>
<td>160(69.57%)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>23(5%)</td>
<td>11(4.78%)</td>
<td>12(5.22%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknow</td>
<td>127(27.61%)</td>
<td>69(30%)</td>
<td>58(25.22%)</td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>N0</td>
<td>297(64.57%)</td>
<td>148(64.35%)</td>
<td>149(64.78%)</td>
<td>0.2301</td>
</tr>
<tr>
<td></td>
<td>N1</td>
<td>86(18.7%)</td>
<td>47(20.43%)</td>
<td>39(16.96%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>64(13.91%)</td>
<td>27(11.74%)</td>
<td>37(16.09%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N3</td>
<td>2(0.43%)</td>
<td>2(0.87%)</td>
<td>0(0%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknow</td>
<td>11(2.39%)</td>
<td>6(2.61%)</td>
<td>5(2.17%)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2

Seven chemokine-related prognostic IncRNAs significantly associated with OS

<table>
<thead>
<tr>
<th>id</th>
<th>coefficient</th>
<th>HR</th>
<th>HR.95L</th>
<th>HR.95H</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL391261.1</td>
<td>0.84</td>
<td>2.81</td>
<td>1.82</td>
<td>4.33</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>AC034223.2</td>
<td>0.27</td>
<td>1.58</td>
<td>1.32</td>
<td>1.89</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>SH3BP5-AS1</td>
<td>-0.22</td>
<td>0.64</td>
<td>0.50</td>
<td>0.82</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>LY86-AS1</td>
<td>-1.38</td>
<td>0.05</td>
<td>0.01</td>
<td>0.53</td>
<td>0.0124</td>
</tr>
<tr>
<td>AC104971.3</td>
<td>-0.43</td>
<td>0.56</td>
<td>0.39</td>
<td>0.80</td>
<td>0.0013</td>
</tr>
<tr>
<td>LINC01843</td>
<td>0.20</td>
<td>1.29</td>
<td>1.08</td>
<td>1.55</td>
<td>0.0042</td>
</tr>
<tr>
<td>AL157388.1</td>
<td>-0.57</td>
<td>0.36</td>
<td>0.19</td>
<td>0.68</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Prognostic values for the training group, the test group, and all groups

Prognostic risk scores were calculated for each patient using the formula above and LUAD patients were divided into high- and low-risk groups based on the median risk score. As can be seen from the risk score point plots, we found that the higher the risk score, the higher the number of deaths (Fig. 3A-C). In addition, the heat map shows the expression of 7 CRLs in the high-risk and low-risk groups (Fig. 3D-F). K-M analysis showed that among LUAD patients, the high-risk group had a worse prognosis than the low-risk group (Fig. 3G-I). We also used ROC curve analysis to determine the prognostic accuracy of this signature. We validated the prognostic power and accuracy of CRLs in the training group with an AUC value of 0.796, significantly higher than gender, age, and stage (Fig. 3J-L). Univariate Cox regression results showed that risk score (p < 0.001), and stage (p < 0.001) were all statistically significant in predicting the prognosis (Fig. 4A). Multivariate Cox regression results showed that risk score [P < 0.001, hazard ratio (HR) = 2.287,95% CI = 1.859–2.812] and stage [P < 0.001, hazard ratio (HR) = 1.549,95% CI = 1.337–1.795] would be independent risk factors for LUAD prognosis (Fig. 4B). We divided the clinicopathological characteristics into subgroups by Age, Gender, T, N, and clinical stage, and then performed a K-M analysis for each subgroup. The results showed that the prognosis of high-risk patients was significantly more unfavorable in several subgroups, including age < 65, age > 65, female, male, N0, N1 + N2 + N3, stage I + II, stage III + IV, T1 + T2, T3 + T4(Fig. 5). This suggests that this prognostic feature is valid in assessing the prognosis of LUAD patients.

Nomogram

To further validate the efficacy of the risk model in predicting prognosis over other clinical features, ROC curve analysis and C-index analysis were conducted at 1, 3, and 5 years. The C-index result indicated that this predictive feature was better than other clinical features in predicting the prognosis of LUAD patients (Fig. 4C). The AUC values of the ROC curves for LUAD patients at 1, 3 and 5 years were 0.792, 0.727, and
0.787 respectively (Fig. 4D). To assess the potential clinical utility of the prognostic features constructed based on the 7 CRLs, we created a nomogram with risk and clinicopathological characteristics to predict 1, 3, and 5-year OS rates in patients with LUAD. We observed that a patient with a total score of 161 had a predicted 1-year OS rate of 0.865, a 3-year OS rate of 0.539, and a 5-year OS rate of 0.274 by nomogram (Fig. 4E), with the higher the patient's total score, the worse the outcome. We then constructed calibration curves to assess the accuracy between the nomogram predicted OS rates and observed OS. The results showed a good agreement with the predicted OS at 1, 3, and 5 years (Fig. 4F).

**PCA and Functional enrichment analysis**

Based on all gene expression groups, chemokine-associated gene groups, CRLs groups, and chemokine-associated prognostic IncRNA groups, PCA analysis was used to find a clear distribution between the high- and low-risk groups (Fig. 6). GO analysis showed that CRL was enriched in humoral immune responses, antimicrobial humoral responses, leukocyte proliferation, and lymphocyte proliferation (Fig. 7A). KEGG analysis showed that these were associated with CRL with primary immunodeficiency, hematopoietic cell lines, complementary and coagulation cascades, and arachidonic acid metabolism (Fig. 7B). We also analyzed immune-related functions to assess the immune status of the low- and high-risk groups and the results showed that the low-risk group was significantly more likely than the high-risk group in these immune functions, including Type II Interferon (IFN) Response, human leukocyte antigen (HLA), Antigen-presenting cells (APC) co-stimulation, CCR, Parainflammation, Cytolytic_activity, Inflammation-promoting, T cell co-inhibition, Checkpoint, T cell co-stimulation (Fig. 7C). Finally, we further compared the expression levels of immune checkpoints and immunosuppressive factors in the high-risk and low-risk groups to assess the potential predictive value of CRLs in current immune checkpoint blockade therapy. We found that the high-risk group had lower levels of immune checkpoint gene expression, including CTLA4 and PDCD1 (Fig. 7D). This suggests that these IncRNAs may regulate tumor progression by being involved in the immune response.

**Tumor mutation burden (TMB) and drug sensitivity analysis**

We used the “maftools” algorithm to assess the differences in mutations between the high-risk and low-risk groups and showed that for most genes, the frequency of mutations was higher in the high-risk group than in the low-risk group (Fig. 8A-B). In addition, the TMB was significantly higher in the high-risk group than in the low-risk group (Fig. 8C), further analyzing the difference in survival between patients with high-TMB and low-TMB. As shown in Fig. 8E-F, the high TMB group had significantly better OS than the low TMB group (p < 0.05), and the low-TMB combined with the high-risk group had the worst OS compared to the other groups. We then also investigated the difference in sensitivity to immunotherapy between patients in the high-risk and low-risk groups. We found higher TIDE scores in the low-risk group than in the high-risk group (Fig. 8D). Finally, we used the “pRRophetic” package to screen for potentially effective antitumor agents, including cisplatin, talazopanib, epramycin, and etanercept (Fig. 9).
Discussion

Lung cancer is a common cause of cancer-related deaths worldwide[30], and due to the lack of technical means for early diagnosis, many patients are diagnosed at an advanced stage and lose the best opportunity for treatment. With extensive research into the pathogenesis of lung cancer, immunotherapy has been shown to extend the overall survival time of many cancer patients. However, there are still a large number of patients who do not benefit from these treatments, and in some cases even develop resistance and relapse during the course of treatment[31]. Therefore, there is an urgent need to find more prepared biomarkers to screen for populations benefiting from immunotherapy and to achieve better-individualized treatment.

Due to the high heterogeneity of LUAD and the complexity of the tumor microenvironment (TME), it is often difficult to effectively assess the prognostic value of a single biomarker in LUAD[32]. The development of multi-gene prognostic assessment models by screening a series of tumor-related functional genes has become a hot research model. Prognostic assessment models related to aspects such as immunity[33], ferroptosis[34], autophagy[35], hypoxia[36], and apoptosis[37] have been established and used as prognostic markers and therapeutic targets for lung cancer. However, chemokine-based related studies still lack further exploration. In this study, we have developed a risk model based on CRLs. 7 CRLs were identified in this risk model, including AL391261.1, AC034223.2, SH3BP5-AS1, LY86-AS1, AC104971.3, LINC01843, AL157388.1. Among these lncRNAs, SH3BP5-AS1, LY86-AS1, and LINC01843 were involved in the construction of other prognostic models in a variety of tumors and were associated with immune responses. In contrast, we found that SH3BP5-AS1 was involved in the construction of an immune-related model, which had an AUC of 0.777[38]. LINC01843 was involved in the construction of an iron death-related model, which had an AUC of 0.741[39]. The AUC of our constructed model exceeded these models. This is not contradictory, and our conclusions can complement each other. Furthermore, it also shows that our model has some superiority. We used the median of the risk scores as a cut-off, dividing all patients into a high-risk group and a low-risk group. Survival analysis revealed that the OS was worse in the high-risk group than in the low-risk group among LUAD patients. The AUC values for the 1, 3, and 5-year ROC curves for LUAD patients were 0.792, 0.727, and 0.787 respectively. Univariate and multivariate Cox regression analyses found that risk score was an independent factor affecting the prognosis of LUAD patients. In addition, across different clinical subgroups, we also found that patients in the same subgroups with high risk remained survival unfavorable. It shows that the model can accurately predict the prognosis of patients, bridging the gap between TNM staging in predicting patient prognosis. To visualize the clinical application of this feature, we created a nomogram based on risk and clinical-pathological characteristics and obtained the OS of patients at 1, 3, and 5 years by summing the scores of each characteristic. The calibration curve shows a high agreement between predicted and actual OS. These results suggest that our risk model is effective in predicting the prognosis of patients with LUAD.

To explore whether the CRL risk model also plays a role in the tumor immune microenvironment. We performed functional enrichment analyses and found that CRLs were enriched in humoral immune
responses, antimicrobial humoral responses, leukocyte proliferation, and lymphocyte proliferation. Chemokines are widely known for their ability to stimulate cell migration, particularly leukocytes[40], and studies have shown that immune monitoring is prone to failure in the absence of chemokine-directed leukocyte migration[41]. In addition, some chemokines have direct antimicrobial activity[42, 43]. This also suggests that chemokine-associated IncRNAs also have immunomodulatory functions. Therefore, we further investigated the differences in immune-related functions between the high-risk and low-risk groups and found that the low-risk group exhibited strong immune functions and we also observed that the expression of immune checkpoints was generally higher in the low-risk patients than in those in the high-risk group. It was shown that excessive immune function activates the expression of immune checkpoints, thus blocking the antigen presentation process in tumor immunity and ultimately inhibiting the immune function of T cells and escaping immune surveillance[44]. This also suggests that patients in the low-risk group, who receive immunotherapy, have a worse outcome. To test this hypothesis, we compared the TIDE scores of the high- and low-risk groups and found that the low-risk group had higher TIDE scores than the high-risk group, which also suggests that the high-risk group is more sensitive to immunotherapy. In addition, TMB is one of the predictors of immune response, and it is generally accepted that patients with high TMB have better outcomes with immunotherapy[45, 46]. Our results also found a higher TMB in the high-risk group of patients than in the low-risk group. We also performed a survival analysis of TMB and showed that patients with high TMB had a longer OS than those with low TMB, and that patients in the high-risk group combined with those with low TMB had the worst survival outcome. Based on these results, we believe that the CRL risk model can accurately predict immunotherapy sensitivity in LUAD. Drug sensitivity screening was also performed based on CRL signature. Sensitivity to cisplatin, talazopanib, epramycin, and etanercept was significantly higher in the high-risk group compared to the low-risk group. These drugs may be used in combination with immunotherapy to further improve outcomes in patients with LUAD. However, further clinical studies are needed to confirm this.

There are also certain limitations in this study. Firstly, it is necessary to explore the mechanisms of chemokine-related IncRNAs in LUAD tumorigenesis and development through basic biological studies. Secondly, this is a retrospective study from a single database and therefore external validation and further prospective studies are needed to help increase the credibility.

Conclusions

We identified a novel CRL signature that has clinical value in predicting prognosis and response to immunotherapy in patients with LUAD. CRLs may have a potential role in antitumor therapy and hold promise as therapeutic targets for LUAD.

Material and methods

Data collection
RNA-Sep data for LUAD patients and their corresponding clinical information and mutation data were obtained from The Cancer Genome Atlas (TCGA). 501 LUAD samples and 54 normal samples were included in the dataset, excluding patients with less than 30 days of survival and those with lost survival time, resulting in a total of 460 LUAD patients. Human GTF files downloaded from the Ensembl database(http://asia.ensembl.org) re-annotated the transcriptome data to obtain mRNA and lncRNA expression data. 64 chemokine-related genes were identified by reviewing four reviews on chemokines or chemokine receptors were obtained[47–50].

**Screening of CRLs in LUAD**

Based on R (ver. 4.1.2), Co-expression correlation analysis of lncRNA and chemokine-related gene expression profiles was performed using the “limma” package to obtain CRLs, with a screening condition of |R| > 0.5, p < 0.001.

**Construction of a CRL Signature for predicting prognostic**

The 460 LUAD patients were randomized in a 1:1 ratio to either the training or test group. In the training group, univariate Cox analyses based on a “survival” package were performed to identify potential prognostic CRLs. 1000 lasso regression analyses and multivariate Cox regression analyses were then performed to determine the best prognostic IncRNA for model building and to calculate a risk score with the formula: Risk Score = [where Xi means the correlation coefficient, Yi means the expression level for IncRNAi, and i means the number of IncRNA]. Based on the median of the risk score, all patients were divided into high-risk and low-risk groups.

**Survival analysis**

Based on R (ver. 4.1.2), we compared the OS between the high-risk and low-risk groups using the “survival” and “survminer”. Then, the ROC and AUC were also performed using the “sevivalROC” package to assess the sensitivity and specificity. Independent prognostic significance of prognostic signature assessed by the univariate and multivariate Cox regression analysis. p < 0.05 was considered statistically different.

**Nomogram**

Based on R (ver. 4.1.2), we constructed a nomogram using the “rms” package to predict 1-, 3-and 5-year OS rates in LUAD patients. We also constructed calibration curves to assess the agreement between the OS rate predicted by the nomogram and the observed OS rate.

**Principal component analysis (PCA) and Functional enrichment analysis**

Based on R (ver. 4.1.2), we used the “scatterplot3d” package to assess the status of the high-risk and low-risk groups by dimensionality reduction. We used the “ggplot2” package to perform Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis to investigate whether these genes were differentially expressed between the high-risk and low-risk groups.
**Immune-related functional analysis and immune checkpoint analysis**

Based on R (ver. 4.1.2), we used the "GSVA" package [51] to analyze differences in immune-related functions in LUAD patients and the "ggplot2" and "ggpubr" packages to analyse the expression levels of multiple immune checkpoints in the high- and low-risk groups. Differences between groups were statistically assessed using Wilcox. test.

**Tumor Mutational Burden (TMB)**

TMB refers to the number of non-synonymous mutations in somatic cells in a particular genomic region, usually expressed in terms of how many mutations per megabase (mut/Mb)[52]. Based on R (ver. 4.1.2), we first screened the top 15 genes with high mutation frequencies between the high-risk and low-risk groups by the “maftools” package, then carried out the differential expression of TMB between the high-risk and low-risk groups, and finally drew the survival curve of TMB combined with risk.

**Immune response and drug sensitivity analysis**

We downloaded tumor immune dysfunction and exclusion (TIDE) data for non-small cell lung cancer (NSCLC) from TIDE (http://tide.dfci.harvard.edu/). TIDE was used to assess the possibility of tumor immune escape in the gene expression profile of tumor samples, with a higher TIDE prediction score indicating a higher likelihood of immune evasion, indicating that patients are less likely to benefit from immunotherapy.[53]. Therapeutic agents were screened and drug sensitivity was assessed using the “pRPophetic”, “ggplot2”, and “ggpubr” packages, |R|>0.3, p < 0.001.

**Statistical Analysis**

All statistics were obtained using R version 4.1.2. A chi-square test was used to compare the categorical variables between the training and test groups and to calculate p-values. Survival curves were plotted using the Kaplan-Meier method. Correlations between risk and immune checkpoint expression, TMB and drug sensitivity were explored by Pearson correlation analysis. Univariate and multivariate Cox regression analyses were performed to determine the independent prognostic value of risk scores in combination with other clinical characteristics. p < 0.05 indicates statistical significance.

**Data Availability**

The data that support the findings of this study are openly available at https://portal.gdc.cancer.gov/repository http://tide.dfci.harvard.edu/, further inquiries can be directed to the corresponding authors.

**Declarations**

**Data Availability**
The data that support the findings of this study are openly available at https://portal.gdc.cancer.gov/repository http://tide.dfci.harvard.edu/ , further inquiries can be directed to the corresponding authors.

Funding Statement

This work was supported by the grants from the National Natural Science Foundation of China (No. 81960424).

Author contribution

DY and YL conceived the study. KZ and LX wrote the paper and drew the figures. JX and YZ analyzed the data. KZ and DY edited and reviewed the paper. All the authors contributed to the study and approved the submitted version of the article.

Declaration of Figures Authenticity

All figures submitted have been created by the authors who confirm that the images are original with no duplication and have not been previously published in whole or in part.

Ethics approval and consent to participate

None.

Conflicts of Interest

there is no conflict of interest regarding the publication of this paper.

References


Figures
Figure 1
Flow chart of this study
Figure 2

Construction of the risk model. The lasso regression analysis was conducted for the construction of the risk model (A, B). A total of 77 chemokine-related IncRNA were identified in the risk model (C). The correlation between chemokines and IncRNAs in the risk model (D). Red means positive correlation, and blue means negative correlation.
Figure 3

Validation of the risk model in the training group test group and all group. Survival status of the patients in high-risk and low-risk groups (A, B, C). The expression of the chemokine-related IncRNAs in the risk model was shown by using a heatmap (D, E, F). Survival analysis of the training group test group and all groups (G, H, I). Roc curves were plotted to assess the accuracy of the risk model in the training group test group and all groups (J, K, L). The Low or high type represents the patients with low risk or high risk.
Figure 4

Independent prognosis value of the risk model. Univariate analysis and multivariable analysis were conducted to validate the independent prognosis value of the model (A, B). ROC curves were performed to validate the superiority of the risk score in predicting patient’ survival (C). Decision curve analysis (DCA) was conducted to confirm the superiority of the risk score in the clinical application (D). Nomogram was plotted for the prediction of overall survival time (E) in LUAD patients. The calibration curves and ROC curves were further plotted to determine the accuracy of the nomogram for OS at 1, 3 and 5 years, respectively (F).
Figure 5

Kaplan–Meier survival curves for OS in LUAD patients with high- and low-risk stratified by different clinical factors. Kaplan–Meier survival for OS in subgroups stratified by age ≤ 65(A), age > 65(B), FEMALE(C), MALE(D), N0 (E), N1-3 (F), stage I-II (G), stage III-IV (H), T1-2 (I) and T3-4 (J) in LUAD patients.
Figure 6

PCA analysis. PCA analysis observed the distribution of patients according to all genes (A). Genes associated with Chemokine (B). IncRNAs associated with chemokine (C). Risk IncRNAs (D).
Figure 7

Correlation between immune microenvironment and risk model. GO (A) and KEGG (B) analysis based on the chemokine-related lncRNAs risk model. The difference in the enrichment of thirteen immune-related pathways between the low-risk group and the high-risk group was assessed (C). Correlation between the risk model and checkpoint in LUAD (D).
Figure 8

Correlation between the risk model and tumor mutation burden (TMB) in LUAD. The top 15 genes’ TMB in the high-risk group and low-risk groups (A–B).

A violin plot was used to visualize the TMB level between the high-risk group and low-risk groups (C). Tumor immune dysfunction and exclusion (TIDE) algorithm analysis for high-risk and low-risk groups (D).
The survival difference between the high-TMB and low-TMB groups (E). The survival status of patients with low or high risk in the high-TMB group and low-TMB group (F).

Figure 9

Observed the drug sensitivity of (A) BMC-509744, (B) BMC-754807, (C) Cisplatin, (D) Epothilone, (E) Talazoparib, (F) GSK-650394, (G) OSU-03012, (H) Pyrimethamine, (I) RO-3306, (J) Phenformin and (K)
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- S1Table.docx
- S2Table.docx
- S3table.docx
- S4Table.docx