Heliyon 8 (2022) e11399

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Comprehensive analysis of chemokines family and related regulatory ceRNA network in lung adenocarcinoma

Yifan Hu [a,b,c,d,1], Mintao Xiao [a,b,c,1], Duoli Zhang [a,b,c], Jing Shen [a,b,c], Yueshui Zhao [a,b,c], Mingxing Li [a,b,c], Xu Wu [a,b,c], Yu Chen [a,b,c], Zhigui Wu [a,b,c,d], Hongli Luo [d], Zhangang Xiao [a,b,c,*], Fukuan Du [a,b,c,**]

[a] Laboratory of Molecular Pharmacology, Department of Pharmacology, School of Pharmacy, Southwest Medical University, Luzhou, Sichuan, China
[b] Cell Therapy & Cell Drugs of Luzhou Key Laboratory, Luzhou, Sichuan, China
[c] South Sichuan Institute of Translational Medicine, Luzhou, Sichuan, China
[d] Department of Pharmacy, Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, China

ARTICLE INFO

Keywords:
Chemokine family
Lung adenocarcinoma
Prognosis
Immune
Competing endogenous RNA
Bioinformatics

ABSTRACT

Lung adenocarcinoma (LUAD) is one of the world's commonest malignancies with a high fatality rate. Chemokines not only regulate immune response but also participate in tumor development and metastasis and yet the mechanism of chemokines in LUAD remains unclear. In this study, transcriptional expression profiles, mutation data, and copy number variation data were downloaded from The Cancer Genome Atlas (TCGA). Risk gene protein expression was assessed by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) and the Human Protein Atlas (HPA). Gene Expression Omnibus (GEO) data was used to validate the prognostic model. We summarized the genetic mutation variation landscape of chemokines. The risk prognosis model was developed based on differentially expressed chemokines, and patients in the high-risk score (RS) group had lower survival rates. Gene Set Enrichment Analysis (GSEA) revealed that high-RS patients were associated with metabolic transformation pathways, while low-RS patients were associated with immune-related pathways. Compared with the high-RS group, the low-RS group had higher immune/stromal/estimate scores calculated by the ESTIMATE package. The proportion of immune cells obtained using the CIBERSORT package was significantly different between the two groups. Most of the immune checkpoints were highly expressed in low-RS samples. Finally, we discovered that the lncRNA MIR17HG/AC009299.3/miR-21–5p/CCL20 regulatory network might be crucial in the pathogenesis of LUAD. In conclusion, we developed a risk signature and chemokine-related competing endogenous RNA (ceRNA) network.

1. Introduction

Globally, the death toll of lung cancer far exceeds that of other cancer types, ranking first in the number of cancer deaths [1]. According to the morphological properties of lung cancer cells, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) can be distinguished [2]. The vast majority of lung cancer types is NSCLC, accounting for about 85% [3]. As the environment changes, the proportion of LUAD increases year by year and has occupied the majority of NSCLC [4]. Despite remarkable clinical achievements, the five-year survival rate for lung cancer remains low at approximately 21% [1]. Individualized treatment plans that take into account the patient's actual condition can improve treatment outcomes and reduce adverse effects. Therefore, it is urgent to explore the characteristics and mechanisms of lung adenocarcinoma to develop individual precise treatment strategies.

Immunotherapy is known to be the most popular therapy for cancers. In recent years, Food and Drug Administration has approved some immune checkpoint inhibitors (ICIs) as first-line therapeutic agents for advanced NSCLC. Tumor cells can upregulate the expression of checkpoint molecules, leading to immune escape by unresponsive cytotoxic T cells in the tumor microenvironment (TME) [5]. ICIs enhance the anticancer immune response by targeting and antagonizing immune checkpoint molecules

* Corresponding author.
** Corresponding author.
E-mail addresses: zhangangxiao@swmu.edu.cn (Z. Xiao), dufukuan@swmu.edu.cn (F. Du).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2022.e11399
Received 21 February 2022; Received in revised form 31 May 2022; Accepted 28 October 2022

such as PD-L1, PD-1, CTLA-4, and LAG-3. Nevertheless, only one-third of patients in NSCLC can benefit from ICI treatment [6], suggesting that TME is heterogeneous and understanding tumor immune heterogeneity is essential for the development of effective therapy.

Inflammatory mediators are important components of TME and promote tumor development [7]. Chemokines are signaling proteins with a small molecular weight that are released by cells, inducing the directional chemotaxis of nearby reaction cells, which play a key role in inflammation and immunity [8]. They can be expressed by tumor cells, immune cells, and stromal cells. Under the action of specific chemokines, different subsets of immune cells migrate to TME to regulate the immune response [9]. Recently, chemokines have emerged as potential targets for cancer immunotherapy and prognostic biomarkers [10, 11]. Low expression of CCL14 is related to a bad outcome and various types of immune cell infiltration in hepatocellular carcinoma [11]. High expression of CXCL13 and CXCR5 in breast cancer is associated with clinical stages, cancer metastasis, and poor prognosis [12]. However, there has been no comprehensive analysis and screening of chemokines as prognostic risk characteristics of LUAD yet. Therefore, we systematically analyzed the chemokines' expression and constructed a risk prognosis model and related regulatory network. Considering the important role of chemokines in immune regulation, the relationship between the risk prognosis model and immune score, immune checkpoint expression, and immune cell infiltration was further analyzed. Our findings may add to the evidence that chemokines play a crucial role in the genesis and progression of LUAD.

2. Materials and methods

2.1. Data procession and differential expression analysis

The microRNAs (miRNA), long non-coding RNA (lncRNA), messenger RNA (mRNA) sequencing data and corresponding clinical data of LUAD were downloaded from the TCGA database (https://portal.gdc.cancer.gov). Somatic mutation data and copy number variation (CNV) data were downloaded from TCGA and the University of California, Santa Cruz (UCSC) *Xena website* (https://xenabrowser.net/datapages/), respectively. Transcriptome data of mRNA and lncRNA included 535 tumor tissues and 59 normal tissues. Transcriptome data of miRNA included 521 tumor tissues and 46 normal tissues. 513 clinical samples were included after excluding those with missing survival time. The data (226 tumor samples) for validating the prognostic model was acquired from the early lung adenocarcinoma expression data uploaded by Kohno T et al in the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31210) (GSE31210) [13,14]. The Chemokine ligands and receptors were derived from previous publications [15]and summarized in Supplementary Table 1. R package "DEseq2" was used to analyze the differential expression of chemokines in lung adenocarcinoma tissues and normal tissues. The thresholds were set as $|\log2 \text{ fold change (FC)}| > 1$, false discovery rate (FDR) < 0.05. To verify whether the risk genes' mRNA expression is consistent with that protein expression, we downloaded the lung adenocarcinoma protein expression data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC, https://proteomics.cancer.gov/data-portal) [16] with study ID: PDC000219 and analyzed CXCL13, CCL20, CX3CR1 protein expression in tumor and normal tissues. Meanwhile, as a complement to the protein detection method, the Human Protein Atlas (HPA, https://www.proteinatlas.org/) [17] demonstrated the results of CXCL13, CX3CR1, CXCL4L1 immunohistochemical staining in LUAD tissues and normal tissues. Unfortunately, CXCL4L1 expression data was not available in CPTAC, and HPA did not have immunohistochemical staining maps for CCL20.

2.2. Mutation and copy number variations analysis of chemokines

The "maftools" package was used to present the mutation landscape of chemokines in LUAD. The location of CNV alteration of chemokines on chromosomes was plotted by the "RCircos" package.

2.3. Construction of risk prognosis model

We conducted a univariate regression analysis on the chemokines with differential expression in TCGA-LUAD and selected the chemokines with that P value less than 0.2 in the univariate analysis for inclusion in the multivariate analysis. Finally, the risk prognosis model of the four genes was established. In both the training (TCGA) and validation (GEO) datasets, the formula was used to generate the risk score for each patient. The formula is shown below:

$$\text{Risk score} = \beta_1 \times Expression_1 + \beta_2 \times Expression_2 + \cdots + \beta_i \times Expression_i$$

(β is the regression coefficient of each risk gene and Expression is the genes' expression).

The median risk score was used to separate patients into the high-RS group and low-RS group in the TCGA (n = 501) and GEO (n = 226) studies. Univariate and multivariate analyses of risk score and clinical parameters were performed to assess the risk ratios for clinical parameters. 3, 5-year receiver operator characteristic curve (ROC) curve was created to ensure the model accurate. Nomogram was conducted by "rms" to predict the 3, 5, and 7-year survival of patients. A vertical line determined the point for each variable, and the points for all variables were added up to give a total point, which allowed to predict survival rates. The calibration curve was set up to verify the prediction performance. The closer the curve slope is to 1, the better the prediction effect is. The random sampling was repeated 600 times.

2.4. Relationships of risk gene signature with immune score and immune cells

GSEA can evaluate the distribution tendency of genes in a predetermined gene set in the gene table sorted by phenotypic correlation to judge their contribution to phenotype [18, 19]. KEGG gene set were from the Molecular Signatures Database (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). GSEA analysis was performed adopting the reference gene set 'c2. cp.kegg.v7.4. symbols.gmt' at 1,000 number of permutations by JAVA program. In this study, the input to GSEA was a matrix of all mRNA expression in TCGA-LUAD, the samples were divided into high-risk and low-risk groups based on the median risk score of 1.0, and all mRNA genes were ranked according to enrichment score (ES). Pathways with ES > 0 were enriched in the high-risk group, and pathways with ES < 0 were enriched in the low-risk group. The ESTIMATE package predicted stromal and immune cell scores from expression profile data to estimate the percentage of each type of cell and the tumor purity of the samples [20]. To learn more about how the immune system changes in cancer patients, the relative proportions of immune cells were calculated using CIBERSORT. We described the proportion of immune cell infiltration with reference to LM22, which includes 547 genes for differentiating twenty-two human hematopoietic cell phenotypes [21]. The perm was set at 1000. Only Samples with $P < 0.05$ were retained.

2.5. Construction of competing endogenous RNA (ceRNA) network

CCL20 was identified as the best prognostic factor among the four risk genes according to prognosis and clinical stage. Therefore, the ceRNA network in this study was used to reversely predict the corresponding miRNA and lncRNA through CCL20. TarBasev.8 [22] (https://dianalab.e-ce.uth.gr/html/diana/web/index.php?r=tarbasev8), miRTarBase [23] (http://mirtarbase.cuhk.edu.cn/) and TargetScan [24] (https://www.targetscan.org/vert_72/) were used to anticipate miRNA targets that bind to mRNA. LncBase Predicted v.2 [25] (https://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbasev2/index-predicted) and StarBase [26] (https://starbase.sysu.edu.cn/) were used to predict the lncRNAs targeting miRNAs. The study workflow was shown in Figure 1.

2.6. Statistical analysis

R software was used to conduct statistical analyses. Wilcoxon test was applied to compare differences between two sets of data while Kruskal-Wallis test was applied to compare three or more sets of data. For survival analysis, the optimal cutoff values of all variables were defined by the "survminer" R package, and P-value was calculated with log-rank test. The correlation between the two groups of continuous variables was analyzed by Pearson test. The meanings represented by statistical symbols are respectively: ns, not significant; *P < 0.05; **P < 0.01; ***P < 0.001. P < 0.05 was regarded as statistically significant.

3. Results

3.1. Chemokine genetic variation landscape in TCGA-LUAD

Figure 2(a) summarized the type and frequency of all chemokine somatic cell mutations. Missense mutation and SNP ranked as the top variant type. C > A was the most common type in SNV class. ACKR1 exhibited the highest mutation frequency with 14% (Figure 2(b)). Further analysis showed that there were significant co-occurrence mutation relationships between CCL20 and CXCR5, CCL19 and CCL17, CCL2 and CCL23, etc (Figure 2(c)). The location of chemokine CNV variations on 23 chromosomes is depicted in Figure 2(d). There is a prevalent CNV alteration in chemokines and more than half of them were focused on the amplification in copy number (Figure 2(e)). On the whole, the mutation rate and CNV frequency of atypical chemokine receptors ACKR1 are highest in the chemokine family.

3.2. Construction of risk prognostic model based on differentially expressed chemokines in TCGA-LUAD

There were 19 up-regulated genes and 12 down-regulated genes in tumor tissues by the standard of |log2(FC)| > 1 and FDR <0.05 (Supplementary Table 2). Heatmaps of these differentially expressed genes were displayed in Figure 3(a). The genes with P < 0.2 in the univariate analysis were selected out for inclusion in the multivariate analysis [27]. In this regard, we believe that appropriately relaxing the inclusion criteria to P < 0.2 can effectively avoid omitting some important variables. Although they are not statistically significant in univariate analysis, their real effects may be underestimated or covered up. The forest plot showed the regression coefficient, hazard ratio and P-value of the chemokines that meet P < 0.2 in the univariate analysis (Figure 3(b)). Then the model was optimized by multivariate Cox regression, and four chemokines were obtained to develop risk prediction model (Table 1). To identify the gene expression of the four chemokines at the protein level, we analyzed the protein expression of LUAD tissues and adjacent normal tissues using CPTAC data. The results showed that CXCL13 and CCL20 protein expression was significantly higher in LUAD tissues than in normal tissues, and the opposite was true for CX3CR1 (Figure 3(c)). As shown in Figure 3(d), immunohistochemical staining from HPA indicated that CXCL13 protein expression was upregulated and CX3CR1 protein expression was downregulated in LUAD tissues, and CXCL4L1 was not significantly different in the two tissues. It is clear from our expression analysis that both mRNA and protein expression of CXCL13 and CCL20 are highly expressed in LUAD while CX3CR is the opposite and the mRNA expression level of CXCL4L1 is highly expressed in lung adenocarcinoma tissues.

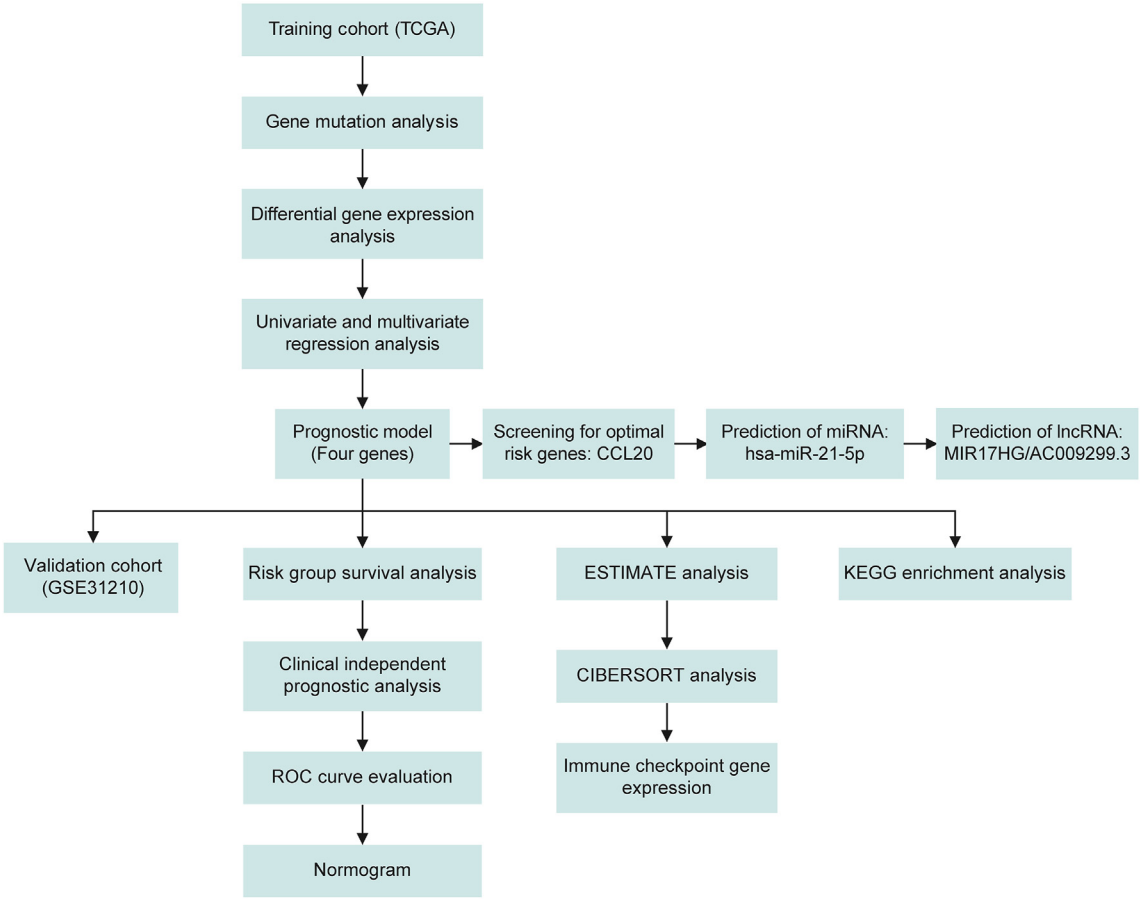

Figure 1. The workflow of the study.

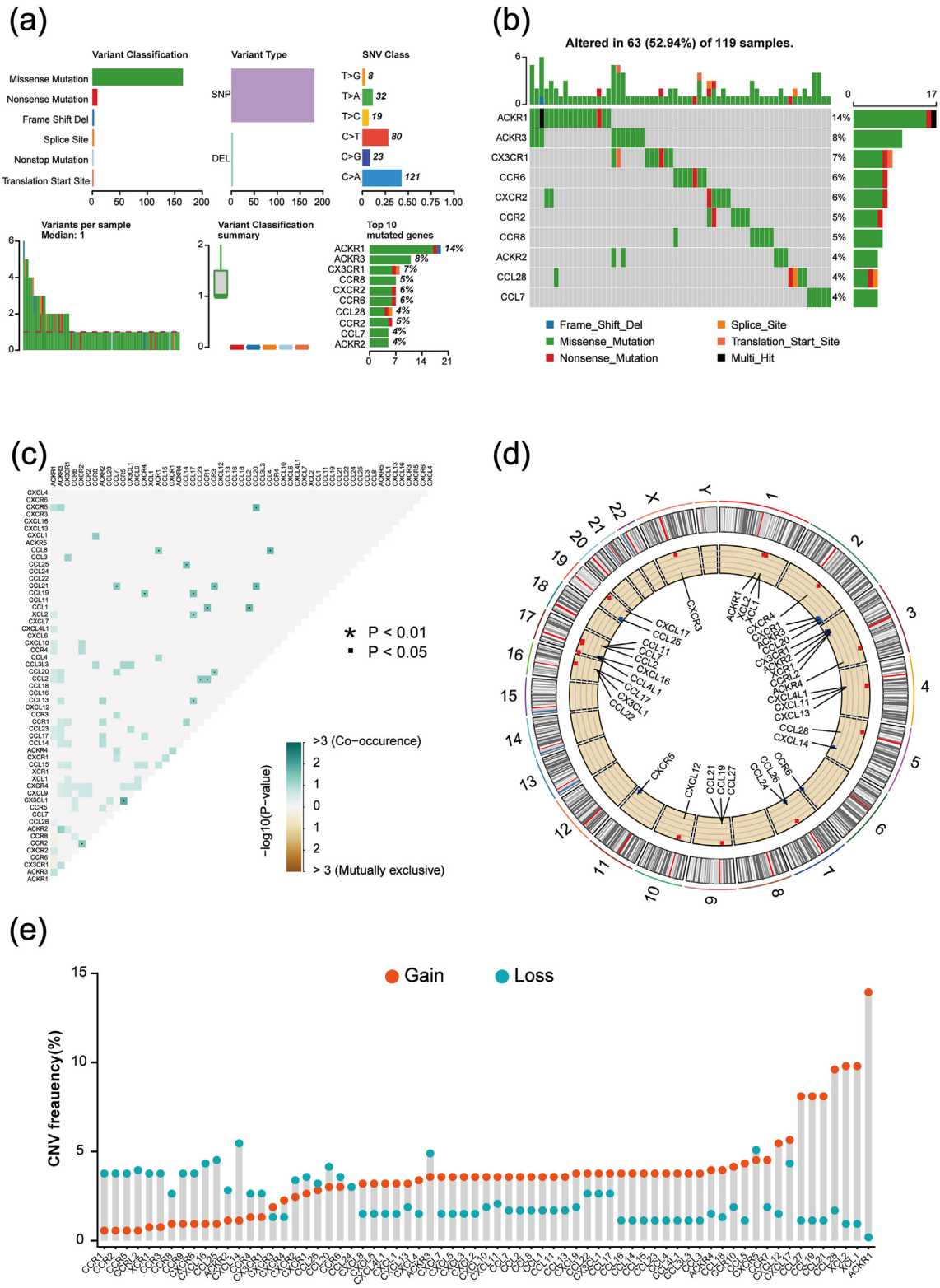

Figure 2. Chemokine genetic variation landscape in TCGA-LUAD. (a) Chemokine mutation classification in LUAD. **(b)** The mutation frequency of the top ten chemokines. Each column on the panel represented an individual patient, with different colors representing variant types. The percentage of mutation frequency for each gene was shown on the right. The total mutation burden (TMB) of each sample showed in the upper bar plot. **(c)** The mutation co-occurrence and exclusion analyses of chemokines. **(d)** Chemokine CNV alterations were observed on twenty-three chromosomes. **(e)** The CNV frequency of all chemokines. DEL deletion; SNP, single nucleotide polymorphism.

In the heat map, the mean expression of CCL20 was higher in high-RS group, while the opposite was true for CXCL13, CX3CR1 and CXCL4L1 (Figure 4(a)). The Kaplan-Meier (KM) survival curve revealed that low-RS patients had consistently higher survival rates than high-RS patients (Figure 4(b)). Patients were ranked from lowest to the highest risk score, with the median risk score of 1.0 represented by the

Table 1. Coefficients of multivariate cox regression analysis.

Gene	Coef	HR	HR.95L	HR.95H	P value
CXCL13	-0.06627	0.935876	0.869642	1.007155	0.076794
CCL20	0.118971	1.126337	1.052461	1.205399	0.000588
CX3CR1	-0.19912	0.819449	0.719571	0.933191	0.002677
CXCL4L1	-0.2435	0.78388	0.65214	0.942232	0.009492

horizontal dotted line (Figure 4(c)). As the RS increased, the duration of patient survival diminished and the number of deaths increased (Figure 4(d)). Regression analysis was used to identify prognostic outcomes of clinical indicators. Univariate regression analysis showed RS, T, N, M and stage were related to the prognosis of patients (Figure 4(e)). Multivariate regression analysis showed RS, T, N, M were independent prognostic factors (Figure 4(f)). We combined RS and clinicopathological features to draw 3-year (Figure 4(g)) and 5-year ROC curves (Figure 4(h)) to predict the accuracy of the model. The larger the area under the curve, the better the accuracy of the model. Both the 3- and 5-year ROC curves showed that the AUC of RS was higher than other clinical features.

The expression of risk genes was employed to create a nomogram for predicting survival time (Figure 5(a)). The "Points" axis has been normalized to 0 to 100, and the total points can be obtained based on the expression of the risk genes to predict the survival rate of 3, 5, 7 years. The calibration curve was designed to demonstrate the nomogram's precision. It can be seen that the red line is close to the gray line, which indicates that the predicted result is consistent with the actual result (Figure 5(b)). Concordance index (C-index = 0.64) could evaluate discrimination power of the nomogram.

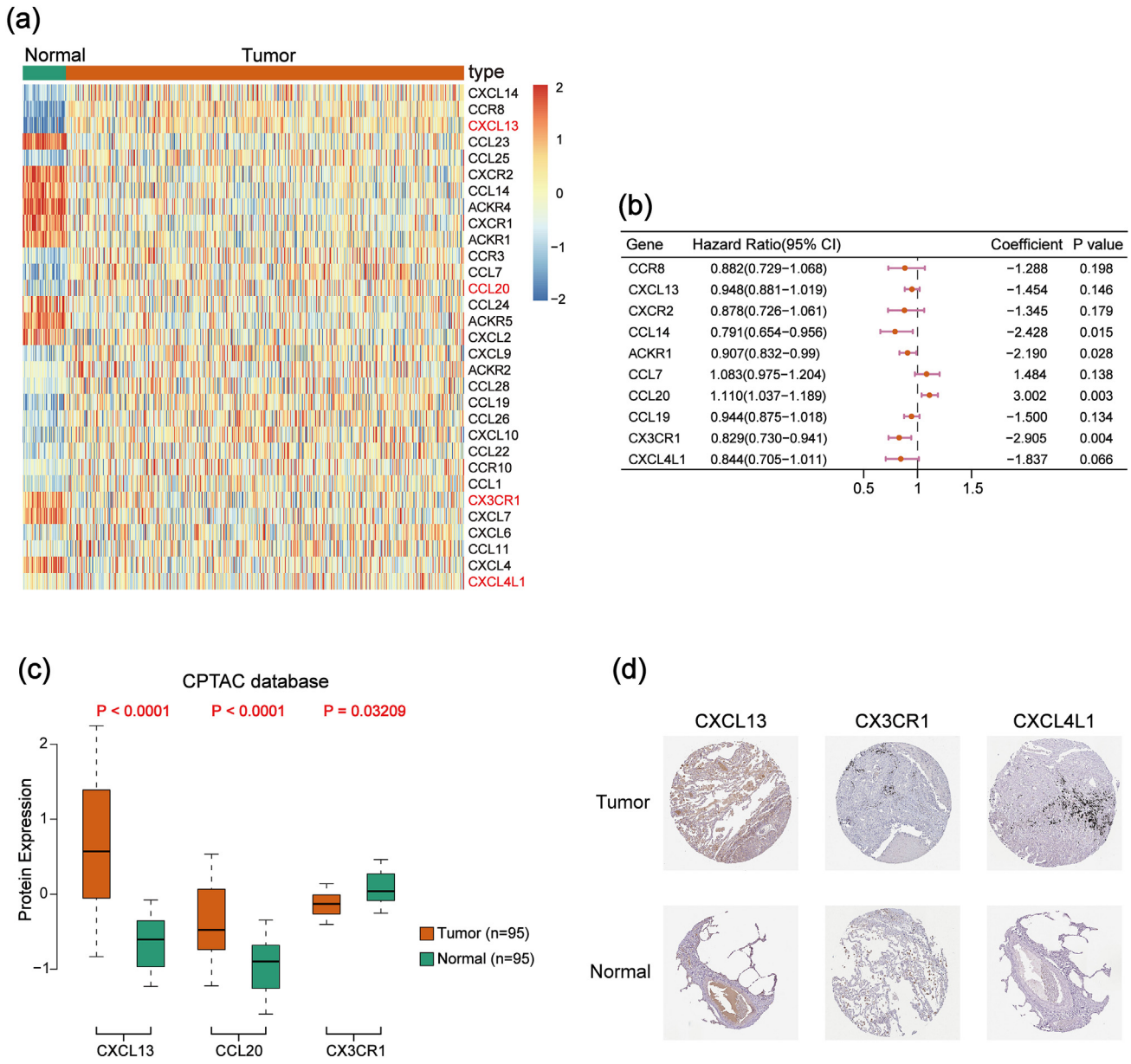

Figure 3. Differential expression analysis between normal and tumor tissues and univariate analysis in LUAD. (a) Heatmap of chemokines that were differentially expressed in TCGA. **(b)** Forest map of univariate analysis in TCGA. HR, the hazard ratios; 95%CI, 95% confidence intervals. **(c)** The protein expression level of CXCL13, CCL20, and CX3CR1 in the CPTAC database (n = 190). **(d)** The protein expression level of CXCL13, CX3CR1, CXCL4L1 in the HPA database. CXCL13 (HPA052613) staining in tumor/normal: low/not detected; CX3CR1(HPA046587) in tumor/normal: not detected/medium; CXCL4L1 (HPA052485) staining in tumor/normal: not detected/not detected.

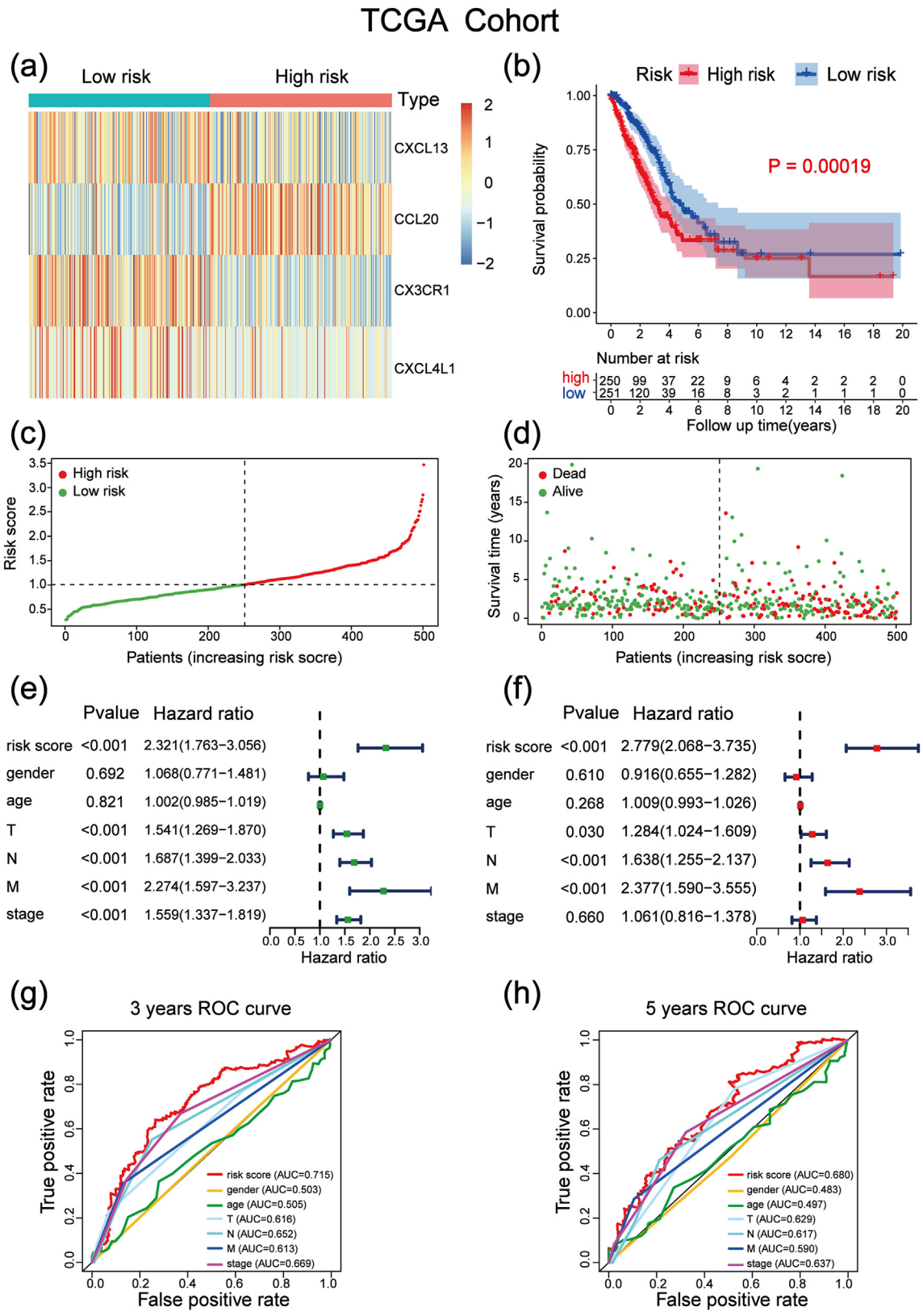

Figure 4. Establishment and evaluation of risk prognosis model in TCGA-LUAD. (a) The heatmap showed the four genes' expression in high/low-RS group. **(b)** The survival curve exhibited prognostic difference between patients in the high/low-RS groups over the period of 20 years. **(c)** The risk scores of 501 patients increased from left to right, with a median risk score of 1.0. The horizontal coordinate is the patient number and the vertical coordinate is the risk score for each patient. **(d)** The scatter plot reflected the survival time and status of patients from low-RS to high-RS. **(e)** Univariate analysis showed that risk score, T, N, M, stage was significantly correlated with prognosis. **(f)** Multivariate analysis showed that risk score, T, N, M were independent factors affecting prognosis **(g-h)** The ROC curves at 3 and 5 years showed the largest AUC values for risk score compared to other common clinical features, indicating that risk score has a stronger performance in predicting prognosis. ROC, Receiver Operator Characteristic; AUC, Area Under the Curve.

3.3. The prognostic model was validated with GEO

The data from GSE31210 was utilized to validate the risk signatures. The median value of the risk score in the GEO cohort of 226 patients was 0.95 according to the risk score formula. 113 of 226 samples were separated into the low-RS group and the rest into the high-RS group. The expression of risk genes and the survival rate of patients in the high/low-RS group were both in line with the TCGA findings (Supplementary Figure 1(a)–(d)). Different from TCGA, RS and stage were both independent prognostic factors through univariate (Supplementary Figure 1(e)) and multivariate regression analysis (Supplementary Figure 1(f)). The area under the curve was almost equal to that of the TCGA cohort (Supplementary Figure 1(g)–(h)). We also use GEO data to draw nomogram (Figure 5(c)) and calibration curve (Figure 5(d)) and the C-index is 0.68. Overall, these four chemokines were used to construct risk models that accurately predict patient outcomes.

3.4. The risk signature defined distinct enrichment pathway, immune score, immune cells, and immune checkpoint genes in TCGA-LUAD

GSEA was used to analyze genes related to the high/low-RS group in TCGA for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis to examine potential biological function (Supplementary

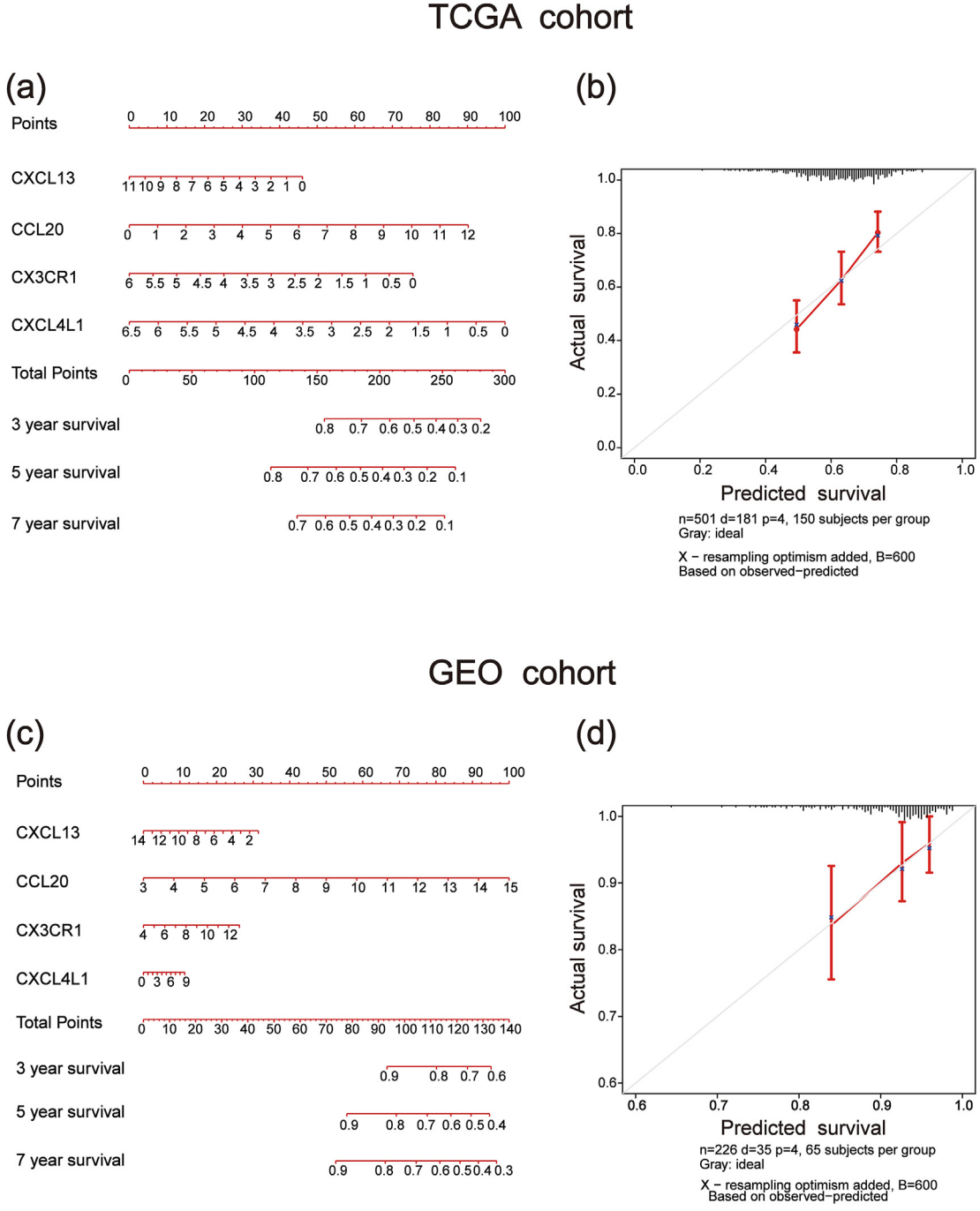

Figure 5. Building a nomogram and evaluating the predictive performance of the nomogram with calibration curves in TCGA and GEO cohort. (a, c) The nomogram was constructed based on risk gene expression. The total points are the sum of the points corresponding to the expression of the four risk genes for each patient, which corresponds to the 3-, 5- and 7-year predicted survival rate of the patient **(b, d)** Calibration curve for testing the predictive power of nomogram. The more the red dots overlap with the diagonal line, the stronger the predictive power of the nomogram.

Figure 2(a)). Enrichment scores (ES) and P-value of pathways with significance were provided in Supplementary Table 3. Genes related to high-RS samples were enriched in protein export, citrate cycle tca cycle, proteasome, etc (Supplementary Figure 2(b)–(d)), while genes related to low-RS samples were enriched in allograft rejection, asthma, chemokine signaling pathway, intestinal immune network for IGA production, primary immunodeficiency (Supplementary Figure 2(e)–(i)). Overall, patients with high-RS participated in pathways of metabolism or translation, while patients with low-RS participated in pathways of immune diseases or immune systems.

To understand the differences between the high-RS group and the low-RS group in TME, we calculated immune/stromal/estimate scores and tumor purity. Immune/stromal/estimate scores were significantly higher in low-RS samples than in high-RS samples and were associated with good prognosis, while the opposite was true for tumor purity (Figure 6(a)–(h)). Next, we calculated the proportion of immune cells (Figure 6(i)). The results indicated macrophages accounted for the most proportion, followed by T cells. The proportion of 10 immune cells between the two groups exhibited significant difference (B cells memory, plasma cells, T cells CD4 memory resting, T cells CD4 memory activated, T cells gamma delta, monocytes, macrophages M0, dendritic cells resting, mast cells resting, mast cells activated) (Figure 6(j)). The relevance between immune cells and RS was shown in Figure 6(k). RS has strong negative correlation with mast cells resting, dendritic cells resting, B cells memory, and strong positive correlation with dendritic cells activated, mast cells activated. Combining Figure 6(j) and Figure 6(k), we found high proportions of mast cells resting and dendritic cells were associated with good prognosis.

To explore the application of immune checkpoint inhibitors in patients with lung adenocarcinoma, we visualized the expression of 19 immune checkpoint genes in the two groups [28, 29, 30] (Supplementary Figure 3). The results revealed that low-RS samples have lower expression of FGL1, but higher expression of BTLA, LAG3, CTLA4, CD274, PDCD1, VTCN1, CD27, CD40, ICOS, TNFRSF4, PDCD1LG2, HHLA2, HAVCR2 and VSIR. This suggests that lung adenocarcinoma patients at low risk could benefit from immune checkpoint inhibitors.

3.5. Construction of a ceRNA network

Regression analysis and KM survival analysis suggested that CCL20 might be an independent biomarker in LUAD. Besides, high expression of CCL20 is related to bad outcome (Supplementary Figure 4(a) and (c)). Moreover, the expression distribution in different clinical stages showed that CCL20 expression was positively correlated to the advanced clinical stage (Supplementary Figure 4(b) and (d)). This suggested that CCL20 might take part in LUAD progression. We reversely predicted miR-21–5p as the targeting miRNA sticking to CCL20 via TarBaseV.8, mirTarBase and TargetScan for figuring out the possible molecular mechanism of CCL20 in lung adenocarcinoma. Then miR-21–5p was used to predict upstream lncRNA, and the intersection of LncBase Predicted v.2 and StarBase was 8 lncRNAs (AC009299.3, FAM66E, OTUD6B-AS1, NUTM2A-AS1, SNHG1, MIR17HG, BRWD1-IT1 and XIST) (Figure 7(a)). The violin diagram displayed that miR-21–5p expression was upregulated in tumors (Figure 7(b)) and correlated with poor prognosis (Figure 7(c)). Among the 8 lncRNAs, only the expression of MIR17HG and AC009299.3 was significantly correlated with survival prognosis (Figure 7(d)–(e)), and survival curves of the remaining 6 lncRNAs were shown in Supplementary Figure 5. Therefore, the two lncRNAs may be upstream targets for the regulation of chemokines in LUAD. The ceRNA network was shown in Figure 7(f).

4. Discussion

Early detection and prompt treatment are crucial to the recovery of lung cancer patients. However, due to the asymptomatic nature of LUAD, many patients are not diagnosed until the middle and advanced clinical stage, which greatly reduces the effectiveness of the treatment [31]. In recent years, precision medicine at the molecular level has developed rapidly, and many targeted therapeutic drugs have emerged, improving the quality of patients' life. But there is a great limitation that only a few LUAD patients are ideal candidates for targeted therapy and patients gradually appear the problem of targeted drug resistance with the extension of treatment time. In order to make more patients have access to treatment, we established a chemokine-related prognostic model to provide the possibility of new therapeutic targets for clinical treatment. Chemokines regulate immune cell chemotaxis and play a crucial role in immunological surveillance. Besides, they participate in angiogenesis, angiogenesis inhibition and tumor development and metastasis [32, 33]. Chemokines in TME have been reported to be of great importance during cancer's tumorigenesis for the past few years [34].

In this study, chemokine mutation, CNV and expression were analyzed. The results revealed that genetic alteration might lead to the dysregulation of chemokines. To build the risk prognostic model, differentially expressed genes were selected. Subsequently, we analyzed the patients' survival prognosis between the high-RS group and the low-RS group as well as clinical independent prognosis and verified it with GSE31210 data. We also applied the risk model to GSEA and immune infiltration analysis and discovered the high-RS group was correlated to metabolism and translation, while the low-RS group was correlated to immune-related pathways. The number, type, and location of tumor immune infiltration are crucial to the prediction of cancer outcome. Patients in the high/low-RS group have different proportions of immune cell infiltration. The results of the combined ESTIMATE and CIBERSORT analysis showed that high-RS might indicate immunosuppressive. Furthermore, the expression of immune checkpoint genes significantly increased in the low-RS group, which meant that low-RS patients might benefit from the immune checkpoint inhibitors (ICIs) therapy.

Prognostic genes including CCL20, CX3CR1, CXCL4L1 and CXCL13 might be associated with the prognosis of LUAD patients. Especially, regression analysis and survival analysis indicated high CCL20 expression was related to poor prognosis. In addition, higher CCL20 expression is positively correlated to the advanced clinical stage. Previous study showed that CCL20 promotes tumor cell growth and metastasis in lung cancer [35, 36, 37]. CXCL4L1 inhibits the growth and metastasis of lung adenocarcinoma by preventing angiogenesis [38]. In serum, CXCL13 expression was significantly higher in the lung adenocarcinoma group than in the healthy control group, which was consistent with our previous analysis of both mRNA and protein sequencing data from solid tissues [39]. Results of CX3CR1 studies in lung cancer are controversial. A study showed the expression of CX3CR1 was significantly higher in lung cancer, promoting the tumor's development [39, 40]. But there are opposing views thought that in the early stages of anti-PD-1 therapy, the increased frequency of CX3CR1 subgroups in circulating CD8 T cells is associated with the improvement of response and survival of patients with NSCLC [41]. We consider the cause of this phenomenon may be the instability of gene expression caused by post-translational modification [42]. In short, we have successfully screened the chemokines to construct a risk model that could accurately forecast the prognosis of LUAD patients. In addition to that, we also found that high expression of CCL20 results in poor prognosis, and the expression is significantly different in variant clinical stages, which may be acted as a prognostic biomarker of LUAD.

LncRNA, as ceRNA, competes with miRNA to regulate biological functions to participate in the development of lung cancer [43]. We predicted and constructed a lncRNA MIR17HG/AC009299.3/-miR-21–5p/CCL20 regulation network. In fact, miR-21–5p is an oncogene with increased expression in many cancers [44, 45, 46]. In contrast, overexpression of MIR17HG inhibited the invasion and migration of cancer cells in NSCLC [47]. On the regulation of genes in the network, MIR17HG/miR-21 and miR21/CCL20 have verified the accuracy of binding sites by luciferase reporter assay. Overexpression of MIR17HG

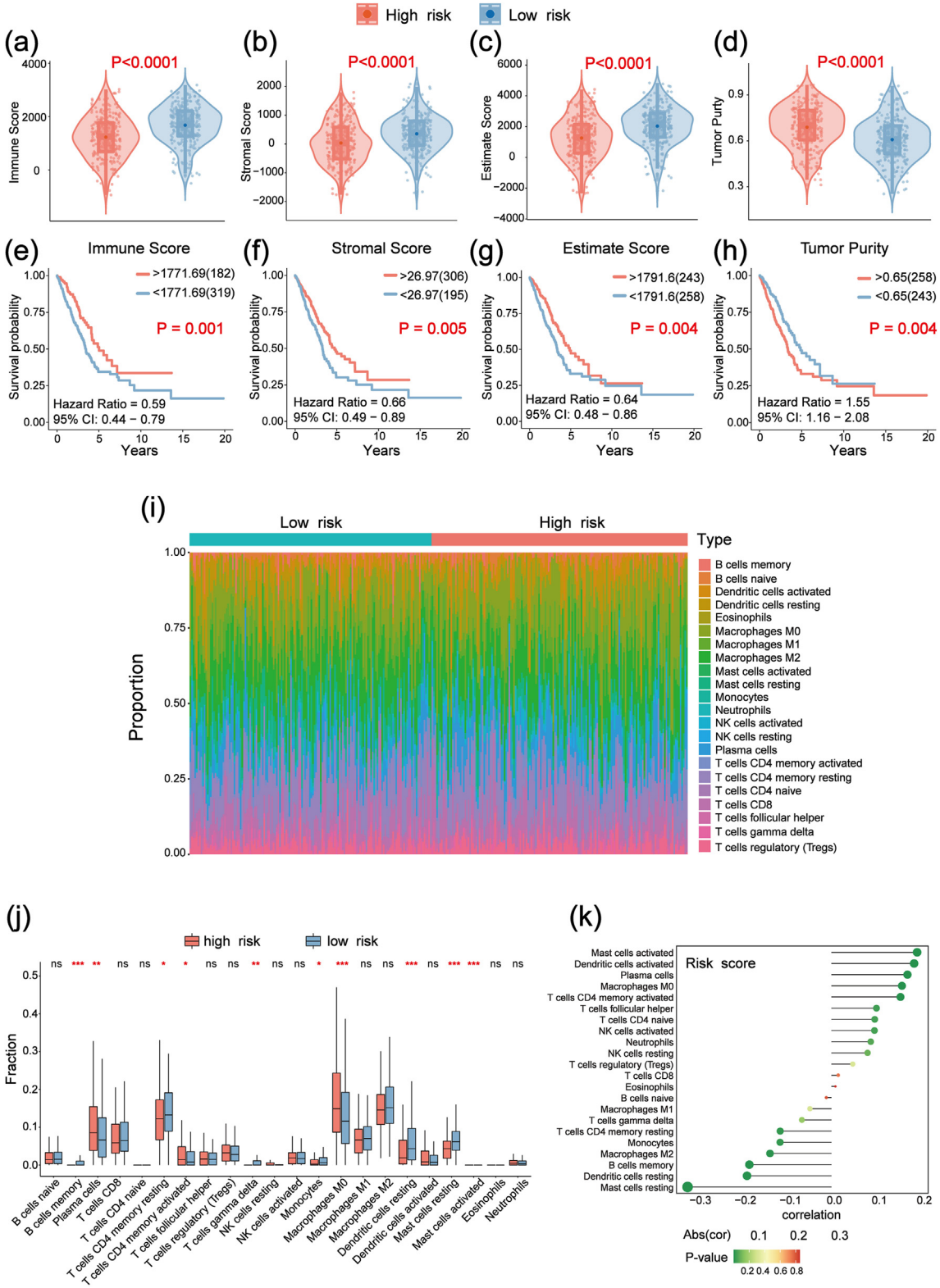

Figure 6. Immune score and immune cells analysis in TCGA-LUAD. (a-d) Correlation of immune score/stromal score/estimate score/tumor purity with respect to the risk group **(e-h)** The survival curve showed the correlation between prognosis and immune score/stromal score/estimate score/tumor purity over the period of 20 years. **(i)** The proportions of immune cells in high-RS and low-RS groups **(j)** The infiltration fraction of 22 immune cells in tumor microenvironment was compared between high-RS group and low-RS group (ns, not significant; *P < 0.05; **P < 0.01; ***P < 0.001) **(k)** Correlation between RS and the ratio of twenty-two immune cells.

suppresses acute myeloid leukemia by sponging miR-21 to promote apoptosis and miR-21 promotes the development of colorectal cancer and cervical squamous carcinoma by regulating CCL20 [48, 49, 50]. These conclusions provide some evidence for our findings. AC009299.3 has not been studied after our investigation. Anyway, the expression of the three RNAs was associated with the prognosis of LUAD, suggesting that the

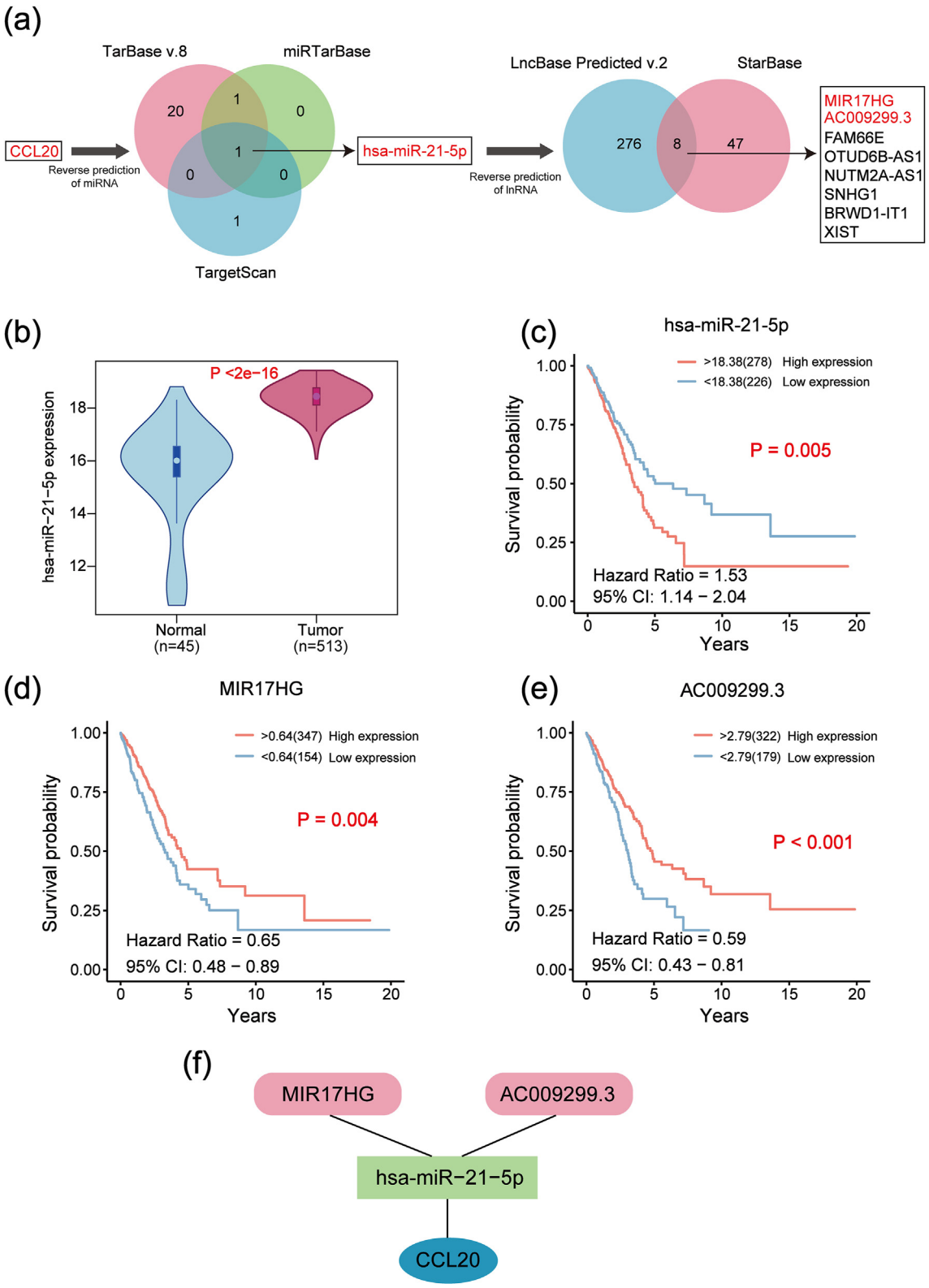

Figure 7. Establishment of a CCL20-associated ceRNA network. (a) The miRNA and lncRNA that regulate risk gene CCL20 were reversely predicted through the database. TarBase v.8, miRTarBase, and TargetScan were used to predict the common miRNA (has-miR-21–5p) regulating the target gene CCL20. LncBase Predicted v.2 and StarBase was used to predict common lncRNAs based on has-miR-21–5p. **(b)** The expression of hsa-miR-21–5p between 513 lung adenocarcinoma tissues and 45 normal tissues in TCGA ($P < 2e\text{-}16$) **(c-e)** The prognosis value of hsa-miR-21–5p and lncRNA (MIR17HG, AC009299.3) in TCGA-LUAD. **(f)** The ceRNA network. ceRNA, competing endogenous RNA; lncRNA, long non-coding RNA; miRNA, microRNA; mRNA, messenger RNA.

lncRNA MIR17HG/AC009299.3/miR-21–5p/CCL20 regulatory network might be of great importance during the progression of LUAD. More research needs to be performed to validate the discovery.

Although the chemokine risk model was confirmed to have strong performance for the prognosis of LUAD, we have to admit that the study still has some limitations. First of all, the use of a single gene set to

establish a predictive model is an unavoidable defect, since the development and progression of lung adenocarcinoma are actually influenced by multiple mechanisms. Secondly, immune score and immune cells in TME were calculated by advanced algorithms, so the predictive power of the model for immunotherapy response is indirect. Thirdly, the data used for analysis were downloaded from public databases. Although the conclusions are supported by some evidence from the literature, the risk genes model and the construction of the ceRNA network for lung adenocarcinoma would be more convincing if fresh human tissue samples could be collected or cells obtained for experimental validation of this study.

5. Conclusion

In conclusion, our study comprehensively elucidated the mutational pattern of chemokines in LUAD and constructed the prognostic signature and lncRNA MIR17HG/AC009299.3/miR-21–5p/CCL20 regulatory network. This study will help enhance our understanding about the role of chemokines, provide new independent prognostic factors and guide effective immunotherapy strategies for lung cancer.

Declarations

Author contribution statement

Yifan Hu, Mintao Xiao, Duoli Zhang: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Jing Shen, Yueshui Zhao, Mingxing Li, Xu Wu: Performed the experiments; Analyzed and interpreted the data.

Yu Chen, Zhigui Wu, Hongli Luo: Conceived and designed the experiments; Wrote the paper.

Zhangang Xiao, Fukuan Du: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

Mr Fukuan Du was supported by the Sichuan Science and Technology Program, China [No. 2022NSFSC0783], Strategic Cooperation Project for Transfer and Transformation of Scientific and Technological achievements of Southwest Medical University and Lu County Government [No. 2019LXXNYKD-07], Joint Founds of Southwest Medical University and Luzhou Government [No. 2020LZXNYDJ08], Funds of talent introduction and scientific research of Southwest Medical University [No. 05-00040140], Science and Technology Program of Luzhou, China [No. 21CGZHPT0001].

Zhigui Wu was supported by Project of Science and Technology Department of Sichuan Province [No. 2021YJ0445].

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e11399.

References

[1] R.L. Siegel, K.D. Miller, H.E. Fuchs, et al., Cancer statistics, 2021, CA A Cancer J. Clin. 71 (2021) 7–33.
[2] C. Gridelli, A. Rossi, D.P. Carbone, et al., Non-small-cell lung cancer, Nat. Rev. Dis. Prim. 1 (2015) 15009.
[3] G. Liu, F. Pei, F. Yang, et al., Role of autophagy and apoptosis in non-small-cell lung cancer, Int. J. Mol. Sci. 18 (2017) 2017.
[4] N. Cancer Genome Atlas Research, Comprehensive molecular profiling of lung adenocarcinoma, Nature 511 (2014) 543–550.
[5] W.W. van Willigen, M. Bloemendal, W.R. Gerritsen, et al., Dendritic cell cancer therapy: vaccinating the right patient at the right time, Front. Immunol. 9 (2018) 2265.
[6] P. Sharma, S. Hu-Lieskovan, J.A. Wargo, et al., Primary, adaptive, and acquired resistance to cancer immunotherapy, Cell 168 (2017) 707–723.
[7] A. Mantovani, P. Allavena, A. Sica, et al., Cancer-related inflammation, Nature 454 (2008) 436–444.
[8] J.W. Griffith, C.L. Sokol, A.D. Luster, Chemokines and chemokine receptors: positioning cells for host defense and immunity, Annu. Rev. Immunol. 32 (2014) 659–702.
[9] N. Nagarsheth, M.S. Wicha, W. Zou, Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy, Nat. Rev. Immunol. 17 (2017) 559–572.
[10] V. Mollica Poeta, M. Massara, A. Capucetti, et al., Chemokines and chemokine receptors: new targets for cancer immunotherapy, Front. Immunol. 10 (2019) 379.
[11] Y. Gu, X. Li, Y. Bi, et al., CCL14 is a prognostic biomarker and correlates with immune infiltrates in hepatocellular carcinoma, Aging (Albany NY) 12 (2020) 784–807.
[12] L. Jiang, D. Wang, M. Sheng, et al., CXCL13/CXCR5 are potential biomarkers for diagnosis and prognosis for breast cancer, J BUON 25 (2020) 2552–2561.
[13] H. Okayama, T. Kohno, Y. Ishii, et al., Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas, Cancer Res. 72 (2012) 100–111.
[14] M. Yamauchi, R. Yamaguchi, A. Nakata, et al., Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma, PLoS One 7 (2012), e43923.
[15] A. Zlotnik, O. Yoshie, The chemokine superfamily revisited, Immunity 36 (2012) 705–716.
[16] N.J. Edwards, M. Oberti, R.R. Thangudu, et al., The CPTAC data portal: a resource for cancer proteomics research, J. Proteome Res. 14 (2015) 2707–2713.
[17] M. Uhlen, L. Fagerberg, B.M. Hallstrom, et al., Proteomics. Tissue-based map of the human proteome, Science 347 (2015) 1260419.
[18] A. Subramanian, P. Tamayo, V.K. Mootha, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.
[19] V.K. Mootha, C.M. Lindgren, K.F. Eriksson, et al., PGC-1 alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes, Nat. Genet. 34 (2003) 267–273.
[20] K. Yoshihara, M. Shahmoradgoli, E. Martinez, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.
[21] A.M. Newman, C.L. Liu, M.R. Green, et al., Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (2015) 453–457.
[22] D. Karagkouni, M.D. Paraskevopoulou, S. Chatzopoulos, et al., DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions, Nucleic Acids Res. 46 (2018) D239–D245.
[23] H.Y. Huang, Y.C. Lin, J. Li, et al., miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database, Nucleic Acids Res. 48 (2020) D148–D154.
[24] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
[25] M.D. Paraskevopoulou, I.S. Vlachos, D. Karagkouni, et al., DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts, Nucleic Acids Res. 44 (2016) D231–238.
[26] J.H. Li, S. Liu, H. Zhou, et al., starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, Nucleic Acids Res. 42 (2014) D92–97.
[27] S.J. Kang, Y.R. Cho, G.M. Park, et al., Predictors for functionally significant in-stent restenosis: an integrated analysis using coronary angiography, IVUS, and myocardial perfusion imaging, JACC Cardiovasc Imaging 6 (2013) 1183–1190.
[28] D.M. Pardoll, The blockade of immune checkpoints in cancer immunotherapy, Nat. Rev. Cancer 12 (2012) 252–264.
[29] J. Wang, M.F. Sanmamed, I. Datar, et al., Fibrinogen-like protein 1 is a major immune inhibitory ligand of LAG-3, Cell 176 (2019) 334–347.
[30] M. Janakiram, U.A. Shah, W. Liu, et al., The third group of the B7-CD28 immune checkpoint family: HHLA2, TMIGD2, B7x, and B7-H3, Immunol. Rev. 276 (2017) 26–39.
[31] Q. Du, E. Li, Y. Liu, et al., CTAPIII/CXCL7: a novel biomarker for early diagnosis of lung cancer, Cancer Med. 7 (2018) 325–335.
[32] D. Raman, T. Sobolik-Delmaire, A. Richmond, Chemokines in health and disease, Exp. Cell Res. 317 (2011) 575–589.
[33] E. Marcuzzi, R. Angioni, B. Molon, et al., Chemokines and chemokine receptors: orchestrating tumor metastasization, Int. J. Mol. Sci. 20 (2018) 2018/12–29.
[34] Y. Cao, H. Huang, Z. Wang, et al., The inflammatory CXC chemokines, GROalpha(high), IP-10(low), and MIG(low), in tumor microenvironment can Be used as new indicators for non-small cell lung cancer progression, Immunol. Invest. 46 (2017) 361–374.
[35] B. Wang, L. Shi, X. Sun, et al., Production of CCL20 from lung cancer cells induces the cell migration and proliferation through PI3K pathway, J. Cell Mol. Med. 20 (2016) 920–929.
[36] G.Z. Wang, X. Cheng, X.C. Li, et al., Tobacco smoke induces production of chemokine CCL20 to promote lung cancer, Cancer Lett. 363 (2015) 60–70.

[37] X.P. Zhang, Z.J. Hu, A.H. Meng, et al., Role of CCL20/CCR6 and the ERK signaling pathway in lung adenocarcinoma, Oncol. Lett. 14 (2017) 8183–8189.
[38] S. Struyf, M.D. Burdick, E. Peeters, et al., Platelet factor-4 variant chemokine CXCL4L1 inhibits melanoma and lung carcinoma growth and metastasis by preventing angiogenesis, Cancer Res. 67 (2007) 5940–5948.
[39] R. Singh, P. Gupta, G.H. Kloecker, et al., Expression and clinical significance of CXCR5/CXCL13 in human nonsmall cell lung carcinoma, Int. J. Oncol. 45 (2014) 2232–2240.
[40] Y. Liu, H. Ma, T. Dong, et al., Clinical significance of expression level of CX3CL1-CX3CR1 axis in bone metastasis of lung cancer, Clin. Transl. Oncol. 23 (2021) 378–388.
[41] T. Yamauchi, T. Hoki, T. Oba, et al., T-cell CX3CR1 expression as a dynamic blood-based biomarker of response to immune checkpoint inhibitors, Nat. Commun. 12 (2021).
[42] L. Zhou, S.H. Li, Y. Wu, et al., Establishment of a prognostic model of four genes in gastric cancer based on multiple data sets, Cancer Med. 10 (2021) 3309–3322.
[43] M. Zhao, J. Feng, L. Tang, Competing endogenous RNAs in lung cancer, Cancer Biol Med 18 (2021) 1–20.
[44] Z.L. Liu, H. Wang, J. Liu, et al., MicroRNA-21 (miR-21) expression promotes growth, metastasis, and chemo- or radioresistance in non-small cell lung cancer cells by targeting PTEN, Mol. Cell. Biochem. 372 (2013) 35–45.
[45] H. Xia, W. Zhang, B. Zhang, et al., miR-21 modulates the effect of EZH2 on the biological behavior of human lung cancer stem cells in vitro, Oncotarget 8 (2017) 85442–85451.
[46] T. Wang, Z. Cai, G. Hong, et al., MicroRNA21 increases cell viability and suppresses cellular apoptosis in nonsmall cell lung cancer by regulating the PI3K/Akt signaling pathway, Mol. Med. Rep. 16 (2017) 6506–6511.
[47] S. Wei, J. Liu, X. Li, et al., LncRNA MIR17HG inhibits non-small cell lung cancer by upregulating miR-142-3p to downregulate Bach-1, BMC Pulm. Med. 20 (2020) 78.
[48] J. Yan, L. Yao, P. Li, et al., Long non-coding RNA MIR17HG sponges microRNA-21 to upregulate PTEN and regulate homoharringtonine-based chemoresistance of acute myeloid leukemia cells, Oncol. Lett. 23 (2022) 24.
[49] B. Vicinus, C. Rubie, S.K. Faust, et al., miR-21 functionally interacts with the 3'UTR of chemokine CCL20 and down-regulates CCL20 expression in miR-21 transfected colorectal cancer cells, Cancer Lett. 316 (2012) 105–112.
[50] T. Yao, Z. Lin, MiR-21 is involved in cervical squamous cell tumorigenesis and regulates CCL20, Biochim. Biophys. Acta 1822 (2012) 248–260.